PUBLIC ADMINISTRATION AND PUBLIC POLICY

A Comprehensive Publication Program

Executive Editor

JACK RABIN
Professor of Public Administration and Public Policy
School of Public Affairs
The Capital College
The Pennsylvania State University—Harrisburg
Middletown, Pennsylvania

1. Public Administration as a Developing Discipline (in two parts), Robert T. Golembiewski
2. Comparative National Policies on Health Care, Milton I. Roemer, M.D.
3. Exclusionary Injustice: The Problem of Illegally Obtained Evidence, Steven R. Schlesinger
5. Organization Development in Public Administration (in two parts), edited by Robert T. Golembiewski and William B. Eddy
7. Approaches to Planned Change (in two parts), Robert T. Golembiewski
8. Program Evaluation at HEW (in three parts), edited by James G. Abert
9. The States and the Metropolis, Patricia S. Florestano and Vincent L. Marando
17. Organizational Behavior and Public Management, Debra W. Stewart and G. David Garson
21. Labor Relations in the Public Sector, Richard C. Kearney
25. Decision Making in the Public Sector, edited by Lloyd G. Nigro
26. Managing Administration, edited by Jack Rabin, Samuel Humes, and Brian S. Morgan
28. State and Local Government Administration, edited by Jack Rabin and Don Dodd
29. Public Administration: A Bibliographic Guide to the Literature, Howard E. McCurdy
32. Public Administration in Developed Democracies: A Comparative Study, edited by Donald C. Rowat
34. Handbook on Human Services Administration, edited by Jack Rabin and Marcia B. Steinhauer
37. The Guide to the Foundations of Public Administration, Daniel W. Martin
41. Handbook of Comparative and Development Public Administration, edited by Ali Farazmand
43. Government Financial Management Theory, Gerald J. Miller
45. Public Productivity Handbook, edited by Marc Holzer
46. Handbook of Public Budgeting, edited by Jack Rabin
48. Handbook of Organizational Consultation, edited by Robert T. Golembiewski
49. Handbook of Court Administration and Management, edited by Steven W. Hays and Cole Blease Graham, Jr.
50. Handbook of Comparative Public Budgeting and Financial Management, edited by Thomas D. Lynch and Lawrence L. Martin
51. Handbook of Organizational Behavior, edited by Robert T. Golembiewski
52. Handbook of Administrative Ethics, edited by Terry L. Cooper
54. Handbook of Regulation and Administrative Law, edited by David H. Rosenbloom and Richard D. Schwartz
55. Handbook of Bureaucracy, edited by Ali Farazmand
57. Practical Public Management, Robert T. Golembiewski
60. **Handbook of Debt Management**, edited by Gerald J. Miller
62. **Handbook of Local Government Administration**, edited by John J. Gargan
63. **Handbook of Administrative Communication**, edited by James L. Garnett and Alexander Kouzmin
67. **Handbook of Public Finance**, edited by Fred Thompson and Mark T. Green
69. **Handbook of Economic Development**, edited by Kuotsai Tom Liou
70. **Handbook of Health Administration and Policy**, edited by Anne Osborne Kilpatrick and James A. Johnson
71. **Handbook of Research Methods in Public Administration**, edited by Gerald J. Miller and Marcia L. Whicker
72. **Handbook on Taxation**, edited by W. Bartley Hildreth and James A. Richardson
73. **Handbook of Comparative Public Administration in the Asia-Pacific Basin**, edited by Hoi-kwok Wong and Hon S. Chan
74. **Handbook of Global Environmental Policy and Administration**, edited by Dennis L. Soden and Brent S. Steel
75. **Handbook of State Government Administration**, edited by John J. Gargan
76. **Handbook of Global Legal Policy**, edited by Stuart S. Nagel
77. **Handbook of Public Information Systems**, edited by G. David Garson
78. **Handbook of Global Economic Policy**, edited by Stuart S. Nagel
80. **Handbook of Global International Policy**, edited by Stuart S. Nagel
82. **Handbook of Global Political Policy**, edited by Stuart S. Nagel
83. **Handbook of Global Technology Policy**, edited by Stuart S. Nagel
84. **Handbook of Criminal Justice Administration**, edited by M. A. DuPont-Morales, Michael K. Hooper, and Judy H. Schmidt
85. **Labor Relations in the Public Sector: Third Edition**, edited by Richard C. Kearney
88. **Handbook of Global Social Policy**, edited by Stuart S. Nagel and Amy Robb
89. **Public Administration: A Comparative Perspective, Sixth Edition**, Ferrel Heady
90. **Handbook of Public Quality Management**, edited by Ronald J. Stupak and Peter M. Leitner
91. **Handbook of Public Management Practice and Reform**, edited by Kuotsai Tom Liou
95. **Financial Planning and Management in Public Organizations**, Alan Walter Steiss and Emeka O. Cyprian Nwagwu
96. **Handbook of International Health Care Systems**, edited by Khi V. Thai, Edward T. Wimberley, and Sharon M. McManus
97. Handbook of Monetary Policy, edited by Jack Rabin and Glenn L. Stevens
98. Handbook of Fiscal Policy, edited by Jack Rabin and Glenn L. Stevens
102. Strategic Management for Public and Nonprofit Organizations, Alan Walter Steiss

Additional Volumes in Preparation

Principles and Practices of Public Administration, edited by Jack Rabin, Robert F. Munzenrider, and Sherrie M. Bartell

Handbook of Developmental Policy Studies, edited by Stuart S. Nagel

Handbook of Conflict Management, edited by William J. Pammer, Jr., and Jerri Killian

ANNALS OF PUBLIC ADMINISTRATION

2. Public Administration Education in Transition, edited by Thomas Vocino and Richard Heimovics
3. Centenary Issues of the Pendleton Act of 1883, edited by David H. Rosenbloom with the assistance of Mark A. Emmert
4. Intergovernmental Relations in the 1980s, edited by Richard H. Leach
To Peg, my companion in moving toward awareness of *my* human irony, which can be expressed in such ways:

I am most myself when we are together
I am strongest when I recognize my multiple dependencies on her
I am most comfortably alone in her presence.

Those variants all say much the same thing, one more time, *moja kochana.*
Preface

ORIENTATION TO THE SECOND EDITION

This second edition of *Ironies in Organizational Development* contains 25 chapters plus an overview and they represent a complex mixture—of the old and the new, of the tried-and-true as well as the speculative, of the here-and-now with a dash or two of the by-and-by. To provide some specificity, most of the major components of this volume did not appear at all in the first edition, and 11 chapters were part of the previous book but are revised here—bringing points up to date, citing new research, and eliminating mistakes. Some questions follow, quite naturally. Why this particular assemblage? And why now?

Some early delineation of what follows and why, has a high priority. Let’s focus immediately on several aspects of why, saving the what for later in this Preface.

Organizational Development (OD) practice has been quite successful, by and large. Nonetheless, there has been too little translation of that success into broad analytical frameworks that not only reflect an intellectual grasp of what we know but also substantially enrich praxis. This expresses the basic irony motivating this book. One reader perceptively expressed the barrier that this book seeks to help surmount—or, perhaps better said, to transform from liability into substantial asset:

Applications have not yet yielded theory. The consequence of this is that too often OD successes rest on the talent and instinct of the intervenor. Generalizability of outcomes, and even more widespread application of successful techniques, will inevitably be constrained to the degree that
Now is a very good time for me to summarize what we have learned, as well as to be clearer about reasonable aspirations for what is still to come. Having passed normal retirement age, I do not sense that time has run out for me. But a new urgency has taken hold, without doubt, and this second edition is one manifestation of this being the time for more integrative summary than was possible a decade ago.

So, let’s get going with an emphasis on what we can expect from OD, with a bit more about why. Four emphases provide further and useful delineation for present purposes, although readers no doubt will have their multiple and additional reactions. Note also that a few central citations can only be suggestive, if only because the full body of my OD work probably encompasses 400 published items. So far, I have not counted.

It has become my habit to respond to a pervasive rhythm in my work in OD, in part by plan and in part by serendipity. Either alone or usually with several OD reference teams, I have gone through a number of approximately 10-year cycles. Typically, each begins with an application—more likely in business than in public agencies, and more likely at middle to executive levels of management than at operating levels. Working materials and article-length pieces will be produced during the early years of each cycle, but I like to express their summary sense and substance in one or more hardcover books at the end of each cycle.

This intent dominated as 2000 closed another full cycle for me, and I trust it has the same double-barreled effects as its predecessors. That is, the hardcover formalizations of the period of applications in organizational and personal learning at once summarize several years of effort both by me and, typically, by several project teams with variable memberships, and those formalizations also serve as a platform whose syntheses and shortfalls should guide the next full cycle of more comprehensive applications. In short, each cycle looks backward, as it were, the better to see the next steps.

In introductory preview, this book targets the results of a number of application ↔ reflection cycles, with an emphasis on ironies that can be exploited to raise OD success rates.

**ABOUT SEVERAL CYCLES**

To be more specific, five cycles seem to adequately cover the territory referred to above. In reverse order, these cycles will be briefly outlined later. Immediately, the underlying philosophy has a direct form. Those who will not learn from their experience, especially from positive experience, will be condemned to relive their
history if they do not regress. Hence, there are special challenges in recognizing OD’s ironies and using them to leverage future performance.

Now for a sketch of the cycles, conveniently in reverse order.

**Cycle V, 1990–2000: Accents on Synthesis**

This cycle has a definite integrative and macro focus, along with several extensions of success rate studies. They represent the fullest detailing of planned change of which I am now capable, and build on the substantial efforts of others, sometimes in research in which I was involved but as a less central figure.

It seems both fitting and convenient to me that this fifth cycle’s expression be a second edition of *Ironies in Organizational Development* (1990), much enlarged and expanded. This assemblage is perhaps especially legitimated by the emphasis on large-organization dynamics, while earlier cycles tend to feature small-group arenas of application. Notably, both micro- and macro-levels feature the same values, methods, and approaches.

Cycle V is characterized by three specific features. First, several macro-programs have been completed, with the signal recognition being receipt of the 1997 Grand Award for OD Applications, Worldwide, from the Organization Development Institute. Second, studies of OD success rates also add great detail to lines of inquiry begun in the late 1970s and early 1980s. Third, much recent research is added. For example, substantial development was reported on the concept of psychological burnout, which showed great integrative potential in relating various themes in the behavioral sciences and OD, and built on early syntheses in the first edition of *Ironies in Organizational Development*. The leading book-length treatment was *Global Burnout*, by Golembiewski and associates (1996).

**Cycle IV, 1975–1990: Applications in Planned Change, Sometimes Macro-Scale and Often Conceptual**

This cycle featured extensive work in several substantial systems, in business and government. Both start-ups and organization renewals were involved. Related evaluations appeared in several sources (especially Golembiewski, 1985, and Golembiewski and Kiepper, 1985). Some of this work also was recognized with two McGregor Awards, the only double-winners in that competition. The major summary treatment appeared in *Ironies in Organization Development* (1990), as well as *Humanizing Public Organizations* (1986), and found most telling attention expression in trinitarian change.

Major emphases in Cycle IV deserve highlighting. Several macro-applications were started; new work in OD success rates saw several start-ups, all of
which, if not conclusive, at least were more ambitious than earlier studies. A large number of micro-applications were made and, in many cases, they also required extension and integration. In sum, Cycle IV was essentially in-process, and several of its themes required extension in Cycle V.

**Cycle III, 1970–1975: Applications in Planned Change, Largely Micro-Scale**

This cycle involved a kind of winnowing of sources in planned change via OD ventures in which I was involved, typically as a member of application teams. A major expression of this work was the two-volume *Approaches to Planned Change* (Golembiewski, 1979), which summarized numerous applications in organizations and introduced others. Key here was a project that combined basic concepts of change with a useful application (Golembiewski, 1976). This project won a McGregor Award.


Cycle II overlaps a bit with Cycle I, and their focus in common is instrumental. The *skill*-building was both quantitative and qualitative. For example, a Ford Foundation Grant in Mathematical Applications in Business reflected the first theme. And the latter theme was largely embodied in skill-building in process analysis and the “laboratory” model of change. A National Science Foundation grant was central here, and my progress in change-agent skills and attitudes was reflected in several sources: a book of readings that went through four editions (Golembiewski and Blumberg, 1970, 1972, 1973, and 1997), a “how to” book (Golembiewski and Blumberg, 1976), and a major effort to move both OD and myself into macro-analysis (Golembiewski, 1972).

**Cycle I, 1955–1965: Gearing Up in Concepts and Ethics**

Here, the emphasis is essentially on defining a career focus, and I have been pretty lucky, since I have had few problems of changing my mind when I made false starts. Almost from the start of my serious development, the issues have been:

- *Organizational*, as in large collective enterprises, both public and private, often approached by various arrays of designs building upon learning in small groups (e.g., Golembiewski, 1962c)
- *Normative*, as in what is right in our organizational lives, and what needs correction or fundamental revision
- *Instrumental*, or *action, research*, as in the design of ways and means of achieving normative goals in large organizations
The major products of this initial cycle appeared in several extended written products. They began with my doctoral dissertation (Golembiewski, 1958) and include several books (Golembiewski, 1962b, 1962c, 1965).

Notably, also, when I was transitioning from one cycle to another, I was very lucky in having available useful resources. Arthur Blumberg helped in early transitions from academic to consulting modes, and numerous others—especially Stokes Carrigan and Rick Hilles—played a central role by locating sites for applications that also had research potential.

CHAPTERS: OLD, NEW, AND OTHERWISE IN THIS SECOND EDITION

The chapters have their own life histories, and four classes of them can be distinguished. First, 15 chapters are published here for the first time in the present form. Second, the essentials of several other chapters were published originally in sources unlikely to gain the attention of ODers, in part because I was then less clear about the significant role the pieces could play in understanding the status of OD and, especially, about how to foster tomorrow’s development. Third, some chapters were in substance published in major sources, as well as in the first edition. Fourth, the basic contents of a few chapters were published originally in books that are now out of print. The hope is that they can gain new life from republication in the present context.

So why combine chapters with these diverse life histories into a single volume? For openers, at least two-thirds of the chapters probably will be news even for the reasonably well-read ODer, and the volume as a totality can provide an upscale view of the state of the art for many possible readers: for managers, for new specialists in OD (of whom there are always legions), and for students of organizations who do not follow the twists and turns of the OD literature but whose own work not only is relevant to OD but also can be enriched by it.

Moreover, the basic rationale for inclusion in this second edition (hereafter referred to as Ironies II) is that its chapters en masse can provide a kind of launching pad for theoretical progress in OD, which is often seen as pleasantly stuck—in other words, robust enough to support great and growing consulting activity, but having a theoretical base that lags far behind practice.

Basically, Ironies II seems to me to enrich both theory and practice, and I believe it does so in ways that will surprise even close readers of the original volume. Indeed, surprises of the good fits below often kept me going. In any case, the chapters whose essentials have been published typically were written as stand-alone pieces, and some people paid less attention to the words than to what they fancied as the underlying music—that OD was seriously inadequate and stuck, even if comfortably so.

A sharp contrast seems fair enough to me. Ironies II shows how OD is
quite effective, and yet can be substantially built upon and beyond. Several of its chapters, as stand-alone publications, seemed to say to some, especially to those applying OD as consultants: “Your baby sure is ugly.” This book corrects that perception by showing how even major criticisms are the foundation for major developments when presented in the context of other materials.

IRONIES AS CONTEXT OR MEDIUM FOR MESSAGES: THE NEW COMBINATION

Let me make the last point again: Ironies II provides a vehicle for rectifying such a misinterpretation of my basic view of OD, both by critics and by friends. This book provides not only the words, and some music, but also reasonably precise directions for producing a better OD tune from our past experiences pushed to outline OD’s future, if at times only dimly.

In addition, Ironies II provides a context that unifies and, even more significantly, highlights the constructive character of much research and application. Moreover, in the aggregate, the volume reflects the prudent hopefulness of that context. Individually, that is, most of the chapters pose real challenges for some OD practitioners and theorists, and may even encourage despondency. For both virginal and previously published chapters, Ironies II provides a developmental context and thrust that saves them from charges of mere carping, as that focus also highlights constructive aspects of positions that some saw as negative and as raining on OD’s parade.

Ironically, this book proposes two main arguments: that OD does quite well, in general, but that in numerous particulars it can do much better, with modest expenditures of wit and will. Ironies II provides a positive context for some individual chapters that could be interpreted as grumpy or even malevolent—as never being satisfied or, worse still, as implying grave doubts about the integrity of the entire OD enterprise. Rather, Ironies II as a totality encourages various “stretches” that will save us from the Dr. Feelgood-ism that sooner or later can only jeopardize vitality and growing comprehensiveness, but those stretches also need to be viewed as ways to improve on a level of general performance that is far from shabby. Those stretches are not cries of despair; they are intended as directions for theoretical development that will raise the level of the conscious practice of intervention in social systems. Consulting competencies and sensitivities often have filled in the gaps in theory, but we cannot always rely on that happy outcome.

So this book has a simple format, although details in specific chapters will be daunting.
Preface

Irony I: Substantial Success but Pessimism About Practice (Chapters 1–3)
Irony II: Substantial Success with Insufficient Attention to Replication (Chapters 4–6)
Irony III: Substantial Success in Global Applications While Neglecting Alternative Ethics at Work (Chapters 7–9)
Irony IV: Substantial Success Without Consensus About a Learning Model (Chapters 10–11)
Irony V: Substantial Success While Inadequately Assessing Large-System Interventions and Their Effects (Chapters 12–15)
Irony VI: Substantial Success Without Specifying Contextual Differences (Chapters 16–17)
Irony VII: Substantial Success Without Differentiating People (Chapters 18–20)
Irony VIII: Substantial Success While Neglecting Easy Pieces (Chapters 21–23)
Irony IX: Substantial Success Without Differentiating Kinds of Change and Designs (Chapters 24 and 25)
Postscripts About Multiple Ironies (Chapter 26)

Each of the first 25 chapters contributes toward the same conclusion, and the last provides a summary of that conclusion. All the chapters provide handy ways to enhance success rates in OD from their already substantial levels.

The final chapter constitutes neither the necessary nor the sufficient catalog of all ironies. Overall, the chapter does provide several “next bites.” They will sate Rabelaisian appetites, although this chapter will challenge those seeking to swallow “the whole thing.”

RIGHT NOW AS THE CONGENIAL MOMENT

Finally, right now seems an opportune time to bring together the present assemblage of chapters under the rubric of ironies. More accurately, I should write that right now is even more opportune than the time of the first edition.

And what are the characteristics of this “right now?” Two dominate. As the introduction and Chapters 1–3 show, OD theory and practice have a solid track record, and those several estimates of success rates cannot be dismissed cavalierly. Concern was expressed about OD’s efficacy even by its proponents and, curiously, a kind of minor despair set in among some aficionados in the face of earlier estimates in Ironies I. Today’s estimate provides a more solid base of optimism on which to build. Sharp increases in the ability to provide detailed illustrations of what can be done—either immediately or as items on variously
distant agendas—constitute the second major sense of right now as the opportune
time. The message to OD intervenors who are more comfortable with lower suc-
cess rates is clear: measure up or muster out.

In sum, the augmented optimism about success rates is here seen as providing
the critical medium in which challenging suggestions can be responded to—
specifically, by reducing defensiveness or despondency. In contrast, the message in Ironies I of several of the chapters—perhaps especially its final chapter—
came too soon. Those essential messages often came through as “Your baby sure
is ugly,” to people who had seen OD as comely, even handsome. Then, they
were unclear as to what was to be done about the critical messages. Ironies II
reflects the desire to test a new combination, better than its predecessor—to re-
spond to an older message in some cases, new contexts and information in all
cases.

THE INTENDED AUDIENCES

As noted above, the intended audiences are OD consultants, both old hands and
the large numbers of entrants that always seem to be with us; managers facing
the challenge of continuous change and seeking a sense of value-guided methods;
and the many researchers who work in organization behavior and theory, includ-
ing social psychologists as well as colleagues from numerous disciplines and
specializations.

Some supportive readers see the actual audience as being largely restricted
to the last-identified specialists, with but a few from among OD consultants and
even fewer from among managers. As one reader concludes: “Much of [Ironies]
is simply too sophisticated to be grasped by the practicing OD professional, and
too technical to be of much practical significance or even terribly interesting to
them.”

This position about OD practitioners rests on some daunting realities, of
course. Many ODers got their training on the job. Their initial experiences as
military officers, ministers, or whatever were variously supplemented, but, in
general, their methodological and theoretical concerns have been less cultivated.
Relatedly, the common wisdom has always given preeminence to personal quali-
ties in OD, as in the dictum that ODers’ own warm bodies are their most effective
tool for intervention. That position has merit. However, a daunting reality takes
precedence: that OD practitioners experience a great turnover. For example, 50%
or more of the attendees at the annual meetings of the Organization Development
Network, year in and year out, are first-timers. Early training and socialization
are thus continuously pressing needs, as well as conveniently available.

But such realities are not sufficient to deter this analysis from also targeting
OD practitioners and managers—just the opposite, in fact. Basically, what exists
provides only herniating guidance for closer approaches to some ideal condition.
In addition, three specific points imply that the targeting of *Ironies* II is not simply a kind of misguided willfulness. Like Panasonic, to begin, *Ironies* II seeks to be a bit ahead of its time. The book is oriented toward where OD is going—say, in five to ten years—and pays attention to where OD has been only in the sense of providing a take-off platform. Specifically, this book seeks to meet in future orbit with the sharp increases in theoretically and methodologically aware ODers now being trained for the first time (e.g., Golembiewski, 2000).

In significant addition, various socio-politico-economic authorities have a real interest in *Ironies* II. Consistently, they pursue more with less in many senses. OD provides an approach to conserving people and resources in the pursuit of burgeoning priorities. Those interested in getting more with less will profit from this book.

Is the proposed reach within OD’s grasp? The immediate past implies an affirmative answer. The stakes have been raised greatly in the training of ODers, as in the increase in conspicuous doctorate and numerous master’s degree programs. Finally, what follows often requires only a bit of a stretch to enrich OD and to heighten its already substantial success rates.

**REFERENCES**


Contents

Preface v
Introduction xix

Irony I: Substantial Success but Pessimism About Practice

1. Undercutting the Irony of Ironies with Early and Recent Data: Increasing Confidence About Domestic Success Rates over the Decades 1

2. Further Weakening the Irony of Ironies: Success Rates in Global Settings 49

3. Challenging a Critical Assumption of the Capstone Irony: Putting “Positive Response Bias” into Reasonable Perspective 81

Irony II: Substantial Success with Insufficient Attention to Replication

4. Toward Building Work Cultures to Order: Illustrating Replications of Generic Designs 95

5. Saving Conceptual Shortfalls from Themselves: Enriching New Public Management (NPM) as Exemplar 131

6. Checking Downstream Progress, Years Later: Replication as a Stream of Events over Time 157
# Contents

**Irony III: Substantial Success in Global Applications While Neglecting Alternative Ethics at Work**

7. Responsible Freedom as the Goal in OD, Part I: Some Basic Conceptual Distinctions 183
8. Responsible Freedom as the Goal in OD, Part II: Two Western Work Ethics as the Base 197
9. Responsible Freedom as the Goal in OD, Part III: Confucian Work Ethic as the Base 211

**Irony IV: Substantial Success Without Consensus About a Learning Model**


**Irony V: Substantial Success While Inadequately Assessing Large-System Interventions and Their Effects**

12. Illustrating Large-System Change in Business: Detailing a Design for Strategic Planning and its Effects 281
13. Illustrating Large-System Change in Government: Examining Some Surprises in Labor/Management Cooperation 321
14. Illustrating Large-System Change in Health Care: Reorganizing a Medical-Surgical Ward 347
15. Illustrating Large-System Change at the Interface: Testing Some Features of the Common Wisdom 381

**Irony VI: Substantial Success Without Specifying Contextual Differences**

17. A Big Contributor to Heightened Success Rates, Almost Undoubtedly: OD Designs Improve Group Properties and Reduce Burnout 429
## Contents

### Irony VII: Substantial Success Without Differentiating People

18. Acknowledging Some Limitations of “One Person, One Vote”: Survey/Feedback Realities and Classes of Respondents 449  
19. Highlighting Differences in Personal Slack for Choice and Change, Part I: A Preliminary Profile for Burnout in OD 475  
20. Highlighting Differences in Personal Slack for Choice and Change, Part II: Burnout as Covariant of Many Managerially Relevant Measures, Just About Everywhere 497

### Irony VIII: Substantial Success While Neglecting Easy Pieces

21. Enlarging the Empowering Potential of the Workweek: Flexible Work Hours as Exemplar 511  
22. Enhancing the Empowering Potential of the Concept “Workplace”: Flexi-Place as Exemplar 531  
23. Enhancing the Empowering Potential of the Concept “Development”: Demotion as Exemplar 545

### Irony IX: Substantial Success Without Differentiating Kinds of Change and Designs

25. Really Appreciating Appreciative Inquiry: Extending OD Technology/Values and Success Rates While Preserving the Essentials 591

### Postscript About Multiple Ironies

26. Reducing Ironies and Increasing Success Rates: Tactics and Strategies 605

*Author Index* 651

*Subject Index* 663
Introduction

THE IRONY OF IRONIES IN ESTIMATING OD CONSULTING COMPETENCIES

This book revels in multiple ironies, and perhaps the basic one matches success with failure. To illustrate: Organization Development (OD) theory typically reflects major analytical gaps that need filling, and yet in practice OD success rates seem to be quite high. Indeed, early estimates of those success rates seem to surprise most of the OD literati, not to mention those suspicious of OD or even overtly hostile to its values and approaches.

Other ironies are related to this central one. Thus, two questions express other important moorings of this book, and this pair of queries also provides the basic structure for this Introduction. These questions direct attention to what this book explores, and why:

How can OD have high success rates and yet rely on format theory that is patently fragmentary and incomplete?
If OD has high success rates, why bother much about improving its theoretical base?

WHAT IS OD?

The attention here to ironies is in large part motivated by OD’s values and approaches, which provide reasons why OD works despite theoretical lacunae—why those success rates provide good targets for further upgrading, as well as why people should be motivated to improve an average performance that is already attractive.
Attention here is barely illustrative, yet unavoidable. OD here is seen as a value-loaded enterprise with an associated technology for intervening in organization and process (e.g., Golembiewski, 1979, especially Vol. 1; Golembiewski, 1995). In short, I see my role as an OD intervenor as helping induce greater responsible freedom (see Chapters 7 through 9). The freedom comes in a concerted effort to meet personal needs at work while meeting work demands, and also in empowering employees as well as unfettering modes.

Three basic approaches to such personal need-meeting can be distinguished: interpersonal and group processes, or how people relate and communicate; structural features, or how people are linked in coordinated and intendedly cooperative networks; and policies or procedures, or the rules of the game that encompass and direct both processes and structure in action. Documenting in any depth the need-meeting potential of these three basic OD approaches would distract this analysis, but the range of issues has been raised elsewhere (e.g., Burke, 1982; Golembiewski and Kiepper, 1988; Golembiewski, 1995), and a brief sketch does the job for present purposes. Consider interaction processes that are both unfettering and empowering. I like to think of regenerative and degenerative interaction in such direct terms (Golembiewski, 1979, Vol. 2, esp. pp. 162–175). See Figure 1.

Not only do individuals prefer regenerative interaction, in general, but it also generates consequences that facilitate responsible behavior in organizations. For example, individuals are less burdened with repressed materials, and real issues tend to surface that can be solved without creating greater problems in the process. Degenerative interaction, in contrast, contributes little to either freedom or its responsible manifestations. Thus, important substance or feeling can remain unexpressed and, at an extreme, norms may develop about “not rocking the boat.” Degenerative interaction can lead to organizational mischief via such consequences, even when—and perhaps especially when—everyone is “trying hard.” More fully, degenerative interaction will be characterized by this pattern of progressively more serious effects between people and groups:

Communication and decision-making processes become increasingly burdened
Individuals become less effective at isolating and resolving substantive issues
The amount of unfinished business increases
Individuals feel diminished interpersonal competence and psychological failure; that is, they fail to solve problems such that they remain solved, and without creating other (and often more formidable) problems
Individuals become more dependent and cautious, which can lead to “don’t rock the boat” attitudes and thus reinforce and deepen the tendencies outlined above
Organization norms develop that reinforce closedness, thus exacerbating normal tensions due to misunderstandings, lack of time, or conflicting missions and roles (line versus staff, subunit loyalties, and so on).

In related ways, interventions with structure and policy/procedures can be unfettering and empowering, while also encouraging more responsible behavior—working better and smarter versus harder, giving more of self to work because it is more need-satisfying, and so on. The clearest reflections-in-action of these tendencies exist at the plant level in “high-involvement organizations” (Perkins, Nieva, and Lawler, 1983; Golembiewski and Kiepper, 1988).

**HOW CAN OD SUCCEED AS IT FAILS?**

This much having been said about OD’s good choices in leading theory and values, one can easily encompass substantial success with plenty of room for improvement. One need not look very far for reminders of the several inadequ-
cies and lacunae in OD theory and research. For example, many observers point to the lack of respect in OD for the canons of empirical science: assignments to treatment conditions are seldom made randomly, for example; control or comparison groups are not rare but neither are they commonplace; and long-term studies are the exception rather than the rule.

Put directly, as one perspective on success ↔ failure, the attractiveness of OD values as well as intervenor competencies for approaching them seem to be running ahead of our present ability to formalize what most intervenors seem to be doing correctly most of the time.

Moreover, in the short and even intermediate run, this outrunning of theory by practice is both understandable and even—well—natural. Consider the four perspectives described below consistent with the notion that OD intervenors have basic competencies that somehow provide crosswalks over—or Band-Aids on—analytical and theoretical gaps, or even chasms.

First, many have emphasized that OD practice is ahead of its theory and this is not particularly unusual in a growth area. The urgencies of practice may encourage the discovery of interventions that work, while at the same time discouraging the articulation of growing networks of theory that express in summary form what sensitive OD intervenors have come to know and rely on. The effect can occur in diverse ways, with practical successes paradoxically contributing to the theoretical lag. Specifically, research with a Flexi-Time pilot study produced positive results that accorded with common-sense expectations. Management wished to extend the program throughout the organization, despite a clear lack of knowledge about why the observed effects occurred. Consultants urged a more detailed study of conditions, with only partial success (Golembiewski, Hilles, and Kagno, 1974). Paradoxically, management proposed to extend the good news, immediately and uniformly and by fiat, even though the success of the pilot study rested on deliberately crafted adaptations of the Flexi-Time model to local conditions via participation and involvement. One can even appreciate what underlay their unenthusiasm.

From an important point of view, then, the present collection of ironies constitutes an elaborated form of the notion that OD practice outdistances its theory because its applications have been too successful, as it were, demotivating the careful search for theory. This book thus provides chapter-and-verse illustration of some ways in which short-run success can be long-run poison. And this book also details a number of approaches to reducing the gap between practice and theory before applications too far outdistance the theoretical base and it collapses on itself.

Second, theoretical progress often lags behind practical achievements because of two curious interactions of complexity and convenience. I am reminded of Einstein’s alleged rationale for the theoretical elegance of physics: “Physics is easier than psychology.” Much the same may be said of OD. Electrons cannot
read and hence cannot modify their behavior in response to an experimenter’s findings. In OD’s action research, precisely just such inclusion of the object-of-research as the subject-of-action not only does take place, but intervenors give it a central place.

The example at once establishes the complexity of what OD seeks to do and reinforces the good sense of focusing on basics—on involving broad ranges of participants, and on eliciting their commitment via participation as a workable surrogate for having a comprehensive theoretical model of what happens, when, and why. Indeed, the very power of involvement and commitment has discouraged the development of precise theoretical relationships, or at least has temporarily assigned them a lower priority than making immediate use of what we do know.

This lower priority is not perverse, if we acknowledge that it is only a temporary convenience. That is, the very emphases on participation and involvement can be useful surrogates for fuller knowledge, and even indispensable in approaching that escalating ideal. Efforts to raise the level of current practice will help motivate the search for more comprehensive theory, as well as test the comprehensiveness of various models. In this sense, early insistence on perfection can be a major enemy of excellence.

The efficacy of the basics—participation, involvement, and commitment—may in turn rest on their human rightness, which transcends theoretical detail. It may just be, I grow increasingly convinced, that although people respond in diverse and intricate patterns, those dynamics may well be triggered by relatively direct and even elemental needs. This is the basic rationale for regenerative interaction: that most people require—or at least much prefer—environments characterized by high openness, owning, and trust, along with low interpersonal risk (Golembiewski, 1979, Vol. 1, chaps. 1–3). OD’s power may well derive from tapping such wellsprings of human motivation, and theoretical lacunae may be troublesome but nevertheless pale in comparison with the human correctness of OD fundaments.

Third, it appears to me that OD just might succeed in basic senses in spite of analytical deficiencies, if not exactly because of those deficiencies. Or at least this position is worth reviewing.

What can this third point of view mean? Much of the potency of OD lies in its values, and its aspirations being so right, and so generally desired as well as desirable, that OD’s inherent worth often carries both intervenors and many of their clients over individual and collective inadequacies and shortcomings. In this sense, OD is valuable not only for what it does, but also—perhaps, especially—for the values represented, as they contrast favorably with what exists in many settings.

A variety of evidence supports this third viewpoint. I have been impressed by how many of those who are utterly convinced of OD’s practical inapplicability
nevertheless lament: “You know, it’s a crying shame that we just can’t do those things around here—like being open with each other. It sure would be a different and a better place if we could.” Or consider some hard evidence. Well over 95% of the several thousands of managers and employees we surveyed had no doubt about the kinds of organizational values they prefer. For them, “Ideal” scores on the Likert Profile of Organizational Characteristics (Likert, 1961) typically cluster in the high System III and System IV range—the Consultative and the Participative Group portions of the Likert system of four basic management types. That’s where people want to be even when the realities are very different. “Now” or “Actual” scores typically cluster in System II to low System III. Relatedly, regenerative versus degenerative interaction is much preferred in my experiences with organizational worlds, even in totalitarian systems (e.g., Golembiewski, 2000).

Perhaps the general point can be put in an elemental way. If you understand and accept a set of values, you can tolerate many glitches in attempts to realize those values in practice. Put another way, the person who accepts a why can tolerate almost any how, even halting or fumbling or long-delayed hows. It may even help a bit if people are not too hung up on what “science tells us,” especially if that undercuts involvement, participation, and commitment. This is no plea for the reinventing of veritable wheels, but it does encourage modest proclamations of what we know and when.

Fourth, although a general stampede to humanistic worksites does not seem to be imminent, it also appears that history is on OD’s side for many reasons and guesses reflected in this volume. OD values more clearly and insistently apply in the kind of society and economy we seem to be becoming: more educated and with a growing range of options and personal agendas; with more complex specialties that require integration right now rather than by-and-by; and living within organization structures that often require high degrees of personal, immediate involvement and identification while putting a premium on the ability to let go quickly and with minimal emotional consequences when conditions change, as they almost certainly will.

To be sure, in many senses, OD is the only game in town oriented toward the humanistic meeting of such apparent irreconcilables. But the force of their values and of historical development may keep OD proponents from stumbling over mere deficiencies in personal skills, as well as from running afoul of the analytical gaps in OD theory illustrated below.

In this case, the OD intervenor’s basic capability is being right on the side of history, as opposed to being correct on some matters of theory or technique. Being right about a cosmic essential, it is easy enough to forgive—or at least to put into appropriate perspective—OD’s shortness on particulars such as theoretical or analytical completeness.
WHY TRY TO FIX IT, SINCE IT’S NOT REALLY BROKEN?

So, why not just let OD be, especially given the associated success rates alleged above, and reserved for later detailed demonstration? Since OD does pretty well, despite its several analytical deficiencies, why not leave well enough alone? Indeed, drawing attention to the several ironies below may only reduce the confidence of those who are already doing quite well, thank you. In fact, some OD intervenors have responded to several of the themes below in just this way. Put more directly: “If it ain’t broken, why fix it?”

Basically, in the present view, an OD not reflective about building on and beyond itself is very likely to implode—to fall in on itself in the absence of expanding its reach and grasp, both practically and theoretically. This elemental motivates doing better than average, when average ain’t at all bad.

In addition, several of the enhancements identified below do not require risking professional hernias. This is not true of all the material in this book, but the generalization applies to most. A high benefit:cost ratio urges while it also eases the effort of making many of the analytical accommodations, which this book recommends.

Finally, in this brief catalog of motivations to reduce the gap between practice and theory, OD intervenors are facing—and should feel—definite pressures to become more efficient, even if OD intervenors already appear to be tolerably effective.

This composite point is a weighty one, and developing it in a bit more detail has much to recommend it. OD seems effective, in three senses: its values seem desired and desirable, generally; OD has evolved a family of technologies that permit approaching those attractive values broadly; and a substantial cadre of intervenors exists whose skills and sensitivities seem to extend beyond available theory.

So far, so good. Excellent, in fact.

But efficiency seems to be another matter. For example, individual OD intervenors working with specific teams seem to have substantial success rates in making intended things happen. But 1:1 ratios imply sharp limits on OD’s reach, if only for reasons of cost. Can some reasonable things be done to extend OD’s reach, then, and to reduce the cost per unit? That is, can OD efficiency be increased without jeopardizing OD values?

This book basically argues that not only is an affirmative answer possible but reasonable next bites are already conveniently within reach. In short, we need not invent or discover anything new. We need only to make an act of the will—to apply reasonable and nonheroic skills as we already know.

The motivation for this effort is direct and substantial, because neither efficiency nor effectiveness will be much increased unless substantial attention is given to reducing the analytical deficiencies highlighted below. The conclusion
applies particularly to the leverage inherent in applying the power of useful theories to guide differentiated diagnosis and prescription, especially in massive efforts—for example, for a hundred work teams. There it typically will be awkward or impossible to arrange for intervenor:team ratios of 1:1. Mass applications are a convenient way of reducing unit costs, but they can be dangerous in the absence of a comprehensive theory that can replace in part the personalized adjustments that experienced intervenors apparently make when they work with individual teams.

Finally, OD in a sense has no alternative to increasing its efficiency. OD has a stake in transcending its present limitations—for example, of being available mostly to organization elites, and to small proportions of the workforce, often at substantial cost. Out of enlightened self-interest, then, OD will have to increase its efficiency on several fronts. Various other motives energize such an effort at extension. Thus, pressing needs dominate in many nonelite settings, as in Third World rural settings. Failure to meet such needs may unleash formidable social and political forces, and failure will always imply forfeited opportunities.

REFERENCES
Undercutting the Irony of Ironies with Early and Recent Data
Increasing Confidence About Domestic Success Rates over the Decades

There was a time, and not very long ago, when only a little literature existed about success rates in organizational development (OD). Moreover, not very much attention was paid to that slim literature. Indeed, the first edition of this volume (Golembiewski, 1990) presented the first substantial collection of success rate studies.

Today, this constitutes a better day and time for our purposes, fortunately, and this first chapter will provide multiple fore-and-aft perspective on the developmental status of OD. The focus here, then, is on the success rates available in the literature, over time, in North America. Later attention will go to global settings.

A CONCEPTUAL CONTEXT FOR IRONY I

Irony I deals with both restrained confidence and substantial success in OD, an unlikely combination that this chapter will help to reject. Before dealing with the success rates, however, this section provides a perspective for viewing this irony and related contrasts or even contradictions. For example, some ODers prefer to walk away from OD values even as they encourage great degrees of acceptance and lead to the substantial success rates that motivated the normative extension in the first place. And why? The devalued OD would allegedly be more useful to managers.
In addition, many observers see OD as limited—largely, if not essentially—to interaction-centered designs. Hence, their expectation of limited success. In contrast, the view here is that OD is more properly conceived as a combination of three realms:

Interaction, as via regenerative vs. degenerative interaction (e.g., Chapters 14, 15, 18, 19); Policy and procedures (e.g., Chapters 14, 15, 18, 19, 22); and Structure (e.g., Chapters 13 and 14).

In each case, the unifying force is provided by the values of OD as variously reflected in this volume. Examples in detail appear throughout this book (e.g., in Chapters 14, 15, 19, and 22).

Paramountly, indeed, OD can be considered a $\frac{1}{3}$ matrix, with a specific set of values giving direction and character to the three substantive categories detailed above. Again, the list below indicates the chapters in which normative or value issues get major attention. In sum, OD can be seen as:

Value-loaded interaction (Chapter 12); Greater responsible freedom in broad cultural terms (Chapters 7, 8, and 9); and Normative or value templates that inspire unique policies and procedures (Chapters 21–23).

Penultimately, the first three chapters deal with success rates in three different senses. The word “penultimately” applies, directly, because low success rates would provide little motivation for a comprehensive analysis. But low rates do not characterize the present review, and hence these three chapters legitimize the later comprehensive treatment of a $1 \times 3$ matrix and attempt to show how to avoid the irony of high success rates and also how to apply substantial sources of leverage to heighten those rates as well as to broaden the use of OD/QWL values and techniques.

FOUR DECADES OF EVIDENCE OF OD IMPACT

How to approach the theme of substantial success rates in OD or its close kin such as QWL, or Quality of Working Life? The decade is the unit of analysis, conveniently, and four intervals are distinguished.

Evaluations Through the 1970s

Attempts to precisely date the initial OD evaluations invite contention, so we will take “through the 1970s” as a starting point. For a review of this zesty contention, there is no better place than the best-selling textbook in OD, and espe-
Success Rates in North America

pecially its later editions (French and Bell, 1978, 1994). For general purposes, the
sense of the matter among professionals is aptly—but not accurately, as it will
be shown—captured in Weisbord’s (1977) journal article “How Do You Know
If It Works, If You Don’t Know What It Is?”

Let us finesse this intriguing issue, however, and begin where we can be
quite definite in two regards. First, by the 1970s substantial evidence exists of
efforts to track the success of OD applications, given some mixed features. Sec-
ond, the core of the literature, and especially the growing cadres of ODers at
conventions and in newsletters, did not pay determinative attention to the clear
bias in those accumulating evaluations.

Survey of Earliest Evaluations

By the close of business in 1979, the literature contained major evidence of the
concerns about the validity and reliability of OD approaches, and a clear leaning
as to success. In order of the date of their publication, that literature includes
these titles, among others. The sizes of their panels of studies and major findings
get summary statement:

In 17 applications for broad scope and rigorous methods, 65–70 percent
are rated “effective” (Dunn and Swierczek, 1977);
With N = 38, Pate, Nielsen, and Bacon (1977) do not evaluate success,
but rather describe the studies in terms of eight “categorized dimensions”
and emphasize the need for “more systemic, longitudinal research”;
30 applications are rated as 73 percent “positive,” with 10 percent “mixed,”
24 percent “no change,” and 3 percent as “negative” (Margulies, Wright,
and Scholl, 1977);
One population of 26 cases includes only 8 percent rated as “failures” (Mor-
risson, 1978);
In 35 cases selected for high methodological rigor, about 50 percent of the
changes fall in the expected direction (Porras and Berg, 1978; Porras,
1979); and
An early collection (N = 44) of public-sector applications in urban settings
generates about 70 percent “success” (Golembiewski and Sink, 1979a,b).

These several efforts had some common features. Thus, they variously de-
fine success, but all reported heady levels of change—beginning above 50 percent
and often trending sharply upward. Moreover, populations tended to be small;
methods of comparison were mixed, but tended to be casual; few long-term or
longitudinal studies were involved; units of analysis tended to be small as well at
lower levels of organization; and such surveys typically attracted little attention,
usually being cited only among the informal network of contributors to the sur-
Chapter 1

veys of evaluations. In common, all evaluations were swimming against a tide of doubt or even denial.

Neglect of 1970 Surveys

Why this noteworthy accumulation of wit and will did not dominate early opinion about OD is impossible to say, but several components of such an explanation seem worth the print. Three features get emphasis here.

A Pessimistic Metaphysical Pathos. In the early days, the state of OD did not inspire real optimism or even contention. Debates were held, reputations were risked, and ominous warnings rumbled. But these do not constitute the signs of a new and vigorous arena of thought and applications. At professional meetings, in sum, the politically correct tone was one of self-abnegation, if not the hair shirt.

In short, the literature took on this dour character almost immediately. That literature had an often humorous but usually brow-beaten quality, suffused with doubt if not self-denigration. Thus, one popular professional piece spoke of “The Wizard of OD” (Weisbord, 1978a), with the text suggesting for this reader variants of “OD”—Oz, Id, Odd, and so on. Other views among professionals (e.g., Weisbord and Goodstein 1978; Weisbord 1978b) left some room for eventual OD triumph, but not much. Then, again, the Academy of Management hosted a debate on the theme: “The OD Intervenor as Nebbish or Superman?” The advocate for Superman—it happened to be this author—had to swim against strong opinio nal currents.

Perhaps it is best to interpret this tone as an in-betweeness (e.g., as beyond innocence but short of comprehensive theory and practical guidelines). And all the while, the OD growth rate was upward-sloping, and sharply so.

In any case, by the mid-1970s, a curious agreement developed. Friend and foe alike tended to have real doubts about OD’s future. Critics pointed to a range of problems—theoretical, methodological, and ethical (e.g., Woodworth, Meyer, and Smallwood, 1980). Even historical supporters saw OD as at a critical life stage—as an adolescent, with quite definite signs of lacking those qualities associated with “most likely to succeed” (e.g., Friedlander, 1976, esp. p. 7). Other supporters saw a kind of academic and applied hardening of the early arteries, with the memories of initial hopes still alive but with a growing sense that the heydays were all but certainly over (Burke, 1976, p. 24). I also have in mind an academic symposium of OD aficionados, who had for several hours zestfully played “can you top this” with pronouncements concerning the deficiencies and all but inevitable doom facing OD—poor research, inadequate underlying theory, and so on. One practitioner had more than enough. “You people give me a headache,” he noted in exasperation. “I know I do good work, but you guys have made galloping variables out of all my constants.”
Effects of Certain Big Ticket Applications. Perhaps very influential in this feeling tone about OD in circles acquainted with the literature were certain major, or big ticket, applications. From the earliest days, for example, one large-system application in a factory had a range of consequences on opinion about success rates, dominantly positive (e.g., Marrow, Bowers, and Seashore, 1967). For close readers, in addition, this application had major effects that seem to have persisted over extended periods of time—even an extraordinary interval (e.g., Seashore and Bowers, 1970).

However, other big-ticket items seem to have been more consequential and in an opposite direction. Thus, despite serious questions with its method and technique (e.g., Pasmore, 1976; Torbert, 1975), one collection of evaluative studies had powerful repercussions among those who took their literature seriously, and especially those who valued the source of evaluation. This Michigan ICL survey (Bowers, 1973a, 1973b) had a major and dour influence, although even its major craftsman clearly acknowledged its methodological limits (Bowers, 1973b). To a similar effect, but with less public fanfare among organizational scientists, the mixed record in Project ACORD—or Action for Organization Development—in the U.S. Department of State also encouraged caution. Desirable effects were reported (e.g., Marrow, 1974), but the casual reader was likely to have reinforced the opinion that public-sector applications posed formidable problems for OD. Later and more-nuanced evaluations (e.g., Marrow, Bowers, and Seashore, 1967; Golembiewski and Kiepper, 1988) seem to have had less influence, except perhaps among close followers of the literature on change.

Finally, and especially for those unlikely to delve into the detailed periodical literature, efforts more or less in the critical tradition deserve note. For example, a collection of OD “failures” appeared (Mirvis and Berg, 1977). Few OD sources used the technique of interest here, but nonetheless the cases were troubling ones for those observers following OD developments, and especially at some distance.

Applications in Bureaucracies, Especially Public Ones. In general, opinion indicated that OD faced special challenges in the case of advanced bureaucratic forms, especially in the public sector. For example, Burke (1980, p. 429) makes the general case, forcefully; and Giblin (1976) adds detailed counterpoint from a single application in an eastern U.S. state.

Two apparent hard facts made this accusation even darker. That is, most observers assume, only a tiny fraction of the assumedly small population of OD applications anywhere involved public-sector applications, which are widely regarded as posing unusually difficult problems for OD and as having low success rates. Perhaps 10–15 percent is the usual proportion of public applications in available databases, and one-tenth of a small database does not provide a very solid foundation for generalizations.
Chapter 1

The OD outlook did not seem bright, then, in the 1970s. Some of the criticisms seem credible; some were ephemeral; but only a few gave substantial weight to the substantial set of surveys of applications detailed above.

Evaluations in the 1980s

A sea change began in the 1980s, given special impetus by two major reinforcemnts for a more positive view of OD prospects coming from persuasive angles. One reinforcer provided a massive population of OD applications, with a large representation of public-sector cases; and the second focused on a separate study of “hard” or “objective” measures of OD outcomes. Attention turns to each—the first at length and the second only briefly. In sum, the first survey raised the ante for size and the second upped the criteria for success, in the panels of evaluative studies. Both inclined toward a generally positive evaluation of OD efforts.

Raising the Ante for an OD Survey, Sharply

The first strategic initiative involved a thorough search for North American OD applications, seeking to remedy one major weakness of the several early surveys of OD effects. As noted, the databases for early summary studies tend to be small, on the order of scores of cases. For example, the study by Porras deals with thirty-five cases, and Morrison’s (1978) methodological overview involves twenty-six cases. In addition, only a small fraction of such databases deal with public-sector applications, which are widely regarded as posing unusually difficult problems for OD and, hence, as having especially low success rates.

Specifically, the earliest raising of the ante for surveys came in 1981 and 1982 publications. In that interval, two assemblages of OD applications generated databases of OD applications that could support useful conclusions about effects (Proehl 1980; Proehl, Golembiewski, and Sink, 1981; Nicholas, 1982).

The first survey had multiple attractions. Its batch of OD applications was isolated via:

Seven specialized bibliographies;
Searches of the several relevant computerized listings were isolated (via, e.g., ERIC) from publications in social-science journals over the prior twenty years;
A review of the prior twenty years of studies reported in eighty-eight journals, including ten from overseas, but published in English;
More than one hundred books surveyed for bibliographic items as well as for reports of interventions; and
Personal letters sent to about fifty well-known change agents, especially soliciting unpublished materials such as internal memos, dissertations or theses, and so on.
Appropriate citations were found as early as 1945, and the search extended into early 1980, when the search was closed to analyze data.

This search process has two gaps, neither of which is seen as damning. First, journals unavailable in English were searched only selectively. Even though 17 percent of the total batch of interventions come from non-U.S. settings, this leaves our database with a dominant Western locus.

In addition, the search did not encompass the twice-yearly meetings of the Organization Development Network or of the OD Division of the American Society for Training and Development (ASTD). Until recently, neither interest group published proceedings. Many interventions reported at these meetings got into our database, however, either after being published or because reports about them were forwarded by about fifty personal contacts.

Through 1980, these two gaps notwithstanding, our search uncovered a substantial number of OD applications—574 cases, to be exact (Proehl, 1980). In sum, this raises the ante for the size of survey populations, obviously and substantially so.

Two Claims about Data Set. We make only two claims about this set of OD applications. First, there seems almost no question that public sector applications get adequate representation. Indeed, the very number of such applications (270) itself constitutes a major finding, since most sources emphasize the paucity of public sector applications (for an exception, see Miller, 1979). Public-sector cases constitute over 47 percent of the present batch of OD studies, which is more-than-proportionate representation, given the proportion of public employees in the total workforce.

Second, we propose—a little more tentatively—that the 574 cases provide a reasonable replica of all OD activity. Early published work may have some bias toward successful applications, but we also include a range of unpublished sources. Moreover, the thirty-five-year collection period, plus the large number of cases, should substantially compensate for any early but artifactual hopes or biases.

All in all, then, this analysis can confidently propose that the 574 cases provide a credible source for seeking answers to two major questions concerning OD: What is the range and diversity of interventions or applications? and What is the probability that an intervention was successful?

Range and Diversity. The range of the 574 interventions is broad, with major representation from all the major classes of interventions associated with OD. Let us build toward this conclusion by providing early details, with Chapters 7–9 especially providing more specifics. In introductory summary, most observers see OD as one of the major derivatives of the “laboratory approach”—a major way of learning to learn. Overall, OD has at least six distinguishing features (Blumberg and Golembiewski, 1976, esp. pp. 22–35):
Rootedness in a definite set of values, which emphasizes openness, trust, and collaborative effort;
Seeking to simultaneously meet individual needs as well as those of several levels of larger systems—small groups, large organizations, and so on;
Grounding in immediate experiences as they occur, this often gets expressed as a here-and-now orientation and is reflected in “process analysis” of the panoply of personal and institutional forces acting on individuals and groups;
Emphasis on feelings and emotions, as well as ideas and concepts;
Preeminence of the individual’s involvement and participation—as subject and object, as generator of data as well as responder to those data—in an action-research sense; and
Heavy reliance on group contexts for choice and change, to validate data, to develop and enforce norms, and to provide emotional support and identification.

Beginning in the 1960s, more or less, OD became the major extension of early work with the laboratory approach. It commonly came to encompass not only interaction-centered designs but, if more slowly, also those designs focusing on structure as well as policies/procedures. Basically, OD builds on the core values and central dynamics of the laboratory approach into several classes of learning designs appropriate for choice or change in large aggregates of people and systems.

Each OD application will be unique to an extent and, typically, will combine several basic designs. As a first cut, however, these alternative designs can be classified in terms of eight sub-classes of activities. The classes are listed here, roughly in order of their complexity and subtlety.

**Process-analysis activities**, or applications of behavioral science perspectives to understand complex and dynamic situations. These perspectives can be simple; for example, as in routine retrospection among task-group members who ask, “How do we feel about what we just did?” The perspectives also can be complex, as in seeking to understand interpersonal conflict as an expression of differing personal predispositions, or as in designing and interpreting opinion surveys.

**Skill-building activities**, or various designs for gaining facility with behaviors and attitudes consistent with OD values—as in giving and receiving feedback, listening, resolving conflict, and so on.

**Diagnostic activities**, often with feedback, which often include process analysis, but which also may employ interviews, psychological instruments, or opinion surveys to generate data from members of some social system, for their use in “action research” activities. These data get fed back into that system, to serve as the raw material for action-research
sequences: diagnosis, prescription of changes, implementation, and evaluation.

*Coaching/counseling activities*, which seek to apply OD values in intimate situations, as between a pair-in-conflict in an organization via “third-party consultation.”

*Team-building activities*, or efforts to increase the efficiency and effectiveness of intact task groups. Variants may use T-group or sensitivity training modes, as well as one or more of the other activities listed here.

*Intergroup activities*, which seek to build effective and satisfying linkages between two or more task groups, such as departments in a large organization.

*Technostructural or sociotechnical activities*, which seek to build need-satisfying roles, jobs, and structures. Typically, these activities rest on a “growth psychology,” such as that of Maslow, Argyris, or Herzberg. These structural or policy approaches—job enlargement, Flexi-Time, and so on—often are coupled with other OD activities.

*System-building or system-renewal activities*, which seek comprehensive changes in a large organization’s climate and values, using complex combinations of the seven activities sketched above, and having time spans in the three-to-five-year range.

These eight classes of activities fit with varying precision into three basic OD modes: interaction-centered, structure, and policies or procedures. Process-analysis, skill-building, and coaching/counseling activities are basically interaction-centered. Technostructural and system-building activities emphasize structure, although not to the exclusion of the other two modes. Team-building and intergroup activities often have dominant interaction emphases, but also deal with structure and especially policies or procedures. Later chapters provide examples of all eight types, often in combinations. For example, Chapter 14 may be labeled “technostructural” or “sociotechnical” and gives attention to all three classes of activities in a comprehensive redesign of a hospital setting.

We can be more specific on the general point, as in describing the distribution of our 574 cases among these eight classes of activities. Table 1.1 implies that our population covers the field of interventions. The most narrow designs—diagnostic activities and process-analysis activities—constitute the dominant intervention mode in less than 5 percent of the cases. OD interventions tend to hunt bigger game, in short. To illustrate, nearly 40 percent of the private-sector cases can be categorized as emphasizing the most complex intervention modes—system-building or system-renewal and technostructural activities. Reading the individual case reports in the public sector also reinforces this impression. The applications there seem to give substantial attention to the tough cases, on balance. Hence, the common emphasis is on racial tension; conflict between individ-
TABLE 1.1 Incidence of Eight Classes of OD Activities in Public and Private Sectors

<table>
<thead>
<tr>
<th>Classes of OD Design</th>
<th>Public sector</th>
<th>Private sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual applications classified by dominant design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process analysis activities</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Skill-building activities</td>
<td>65</td>
<td>57</td>
</tr>
<tr>
<td>Diagnostic activities</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Coaching/counseling activities</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>Team-building activities</td>
<td>51</td>
<td>56</td>
</tr>
<tr>
<td>Intergroup activities</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>System-building or system-renewal activities</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>Technostructural or sociotechnical activities</td>
<td>44</td>
<td>84</td>
</tr>
<tr>
<td>Totals</td>
<td>270</td>
<td>304</td>
</tr>
<tr>
<td>%</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Intervene where the pain is felt!

In sum, the 574 cases do not constitute a collection of easy pieces. Significantly, then, it is of note that the data suggest no large differences between the reliance on specific modes of OD interventions in public and business sectors.

The classification of the 574 published OD reports by dominant mode of intervention has a high reliability. Two independent observers classified all cases and had a very high degree of agreement. A 10 percent sample (approximately) places that agreement at nearly 98 percent of the cases. All differences were reconciled before summation either in Table 1.1 or following tables.

Two Estimates of Success. How can we estimate the specific efficacy of OD interventions? Two basic estimates are available in the present approach, and they distinguish “global indicators” and “multiple indicators.”

GLOBAL INDICATORS. A few details provide perspective on the present global evaluation of OD interventions. Two independent readers reviewed each of the 574 interventions and assigned each set of effects to one of four categories whose content the observers had discussed and illustrated in detail. The evaluative categories are:

Highly positive and intended effects on the efficacy and effectiveness of some aspects of a discrete system, as in improving the ability of individu-
als to hear one another without distortion, or in reducing the degree of hostility between conflicting actors or units;

*Definite balance of positive and intended effects*, defined in terms of mixed but generally favorable effects—e.g., most but not all intended effects are achieved on a number of variables, or major positive effects occur in one system, while some negative but not counterbalancing effects occur in another system;

*No appreciable effects*;

*Negative effects*, or a case in which there are substantial reductions in the efficiency and effectiveness of some subsystem or of some border system of which it is a part.

What did this laborious rating and cross-checking reveal? Four points summarize the major findings. First, by and large, the observers saw the same effects. Specifically, the observers’ ratings correlated 0.78, which indicates substantial agreement between raters. Almost all cases of disagreements involve the first two rating categories. All differences were reconciled after this reliability check but, in all cases, the ratings in Table 1.2 reflect the lowest contending score.

Second, given this conservative convention, one can then conclude with some confidence that in this population of studies huge majorities of the interventions had at least a definite balance of positive and intended effects, as raters judged them. See Tables 1.2 and 1.3.

Third, global estimates of the efficacy of OD interventions do not vary much between the public and business sectors. Table 1.2 supports this major point.

**Table 1.2** Global Estimates of the Success of 574 OD Applications

<table>
<thead>
<tr>
<th>Rating category</th>
<th>Public sector</th>
<th>Private sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Highly positive and intended effects</td>
<td>110</td>
<td>41</td>
</tr>
<tr>
<td>Definite balance of positive and intended effects</td>
<td>116</td>
<td>43</td>
</tr>
<tr>
<td>No appreciable effect</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Negative effects</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>Totals</td>
<td>270</td>
<td>100</td>
</tr>
</tbody>
</table>
TABLE 1.3  Global Estimates of Efficacy, Private Sector Cases

<table>
<thead>
<tr>
<th>Class of OD design</th>
<th>Highly positive and intended effects</th>
<th>Definite balance of positive and intended effects</th>
<th>No appreciable effect</th>
<th>Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process analysis activities</td>
<td>16.7%</td>
<td>50.0%</td>
<td>16.7%</td>
<td>16.7%</td>
</tr>
<tr>
<td>Skill-building activities</td>
<td>40.4%</td>
<td>52.6%</td>
<td>3.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Diagnostic activities</td>
<td>33.3%</td>
<td>44.4%</td>
<td>5.6%</td>
<td>16.7%</td>
</tr>
<tr>
<td>Coaching/ counseling activities</td>
<td>40.0%</td>
<td>46.7%</td>
<td>6.7%</td>
<td>6.7%</td>
</tr>
<tr>
<td>Team-building activities</td>
<td>39.3%</td>
<td>51.8%</td>
<td>3.6%</td>
<td>5.4%</td>
</tr>
<tr>
<td>Intergroup activities</td>
<td>44.4%</td>
<td>39.0%</td>
<td>5.6%</td>
<td>11.1%</td>
</tr>
<tr>
<td>System-building or system-renewal activities</td>
<td>45.7%</td>
<td>40.0%</td>
<td>5.7%</td>
<td>8.7%</td>
</tr>
<tr>
<td>Technostructural activities</td>
<td>40.5%</td>
<td>51.2%</td>
<td>3.6%</td>
<td>4.8%</td>
</tr>
</tbody>
</table>

Notes: Estimated effects in percent. Due to rounding, totals may accumulate to 100%. N = 304.

Fourth, global estimates of success vary somewhat by dominant mode of intervention. Table 1.3 summarizes the experience for private-sector interventions, which do not differ markedly from public-sector experience. Except for two classes of OD activities—process analysis and diagnostic—the efficacy estimates indicate that 83 percent or more cases generate at least a definite balance of positive and intended effects.

**Multiple indicators.** A second approach to estimating the efficacy of OD interventions relies on multiple indicators—308 variables, in fact. Proehl (1980) coded each of the 574 cases in the present batch of studies in terms of the comprehensive set of indicators developed by Porras and Berg (1978). Proehl describes his procedures in these terms:

Each of the . . . studies in this research’s data-base was searched for the 308 variables developed by Porras and Berg. When one of the variables was found, it was coded according to whether it had improved (1) or not improved (0) during the course of the change project. Once all of the variables present in each study were identified and coded, the
“percentage of positive reported change” was calculated for each level (individual, leader, group or organization) of study. This was accomplished by dividing the number of positive variables by the total number of variables in which change was desired in each organizational level of each study. For example, a change effort which sought to change five individual-level variables and reported three of them having changed positively was given a score of 60 percent. Scores ranged from zero percent in a change effort, which failed to produce any positive change in process and outcome variables, to 100 percent for a case in which positive change was reported on all variables for which change was desired (1980, p. 58).

The positive change estimate is 70.5, overall, when 574 cases are scored for all of the 308 Porras/Berg variables applicable in each case. The efficacy of the 574 applications also can be arrayed according to levels of analysis, and the percentages of positive reported change are

- Individual: 78.1 percent for 243 cases;
- Leader: 68.1 percent for 173 cases;
- Group: 77.9 percent for 161 cases;
- Organization: 72.4 percent for 206 cases.

As the best informed estimate from the standpoint of multiple indicators, then, at least seven of ten variables show that OD applications induce a balance of positive effects. In addition, no major differences characterize public versus business applications.

By an ample margin, these 1980 results confirm a substantial success rate for a large batch of OD interventions. To be conservative, the two estimating approaches imply a success rate of at least seven in ten cases. More ebullient observers can rightly emphasize the global estimate of efficacy, which approximates 85 percent.

These data powerfully imply that both critics and previously pessimistic supporters of OD must sing a different tune in the future, or at least a more complicated one. That tune below will have a single theme, despite numerous variations: existing OD success rates are already substantial, and it will be shown in various ways that this overall record can be improved upon in important particulars by fine-tuning OD designs and approaches.

Raising the Criteria for “Success”

In significant addition, the early 1980s also saw the publication of another significant evaluative survey of OD effects. The panel contained 65 carefully selected studies, which tightened criteria for estimating “success” (Nicholas, 1982).
Directly, only “objective” or “hard” measures were targeted and, while this reduced population size, the survey focused on a class of variables of direct interest to scientists, and perhaps especially to management. Nicholas used three major categories to encompass the OD designs in his panel of applications:

- **Human-processed approaches**
  - structured laboratory training, etc.
  - team building
  - survey/feedback
- **Technostructural approaches**
  - job design
  - sociotechnical system, etc.
- **Multifaceted approaches**
  - several designs from the categories above

Illustratively, Nicholas focuses on a range of variables: Work Force (e.g., turnover); Monetary, (e.g., costs); Productivity (e.g., efficiency); and Quality.

Table 1.4 summarizes Nicholas’ findings, and they clearly do not require any basic adjustments of the expectations of effects based on the summary of the numerous populations surveyed in this chapter and the two following.

One special characteristic of Nicholas’ work encourages that it gets a premium of attention. If only for the sake of completeness, indeed, one may reasonably propose that Nicholas’ data represent conservative estimates. Consider only three factors in such a case. First, Nicholas intended to select cases that would provide the strictest test of effects, and hence his estimates of success should be the most demanding and the apparently least successful. Second, Nicholas includes cases that are arguably not OD; for example, he includes “job enrichment without participation,” which Nicholas finds has very low success estimates. This is not surprising; indeed, this author would not even include “job enrichment without participation” in his list of OD applications. Third, Nicholas’ basic criterion of change is more limited than the conventions used in most of the studies summarized in Chapters 1–3. Thus, some of those evaluations in the section include as “successes” large changes in target variables in studies that did not perform statistical analysis. They would not be included in Nicholas’ summary because of the lack of statistical analysis.

**Growing Attention to 1980 OD Evaluations**

The two major studies reviewed above constituted a kind of double-play in OD, to rely on baseball terminology. They deprived critics of two major points of argument. First, no longer could critics argue that few evaluative studies existed; and second, no longer could one argue the point that any positive OD effects might only reflect the weakness of operational definitions of “success.” Virtually,
<table>
<thead>
<tr>
<th></th>
<th>Human processed approaches</th>
<th>Technostructural approaches</th>
<th>Multifaceted approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of studies</td>
<td>29</td>
<td>56</td>
<td>14</td>
</tr>
<tr>
<td>Total % of targets achieving statistical significance</td>
<td>54%</td>
<td>53%</td>
<td>47%</td>
</tr>
</tbody>
</table>

all earlier surveys utilized self-reports or “soft” estimates of effects—changes in reported satisfaction, in cooperation, and so on.

This growing attention to OD’s success rates was perhaps most dramatically illustrated by the intrusion of one of the two major surveys of evaluations—that of Golembiewski and his colleagues—in the major book of readings about OD (French, Bell, and Zawacki, 1989; 1994; and 2000). This inclusion suggests a notable degree of legitimization in what is undoubtedly one of the leading teaching resources in OD.

Such growing attention to “success rates” by conventional ODers was balanced by a significant development, which in effect assumed substantial OD success but also may have tended to create problems for OD applications. OD success, that is, led to the broader adoption of OD techniques and (sometimes) values by a range of those not in the original parade of OD enthusiasts. This might be labeled adoption by popularity rather than adoption by commitment.

In such cases, the parties seek advantage by association. They seek to be included in the popular parade, but while making few or no changes in what they do, or how, or why. Typically, as in the case of OD approached via “organization culture,” the new adoptees brought with them the methods and concerns familiar to the training and professional training of the new adoptees (e.g., Golembiewski, 1987). In the case of “culture,” that is to say, the home grounds of many new adoptees of OD were sociology or anthropology, with a sprinkling from other social-variants—social psychology, social geography, and so on. OD values, particularly, might be finessed by the new adoptees.

Evaluations of the 1990s and Beyond

The immediately prior decade gave useful attention to another source of vulnerability of the OD literature on “success”—that the literature was the creation of a few enthusiasts, who fed on one another. This section, in effect, cuts the ground from beneath such a defensive approach.

Despite the inroads of the evaluative literature of the 1980s, aspects of a dour view of OD survived and do so unto our day. Periodically, the literature breaks out with a kind of scholarly zit—an argument proposing that the prospects of organization change are so sorry that doubt and cynicism constitute the only responses (e.g., Burke, 1980; Kanter and Mirvis, 1989). And so it recently came to be on the pages of Executive, which not only detailed sources of that doubt and cynicism but also suggested ways and means for moving beyond the derivative “negative consequences for the commitment, satisfaction and motivation for employees” (Reichers, Wanous, and Austin, 1997, p. 48).

This section seeks to show that the success rates above have been verified by many observers using many batches of applications. In the new millennium,
the loci for such conclusions about OD appears even in prestigious sources like the *Harvard Business Review* (e.g., Beer and Nohria, 2000, esp. p. 133).

**An Overview Through 1989: Why Slow Recognition of Success Rates in OD?**

This basic condition—increasingly robust success rates and their underappreciation—seems a curious state of affairs and may be beyond anyone’s powers to explain, but let’s give it a try, even though clearly conjectural. This author might be wiser to take a minimalist strategy, to be sure. Thus, he could always take refuge in the proposition that OD does work or it would not be widely used. But let’s take on the broader argument.

This version of a good try has three emphases. High success rates have a negative aspect—a thorn among the roses, as it were. Such rates heighten expectations, and hence imply greater psychological loss when failure occurs. If that is our problem in OD, well, we will just have to learn how to live while raising the ante.

Beyond this possibility, let us focus on a single, if involved, point: why both OD optimists and pessimists might have given the silent treatment to the success-rate literature. The line of argument below is a bit involved and is speculative, in the bargain. But the effort needs to be made. For supporting details, see Golembiewski (1986).

Obviously, the reaction to the success-rate literature might well derive from a reasonable caution that the results are spurious or somehow artifactual. Observers could have come to this conclusion in two different ways, however. Let us distinguish two classes of ODers responding to the growing literature on success rates: those who had an interest in the matter, but whose minds were still open as to the results of then-unavailable aggregate analysis of success/failures in individual studies; and those who not only had made up their minds on the matter, but who also reflected a public pessimism about success rates in the absence of comprehensive evidence. This latter camp all but dominated the literature and conferences, taking two strong (if curious for ODers) positions that organizational success rates are modest, and particularly so in the public service.

Parenthetically, one wonders why the pessimists kept doing what they did—that is, OD. Most of the pessimism, without doubt, came from within the OD tent. Given really low success rates, the obvious conclusion should have been: Back to the drawing board!

But I digress. To move on, consider a second possibility. Directly, ODers who had come to pessimistic (and public) conclusions about low success rates, and especially in the public sector, seem to have an easier case in some senses. But that case nevertheless poses some subtle issues.

How might one behave who had earlier taken a position—often and rock-solid—that seemed to be substantially off-base, to judge from later evaluative
studies? Silence is a reasonable enough reaction, and for several possible motives that might be working alone or in various combinations. Hence, silence might be motivated by watchful waiting for contradictory evidence that others might supply. Or silence might simply be the temporary outward sign of an industriousness devoted to developing a careful critique of the evaluation studies, or to generating fresh data supporting the contrary point. Or silence for the public pessimists might simply reflect a reasonable hunkering-down—a kind of “cover your ass” (CYA) by appearing to look elsewhere, hard. During this period of active neglect, the mass of ODers at least would be less likely to have the evaluative literature impressed on their minds by vigorous debate. Alternatively or complementarily, the underlying hope of silence might be that the world would soon enough forget who had made which pessimistic assessments about success rates.

The pessimistic camp also encourages subtler analysis. Consider here only that nobody likes to have their baby called ugly, and any concerted attention to the emerging success-rate literature would perforce call attention to the uncomeliness of someone’s baby. Silence is a reasonable way to avoid unpleasantness, especially for that large proportion of ODers who are far more intervenors than contributors to the literature. In sum, such ODers would not have a dog in the fight, so why draw attention to it.

A third possibility in the pessimistic view of OD involves a major conceptual wrinkle: that the OD record was worst when applications occurred in large bureaucracies and especially those in the public sector. Thus, a well-known ODer (Burke, 1980) developed the major proposition that OD’s record of change in large bureaucracies was a sorry one, with the exception of several of my publications. Burke notes:

Robert Golembiewski is one of the most experienced consultants in the field of OD with bureaucracies, especially in the public sector. He tends to be optimistic about such consultation. . . . To the extent that he is successful as a consultant, he may be an exception to my remarks about . . . OD in the public sector. Most OD consultants find working with bureaucracies, especially public ones, to be difficult at best (Burke 1980, 428–29).

Note that Burke did not yet have available to him some of the various early success rate evaluations referred to in this chapter when he wrote these observations—probably in 1979 and perhaps even earlier.

The common extension of this position created a kind of cul-de-sac for OD pessimists in relation to the emerging success-rate literature. To put the implied point in its boldest terms, the substantial success rates left only uncomfortable conceptual territory for the pessimists. Their once-dominant view could now be explained as having its roots in the somehow peculiarly difficult situations that pessimists encoun-
Success Rates in North America

pterated in their practice. More directly still, this view opens the possibility that the problem for the pessimists was not in OD but in their particular approaches to it.

Both conceptual possibilities might well encourage silence, for at least two reasons. Thus, it would be difficult to establish that—or to understand how and why—the pessimists had usually selected difficult cases for analysis and hence had unfavorable experiences, on the whole. Moreover, precious little motivation would exist among pessimists to embrace the second possibility; this nettle discourages grasping it, let alone cleaving to it.

The Post-1990s Era

By now, in any case—and beginning with the 1990s—the evaluation of domestic OD reflected a growing appreciation of what the success-rate literature had to say. Those success rates, variously viewed and measured, were substantial, even robust.

Further boosts to OD came from several important sources, which get summary treatment below in terms of four themes. Thus, Chapter 2 takes the presentation to a global level. Immediately below, this chapter concludes with three additional perspectives that reflect 1990s inputs. These include, in order:

several propositions about OD and its close kin—Quality of Working Life, or QWL;
additional literature in the 1990s about OD and QWL success rates; and
some concluding comments.

Beyond 1990: Planned Change in OD and QWL

So what do the 1990s and beyond add to what we are learning about planned change of the OD and QWL varieties? OD and QWL vary in a number of specifics (e.g., Skelley 1989), but they tend to differ mostly in terms of the insignificant issues. Primarily, both OD and QWL are substantially congruent in the values they represent.

Several Propositions About Planned Change. Considering both OD and QWL, six propositions at once highlight major sources of doubt and even cynicism about organizational change, as well as summarize post-1990 progress on them. Brief rationales for each of the components of the common understanding are included.

SUCCESS RATES ARE MODEST, AT BEST. The organizational literature tends to be characterized by purplish prose in this regard, but estimates tend to cluster in the poor-to-no-change range. The post-1990s literature on success rates in part undercuts this dour position, and the later research goes further still, as the two sections below and Chapter 2 establish.
There was a long way to go in reversing the clear flood tide. For example, Kaufman (1971, 1995) represents the more dour observers: “... many organization members discover that nothing seriously interferes with their impulse to go on doing just what they were doing previously” (1971, p. 76). The earlier Kaufman (1971, pp. 41–67) acknowledges that organizations do in fact change, but he concludes that, on balance, these barriers usually overwhelm even determined resolve to depart from the status quo. Why is change so frequently dampened? Kaufman emphasizes three major classes of barriers (Kaufman, 1995, pp. 68–91): limitations that are rooted in the organizational aversion to unpredictable events; limitations generated by the relationships characterizing organizations, and limitations inherent in the environments in which organizations exist.

Revealingly, the later Kaufman (1985) surpasses even his earlier self as he espoused a kind of “natural selection” perspective in which the very idea of planned change became presumptuous, if not a delusion. To put it in a conceptual nutshell, but not at all unfairly, organizational change in Kaufman’s later view is essentially random. He fixates on the role of luck or chance, with some interesting (even curious) wrinkles that so far have all but escaped detailed inquiry (e.g., Golembiewski, 1995, pp. 144–167).

SUCCESS RATES ARE ESPECIALLY LOW IN PUBLIC SETTINGS, AND APPLICATIONS LESS COMMON. Quite commonly, observers still see public sector organizations as notably recalcitrant to change or learning, for numerous reasons. In recent days, the usual supporting rationale argues that public agencies typically lack both a competitive market as well as a “bottom line,” or estimate of profit, despite the many reservations that seem appropriate concerning the capacities of “markets” (e.g., Conrad, 1997). From a constitutional perspective, many commentators (e.g., Goodsell, 1994) propose that where separation of powers dominates, the integrative thrust of OD and QWL will be deliberately limited, if not positively inappropriate. A few observers leave open a bit of a margin for doubts and differences, but not much (e.g., Burke, 1980).

In general, the pessimistic view has substantially receded from its high water mark, based on the success-rate literature reviewed above. Additional research concluding this first chapter and extending through Chapter 3 add to the force of this counterargument! On balance, the viewpoint on public sector recalcitrance to definitely planned change has moderated in recent years (e.g., Rainey 1991, pp. 223–249), but the bias still exists.

SUCCESS RATES ARE LOW IN UNIONIZED SETTINGS, AND APPLICATIONS ARE RARE. Two sub-rationales apply here, one for OD and a substantially different one for QWL. For OD, the view builds on the fact that the usual target populations are managerial or executive. To stretch that reality, applications in unionized settings are not only limited but, given the implied OD design bias toward the nonorganized, optimism is not appropriate concerning success rates. Some observers even propose that OD activities are deliberately oriented to inhibiting,
replacing, or even busting labor unions. Illustratively, some court cases try to draw boundaries for OD activities that avoid the charge that they are inherently unfair labor practices (e.g., Electromation, Inc., and International Brotherhood of Teamsters, 309 NLBR No. 163).

For QWL, the issue concerning unionization has been muted. What we know is that QWL activities are oriented toward operating levels, which are more likely to be organized. Relatedly, QWL applications often occur in European—better still, Scandinavian—sites, and QWL ideology emphasizes direct associations with political policies and practices. To the degree that OD is “North American,” this suggests a poor match with QWL activities.

Overall, then, major factors encourage low estimates for success in North American OD, and few applications involving unions exist in the early evaluative literature. Chapters 2 and 3 contain useful detail about the post-1990 situation, which is far more hopeful. Again, this encourages low estimates for success in North American OD.

SUCCESS RATES ARE HIGHEST FOR APPLICATIONS LACKING METHODOLOGICAL RIGOR. This fourth proposition essentially argues that “success” in planned change often in reality is more properly “sloppiness.” In this variant of “all that glitters is not gold,” all or most “positive consequences” of planned change are subject to serious doubt, if they are not baldly artifactual. That effect is said to become progressively more likely as threats to validity and reliability are removed from research designs and measuring instruments (e.g., Campbell and Stanley, 1963; Golembiewski and Sun, 1990a); OD or QWL effects, in this view, will tend to disappear as rigorous research becomes more common.

This general proposition about a “positive response bias” has been a common one in the behavioral sciences and has been no less influential in the literatures on planned change. For example, Morrison (1978) provides an early, and well-done, critique of a small population of OD studies because they lack methodological rigor. She isolates several classes of rigor, with only a minority of evaluations scoring high on their care and completeness in reducing threats to validity.

New evidence presented in Chapter 3, however, puts this concern in a more realistic perspective. In sum, one could go broke quickly in betting that a “positive response bias” explains the success rates reflected above this point, and later in this book.

SUCCESS RATES ARE HIGHER ON SELF-REPORTS OR SUBJECTIVE EVALUATIONS. This next-to-last proposition, for present purposes, makes a compound argument: that evaluations of planned change typically use self-reports to measure effects; and that such “soft” variables are basically inferior to “hard” variables for estimating effects.

Obviously, this fifth proposition is one of the species dealing with degree of methodological rigor. Self-report data present problems of both validity and reliability, and are often consequently said to be punier than “objective” evalua-
tions. Thus, “the dollar cost of materials” used in some OD applications is a hard or objective measure. Illustratively, the awkward kind of soft or self-report data will be typified by this spurious kind of attribution; for example, allegedly well over 90 percent of respondents believe (or report) that they are better-than-average lovers.

Pessimism about OD or QWL effects, however, faces growing limits in urging this fifth proposal. Thus, Nicholas’ work indicates real limits on quick and easy assumptions (later chapters will add useful detail, especially in Chapters 11 and 14).

SUCCESS RATES IN PLANNED CHANGE ARE GREATER IN MICRO THAN MACRO SYSTEMS. This proposition implies that much planned change focuses on individuals and small groups. And whatever success estimates characterize such “low” levels, the real pay-offs relate to “big” changes at “high” levels. Here, pessimists urge that OD and QWL either do not focus much attention or, if they do, the poor success rates at such macro levels are unrealistically blended into higher estimates by the dominance of interventions at lower levels.

Again, caution is appropriate. Evidence already reviewed suggests that micro- versus macro-differences are neither simple nor consistent. Readers should also, in their turn, give close attention to Chapters 12–15, among others.

Some Contemporary Evidence About Success Rates in Planned Change as OD or QWL. A substantial post-1990s literature adds substance to the estimates of the consequences of planned change; and this section summarizes that evidence. The record is incomplete, but the propositions just sketched get little support from the evidence, as Exhibit 1.1 establishes.

How comprehensive is Exhibit 1.1? No doubt, it falls short of being exhaustive, but the author strongly believes that any shortfall is not great. Sixteen data sets are isolated here; two were assembled for present purposes for the first time; and only small numbers of those 16 are referenced in any other single source.

Note also several other features of Exhibit 1.1. It employs several operational definitions of “success.” This is inelegant, from one point of view, perhaps, but no doubt it also serves as a conservative estimate of success rates, if anything. Moreover, in several cases, two or more estimates of “success” are involved, and they uniformly generate similar estimates. Of course, this adds to the weight of present generalizations, even if it clearly does not resolve all issues. In addition, the panel does not rely exclusively on published research. To be specific, the largest survey of applications, entry VI in Exhibit 1.1, has about 15 percent unpublished entries. This reduces the likelihood of a “positive findings bias.” See also the discussion related to the fourth generalization above. Finally, as noted earlier, circumscriptions of OD and QWL as similar appear elsewhere (e.g., Skelley, 1989; Golembiewski and Sun, 1990a), but are not repeated here. Generally,
### Exhibit 1.1 Success Rates in Panels of OD and QWL Applications

<table>
<thead>
<tr>
<th>Evaluative study</th>
<th>No. of cases</th>
<th>Success rate summaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Dunn and Swierczek, 1977</td>
<td>17</td>
<td>About 70 percent of these highly selected cases are considered “effective.” See also Table 1.7. The focus here was on situational features of OD applications.</td>
</tr>
<tr>
<td>II. Katzell, Bienstock, and Faerstein, 1977</td>
<td>28</td>
<td>From 103 productivity experiments, these authors selected those which seem to have an OD or QWL base. The judgments may be unreliable, since only one rater was involved. 85.7 percent of the entries are rated as “Studies Finding Improvement,” and the rest fall in the category “Studies Not Finding Improvement.” This distribution is similar to that for the 103 cases abstracted by editors.</td>
</tr>
<tr>
<td>III. Margulies, Wright, and Scholl, 1977</td>
<td>30</td>
<td>Their ratings can be classified in four success categories: 70-plus percent of the observed effects are labeled as “positive,” 6.3 percent as “mixed,” 18 percent as “no change,” and 3 percent as “negative.”</td>
</tr>
<tr>
<td>IV. Morrison, 1978</td>
<td>26</td>
<td>Eight percent are considered “failures,” but most of the studies “did not utilize designs rigorous enough to adequately determine the outcomes of the OD process” (1978, p. 42). Although Morrison’s focus was on methodological rigor, rather than success, it seems that a bit over half of the interventions had “an impact on the processes of the total organization or system” and can thus be tentatively considered as substantially successful.</td>
</tr>
<tr>
<td>V. Porras and Berg, 1978; Porras, 1979</td>
<td>35</td>
<td>Variables changed in predicted directions about 50 percent of the time. Porras’ focus is on hypotheses relating to the comparative effectiveness of different OD interventions, and he finds few consistent differences. Typically, indeed, Porras finds high success rates for both of all pairs of comparisons—50 vs. 37.8 percent, and so on (Porras, 1979, pp. 162, 167, and 171).</td>
</tr>
<tr>
<td>Evaluative study</td>
<td>No. of cases</td>
<td>Success rate summaries</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>VI. Golembiewski and Sink, 1979a, 1979b</td>
<td>44</td>
<td>Interventions came from urban settings, and two independent observers rated each application in one of four success categories: I, highly positive effects; II, definite balance of positive effects; III, no appreciable effects; and IV, negative effects. A single case did not permit classifications, by agreement of the two raters. Other than that, individual ratings had a mean rank order correlation of .73; 27.3 percent of the cases were assigned to success category I, and 61.4 percent to category II, with the balance being roughly split between III and IV.</td>
</tr>
<tr>
<td>VII. Proehl, 1980; Golembiewski, Proehl, and Sink, 1981, 1982</td>
<td>574</td>
<td>Distinguishes 8 classes of OD interventions, in public as well as business arenas, and relies on two indicators of success. A “global” measure relies on independent raters who put each OD application in one of four categories on largely subjective criteria: highly positive and intended effects; no appreciable effect; and negative effects, on balance. Distributions summarized above.</td>
</tr>
<tr>
<td>VIII. Terpstra, 1981</td>
<td>52</td>
<td>Focus on methodological rigor in 52 OD studies, and outcomes are assigned to one of three categories: Uniformly Positive (67.3%); Mixed or Nonsignificant (23.1%); and Uniformly Negative (9.6%). Inter-rater reliability was 100 percent.</td>
</tr>
</tbody>
</table>
IX. Nicholas, 1982

All evaluative studies provide hard-criteria estimates of interventions, and are defined as the proposition of total cases that in post-test show a statistically-significant increase in the expected direction. For 14 targets of change, the overall percentages of positive change are .52, .43, .53, .42, .42, .70, .67, .60, .60, .54, .80, .50, .33, and .45 (pp. 534–536). In addition, not all observers would include all of the classes of OD interventions Nicholas employs—for example, job enlargement and job redesign “without participation.” As it happens, those two classes contain nearly one-third of all cases analyzed (pp. 534–535). The overall success rate might well have been higher, consequently. To illustrate, job enlargement relies on “horizontal” design which, in many cases, merely increases work without corresponding effects on autonomy or control.

X. Guzzo and Bondy, 1983

From 104 productivity experiments, this author selected those that seem to have an OD or QWL base. The judgments may be unreliable, since only one rater was involved. Using the editors’ categories, 90 percent of the entries are “Studies Finding Improvement” and the remainder got assigned to “Studies Not Finding Improvement.” This distribution is similar to that for the 104 cases abstracted by authors.

XI. Guzzo, Jette, and Katzell, 1985

Focuses on worker productivity as it is influenced by “11 types of psychologically based organizational interventions.” Meta-analysis shows that these interventions, on average, “raised worker productivity by nearly one-half standard deviation” which is, of course, a considerable increase. However, not all 11 types are clearly in the OD or QWL traditions, with only five fitting that bill substantially or completely—Appraisal and Feedback, Work Redesign, Supervisory Methods, Work Scheduling, and Social-Technical Interventions. Two other types—Management by Objectives and Goal Setting—are on occasion rooted in the OD/QWL literature. Three of the five best-fit designs approximate or surpass the overall effect size for all 11 classes of interventions. Situational covariants also were assessed.
**EXHIBIT 1.1** Continued

<table>
<thead>
<tr>
<th>Evaluative study</th>
<th>No. of cases</th>
<th>Success rate summaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>XII. Woodman and Wayne, 1985</td>
<td>50</td>
<td>Using several raters, OD applications were assigned to one of 3 categories of success of outcomes: I, Uniformly Positive; II, Mixed or Nonsignificant; and III, Uniformly Negative. 46 percent of the cases were assigned to I, 48 percent to category II, and 6 percent to III (p. 904). To this author, it seems awkward to combine both “mixed” and “nonsignificant.” “Uniformly” is very high standard, both for category I as well as III. No wonder that authors report nearly 50 percent of the cases in category II. Even as is, this study fits comfortably in the range of overall success rates in evaluative surveys.</td>
</tr>
<tr>
<td>XIII. Golembiewski, 1987, 1991; Golembiewski and Luo, 1994; Golembiewski and Bradbury, 1997</td>
<td>133</td>
<td>Interventions are drawn from nation states with GNP of less than $5000 per capita, 1980 US dollars. Subjective estimates were made by multiple observers, and self-reports dominate among variables studied. Also 76 percent of the cases were rated as: highly positive and intended effects; or definite balance of positive and intended effects. See Table 3.</td>
</tr>
</tbody>
</table>
XIV. Golembiewski and Sun, 1989, 1990a 229 A large population of QWL studies provides both “hard” and “soft” measures of productivity. Published studies dominate, but multiple sources were solicited. Overall success rates were very high for both hard criteria as well as global assessment. See Table 1, 2, and 5. Four categories of estimates are used, with highly positive and intended effects requiring that two-thirds or more of all pre- vs. post-comparisons either attain the .05 level or represent at least a 10 percent increase in the intended direction.

XV. Neuman, Edwards, and Raju, 1989 126 Four classes of OD interventions are focal here, and a variety of meta-analysis is employed. The complex results cannot be reviewed briefly, but researchers conclude that their results are comparable to other evaluations: “. . . OD’s effect on overall satisfaction and other attitudes has been moderate but variable” (p. 477). Several special analyses also were performed, and they are relevant “for judging the potential effectiveness and generalizability of OD interventions.” Organization levels of participants and degree of methodological rigor are noteworthy moderators in many of these analyses (p. 477).

XVI. Roberts and Porras, 1992 52 The focus here is on the effects of different OD interventions, and the results are moderate and mixed. For the full set of OD interventions on all dependent variables, the findings seem more arresting, though they get little attention. The authors conclude: “OD, on the whole, has a significant positive impact on the broad range of variables measured in these studies” (p. 202).
Chapter 1

QWL and OD share a basic sense of underlying values while differing mostly in operational versus managerial levels of applications and in the learning designs utilized. See also Chapter 11; this chapter combines attention to the success rates of QWL designs, a close kin to OD, as well as to the situational features associated with QWL applications.

**SUCCESS RATES SEEM SUBSTANTIAL, AND EVEN FORMIDABLE.** As even the brief summaries in Exhibit 1.1 suggest, little justification exists for a rush to judgment about “low” success rates for either OD or QWL interventions. Fifty percent seems to bound the low side of such estimates and most extend into the 70th percentile and beyond. Not all success measures permit such quantitative summary, to be sure, but even their overall thrust is clear enough. To illustrate, three researchers describe their population of evaluative studies focusing on productivity—see entry XI in Exhibit 1.1—in terms that seem quite typical: “Behavioral science techniques for increasing worker productivity are on the whole, effective” (Guzzo, Jette, and Katzell, 1985, p. 291, their emphasis). See also entry VIII in Exhibit 1.1.

The number of separate surveys of OD/QWL effects in Exhibit 1.1 now stands at 16. This approximately doubles the catalog that could be assembled before the 1990s, and obviously reduces the impact of pessimistic views of success rates. Note that Exhibit 1.1 includes several of the populations referred to above, and especially Nicholas and Golembiewski, but several of those earlier cites could not be accommodated to the format in Exhibit 1.1.

Table 1.5 conveniently illustrates these high-range results from the two largest available populations—one for over 500 OD evaluative studies, and the other for more than 200 QWL cases. The category “positive and intended effects” there clearly justifies the label “substantial, and even formidable” for success rates.

All other details in Exhibit 1.1 generally support the clear, even dramatic, bottom line of Table 1.5. Differences in operational definitions notwithstanding, the success rate estimates reasonably vary somewhat. But the range of that variation seems an attractive one, overall.

**PUBLIC VERSUS BUSINESS SUCCESS RATES SEEM COMPARABLE, AND APPLICATIONS SEEM PROPORTIONAL.** The common wisdom seems off the mark. First, in the two largest available populations of evaluative studies, many entries come from the public sector. They approximate the proportion of U.S. public employees to total employment. See entries VII and XIV in Exhibit 1.1, and Table 1.6B.

Second, Table 1.6 shows that the success rates of public versus private applications are both substantial, even as they differ. In addition, while business applications more frequently fall in the highest success category (71.4 versus 59.1%), that difference is narrowed when the two categories of positive effects are compared. Specifically, 86.4 percent of public sector applications fall in these two categories, while 91.5 percent of the business applications do so.
### TABLE 1.5 Summaries of Overall OD and QWL Outcomes

<table>
<thead>
<tr>
<th>Classes of success rates, in percent</th>
<th>I. Highly positive and intended effects</th>
<th>II. Definite balance of positive and intended effects</th>
<th>III. No appreciable effects</th>
<th>IV. Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. OD Outcomes, Global Assessment*</td>
<td>40.1%</td>
<td>46.0%</td>
<td>5.6%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Cases = 574</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II. QWL Outcomes, Global Assessment**</td>
<td>63.3%</td>
<td>28.4%</td>
<td>3.1%</td>
<td>5.2%</td>
</tr>
<tr>
<td>Cases = 229</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1.6</th>
<th>Planned Change Outcomes, by Public versus Business Sectors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I. Highly positive and intended effects</td>
</tr>
<tr>
<td>A. OD Outcomes, Public Sector, Subjective Assessment*</td>
<td>41%</td>
</tr>
<tr>
<td>Cases = 270</td>
<td></td>
</tr>
<tr>
<td>B. QWL Outcomes, Hard Criteria Assessment**</td>
<td></td>
</tr>
<tr>
<td>Cases =</td>
<td></td>
</tr>
<tr>
<td>Public sector</td>
<td>44</td>
</tr>
<tr>
<td>Business</td>
<td>185</td>
</tr>
</tbody>
</table>

Overall, the common wisdom of the second proposition simply does not derive basic support from the data, as two points illustrate. Thus, confirming evidence of high public sector success rates exists, as in entry VI in Exhibit 1.1. Using similar standards to those in Table 1.6B, the four successive categories of success for 44 urban public applications approximate 27.3, 61.4, 6.8, and 2.3 percent, respectively. In addition, some studies even give the advantage in success rates to the public sector (e.g., Guzzo, Jette, and Katzell, 1985, p. 287, or entry VII in Exhibit 1.1). This, of course, is contrary to common expectations.

Sparse data and mixed/weak associations characterize situational features. Little research exists, but it shows inconsistent and modest associations between a broad range of situational features and success rates of planned change (e.g., Dunn and Swierczek, 1977; Golembiewski and Sun, 1990a, 1990b; Roberts and Robertson, 1992). The relevant research studies are not peas in a pod, but these differences do not seem to matter much. Overall, the cumulative associations are impressive only in their scattered and puny quality. Moreover, even where situational associations with success are isolated, the pattern is mixed and inconsistent. In one meta-analytic study, for example, government versus business applications have higher success rates (Guzzo, Jette, and Katzell, 1985, p. 287). See also entry VIII in Exhibit 1.1.

Let us illustrate here with a comparison between evaluations of QWL applications in unionized versus nonunionized locations. As Table 1.7 shows, success rates seem substantial, whatever the organized status of hosts of QWL applications, with the N being large enough to provide a reasonable test, but even the developing nations applications in Table 1.8 outnumber the QWL applications in union settings.

Methodological rigor is no reverse covariant of success in planned change. The underlying argument here has an attractive character and has been developed early and elegantly (e.g., Morrison, 1978), but the sparse literature suggests that the facts provide inconsistent support for the negative association of methodological rigor and success. The work by Golembiewski and Sun (e.g., 1989, 1990a, 1990b, 1992) as well as Roberts and Robertson (1992)

<table>
<thead>
<tr>
<th>Table 1.7 Organized Status and Success Rates, QWL Studies, N = 106</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Highly positive and intended effects</td>
</tr>
<tr>
<td>II. Definite balance of positive and intended effects</td>
</tr>
<tr>
<td>III. No appreciable effects</td>
</tr>
<tr>
<td>IV. Negative effects</td>
</tr>
</tbody>
</table>
TABLE 1.8  "Developing Nations" OD Applications Estimates of Success
Rates, \( N = 133 \)

<table>
<thead>
<tr>
<th>Success Effects, in Percent</th>
<th>I. Highly positive and intended effects</th>
<th>II. Definite balance of positive and intended effects</th>
<th>III. No appreciable effects</th>
<th>IV. Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17.3%</td>
<td>59.4%</td>
<td>20.3%</td>
<td>3.0%</td>
</tr>
</tbody>
</table>

Source: Golembiewski and Luo, 1994; Golembiewski and Bradbury, 1997.

will be relied on to provide broad support for two major points. See especially entries XIV and XVI in Exhibit 1.1.

One conclusion is derived from several statistical studies that deal with the rigor of research designs and success rates of QWL applications. To begin, inconsistent and, at best, modest support exists for the proposition that estimates of success rates decline as methodological rigor increases. Thus, Terpstra (1981) reports inverse relationships; other researchers report that “four of this study’s five significant moderator effects were in the opposite direction” (Neuman, Edwards, and Raju, 1989, p. 478); and Woodman and Wayne (1985) propose flatly that “a significant positive-findings bias does not exist in OD evaluations” (p. 909).

Two caveats apply to this fourth generalization, in addition to urging the reader to wait on Chapter 3. Thus, the evidence is limited, in both business as well as in public applications. Moreover, existing research deals mostly with a limited portion of the full range of methodological rigor, with one major exception (Morrison, 1978). Possibly, then, the expected inverse association of success and rigor might show up if a fuller range of differences in methodological rigor were studied. Indeed, some support for this position (Roberts and Robertson 1992, pp. 918, 923) exists. Due to methodological complexities, however, even the latter researchers conclude that “... we are not optimistic about a final empirical resolution of the positive-finding bias issue... In the end, evaluators who are knowledgeable and conscientious are the most important safeguards against bias” (Roberts and Robertson, 1992, p. 923).

Moreover, the existing research suggests an interesting limitation on the simple pursuit of heightening methodological rigor. In a population of QWL studies (Golembiewski and Sun, 1992), for example, success rate estimates remain stable and high over time. See for example, Table 1.6B. But methodological rigor, as estimated by three scales, varied: first low, then increasing sharply and leveling off, and, finally, decreasing to earlier levels or below.
Why did this pattern occur? Plausibly, the early success rates may well have inspired suspicion, given middling-to-poor early attention to issues of validity. Hence, the sharp increase in methodological rigor, and a persistence of that tendency, as researchers successfully sought to improve the methodological rigor of their designs. However, given the maintenance of success rate estimates at a very high level, the new persistence soon enough waned.

**Mixed evidence about “hard” versus “soft” criteria.** Most surveys of planned change use only self-report data, but available evidence does not uniformly or even regularly associate “hard” data with low success rates. Two sources of evidence support this compound generalization.

As for the first source, as is illustrated above, an early study (Nicholas, 1982) selectively admitted to analysis only evaluative studies of planned change using “hard” or “objective” criteria, with the assumption being that success rates would be lower than then-available studies suggested (especially see entries I through VI in Exhibit 1.1), because the latter highlighted self-report measures. Nonetheless, Nicholas generated fourteen estimates of percentages of success rates—for example, for various classes of interventions—and those success estimates average 53.6 percent.

What is a reasonable interpretation of such average success rates? Nicholas’ average is both noteworthy and, moreover, not grossly different from the overall estimates in such evaluative surveys as those in Exhibit 1.1. Moreover, this generalization is especially significant because Nicholas includes designs—job enrichment “without participation”—that many/most evaluative researchers would exclude from the category “OD” for several reasons. OD does not favor either mere enlargement or “without participation” as earlier details explain.

The implications seem clear: Nicholas’ success estimates can reasonably be said to be underestimated. That is, since the questionable designs are well-represented in Nicholas’ population, the general effect is to lower his estimates of success. Consistently, Nicholas reports estimated success rates for job enrichment “without participation” that average only about two-thirds of the estimates he reports for “job enrichment with participation” (see Nicholas, 1982, esp. p. 535).

Similarly, for a second source of evidence about “hard” versus “soft” indicators and success rates, data justify no rush to conclusions, at the very least. To this point, a single evaluative population was assembled that included both types of indicators for each site of application. QWL interventions were the focus, and relevant data are reported in Table 1.8. That table also provides data for public versus business applications. See also entry XIV in Exhibit 1.1.

Recall the sense of proposition in the conceptual section above: that hard criteria will generate lower estimates of success than self-report criteria. Table 1.9 provides no comfort to this aspect of the common wisdom. To be sure, all success rate estimates (columns 1 plus 2) are attractive. However, for success
<table>
<thead>
<tr>
<th></th>
<th>I. Highly positive and intended effects</th>
<th>II. Definite balance of positive and intended effects</th>
<th>III. No appreciable effect</th>
<th>IV. Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sub-N</td>
<td>%</td>
<td>Sub-N</td>
<td>%</td>
</tr>
<tr>
<td>I. Hard-Criteria Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Sector</td>
<td>26</td>
<td>59.1</td>
<td>12</td>
<td>27.3</td>
</tr>
<tr>
<td>Business Sector</td>
<td>132</td>
<td>71.4</td>
<td>38</td>
<td>20.5</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>II. Subjective Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Sector</td>
<td>26</td>
<td>50.0</td>
<td>12</td>
<td>38.6</td>
</tr>
<tr>
<td>Business Sector</td>
<td>132</td>
<td>66.5</td>
<td>38</td>
<td>28.9</td>
</tr>
</tbody>
</table>

Source: Golembiewski and Sun, 1990a, pp. 666–667.
Success Rates in North America

category I, the hard criteria generate higher success rate estimates for applications in both public as well as business settings. For success categories I plus II, the two estimates are quite similar for public and for business settings, comparing hard criteria to subjective assessments.

This brief review does not propose that self-reports are inevitably tougher criteria of success than objective criteria, but it should discourage knee-jerk repetitions of conventional mantra. For the record, multiple measures are ideal in all evaluations, and they should include both self-report as well as objective measures. Further, objective measures can be very impressive, but such measures are more lauded than readily available. Finally, for some purposes, self-report data may constitute not only the only-available measures, but also the most appropriate ones.

No surprises exist in a further comparison of public versus business applications, also permitted by Table 1.9, which reinforces the analysis introduced earlier in connection with proposition 2. Even as the business and public sector generates higher success rates in category I, for example, both business and public sector estimates are attractive. In the subjective assessment, to be specific, 88.6 percent of public sector applications fall in categories I and II, while the corresponding estimate for business is 92.4 percent.

SUCCESS RATES SEEM COMPAREABLE AT DIFFERENT LEVELS OF ANALYSIS. The evidence is hardly exhaustive, but what exists gives little comfort to the final proposition urging that success will vary inversely with the micro → macro character of the level of analysis. In sum, individual or small-group behavior will be more amenable to change, on progressive average, than change at the level of departments, divisions, organizations, and clusters of organizations, in turn.

What do available data show? There seems no consistent association. For example, consider the largest OD population of evaluative studies so far assembled. Relying on 308 possible targets for change, and distinguishing four levels of analysis, the following percentages of cases at four levels of analysis fell in the expected directions following interventions (Golembiewski, 1990, pp. 22–24).

- Individual level: 78.1 percent
- Leader level: 68.1 percent
- Group level: 77.9 percent
- Organization level: 72.4 percent

Similarly, Nicholas (1982, p. 538) reports that the percentages of statistically significant changes observed after three classes of OD interventions were quite similar at group versus organization levels. Respectively, his percentages of “hits” or successes approximated 50 and 54 percent. Much the same conclusion also seems to apply to most other work (e.g., Roberts and Porras, 1992), although some contrary evidence also exists (e.g., Neuman, Edwards, and Raju, 1989, p. 477).
These summary data do not imply that the several levels of analysis constitute similarly-difficult targets for similar quanta of planned change values and skills. Rather, the view is more limited. When all is said and done, similar effects seem to be achievable by OD or QWL efforts at different levels of analysis. The bottom line is that, planned change theory, experience, and designs seem sufficiently variegated and elaborated to permit similarly effective interventions at several levels of analysis, given the probability that different degrees of difficulty exist in various settings at different levels.

AND HOW LONG DO EFFECTS LAST?

The general opinion has been that, whether or not OD/QWL designs “work” on application, they have short lives; but a family of ironies seems more appropriate than the general certitude. In contrast to the common view, in sum, several lines of evidence suggest that effects tend to last for appreciable periods in some cases, and in other cases some positive effects appear later in an indeterminate time frame even when initial indications suggest a failure. To be sure, few designs have longitudinal features, so certainty is not appropriate. But the evidence seems sufficient to note that the substantial OD/QWL effects reported in this chapter and elsewhere would not be reduced markedly if the effects of designs were often tracked over long periods of time.

Three points sketch the present view of ironies concerning longitudinal effects in OD. In turn, attention goes to high decay rates, then fade-in effects, and finally to an overview of longitudinal research.

Fade-Out Effects in Intermediate Range

Common opinion associates high rates of fade-out with OD/QWL interventions, but that judgment is too primitive, as direct analysis will show. Almost from the start, the OD literature warned about a kind of mixed success/failure. That is, intervention might work at off-site or training sites, for individuals—a general design common then and too much with us even today. But even marked intended effects might be recalcitrant to transfer into back-home settings. This was the case, for example, with an early effort to change autocratic leadership styles (Fleischman, 1962). At the training site, marked and expected effects occurred: trainees much preferred the new, more open and participative style, as rigorous research showed. No sooner were the trainees back at the worksite, however, when training effects dramatically faded out: indeed, so great was the worksite rejection of the new style that trainee attitudes not only regressed but soon became worse than the benchmark levels. At the worksite, then, training effects not only were reduced but, on definite average, trainee attitudes had cascaded below pre-intervention levels! This is an extreme form of failure.
Success Rates in North America

Such effects clearly complicate judgments about success/failure. Depending upon the timing of estimates, researchers could get evidence supporting either success or failure. Of course, more complete interpretations should deal with both, and in credible ways, but that was and is rare.

Even if slowly, OD/QWL theory/practice moved to the position that transfer effects need accounting for, and that the most direct approach to that nest of issues is to make the work group along with supervisor(s) the site of the training. Early demonstrations of this direct way of dealing with transfer effects were available (e.g., Golembiewski and Carrigan, 1970a, 1970b).

No one can say for certain what percentage of “failures” in application fit this first class of fade-out. Nonetheless, it does not seem that fuller attention to time would much change the success estimates in Chapters 1 and 2. Why? Among other reasons, counter fade-in effects occur in OD applications, although with an unknown frequency.

Fade-In Effects After Short-Run “Failure”

Further complicating the judgment about OD/QWL failure, but this time with a clear bias toward requiring higher success rates, fade-in effects need explicit attention. Several examples come to mind (e.g., Golembiewski, 1979), but attention here focuses on the well-known OD application in the U.S. Department of State—Project ACORD or Action for Organization Development (e.g., Harmon, 1975).

Several in-process estimates (largely qualitative, but not exclusively so) suggest clear failure in ACORD—or perhaps highly qualified success is the better descriptor. Again, the basic design focused on “strangers” or “cousins” in off-site training featuring T-groups, in which large proportions of trainees delighted, and for whose values/skills many sought transfer into real-life state settings. At times, truly monumental transfers were attempted, with successes at one or more levels of analysis but not all. In some cases (e.g., Golembiewski and Scott, 1989), the intervenors were so subtle that truly monumental macro-system changes in the last round of applications went unremarked or were only gently connected to OD. In other cases, micro-applications were reported and their successes highlighted (e.g., Crockett, 1970). In most cases, early analyses of ACORD saw far more “failure” than “success.” In reality, probably a more nuanced view would have reported “successes” at one or more levels of application, while justly attributing failure to penultimate systemic changes; an example is the attempt to extend a Planning Programming Budgeting System design to the full U. S. federal government after a successful application at a county level. This result startled politicians such as President Johnson into demanding similar effects everywhere—at the same time with an inadequate sense of the underlying learning designs and values.
Chapter 1

The last story is sad and short, if little appreciated. OD approaches that generated a successful county-wide PPBS application never had a chance of extrapolation to the full federal government. Top political leaders were delighted with the county-wide applications but were apparently innocent of the fact that federal-wide applications would require building rather than mandating (Golembiewski and Scott, 1989). Hence, it is an oversimplification to attribute the national level fiasco to OD “failure.” Rather, OD was not directly applied there.

Several years beyond the ACORD intervention, nonetheless, OD-like effects began to show up in State. These fade-in effects seem largely due—in my view and that of others (e.g., Harmon, 1975)—to the retirement of a substantial number of “old hands,” which left their juniors more free to do in State what ACORD had impressed upon them as a more desirable alternative for organizational relationships. Other interpretations of the same fade-in effects have been made (e.g., Levinson, 1973), but those alternatives do not convince this observer.

To the degree that similar fade-in effects have occurred as a result of OD/QWL interventions, of course, the other success rate estimates in this volume understate reality. No one knows how commonly fade-in effects have occurred, but similar cases—at both workaday and executive levels (e.g., Golembiewski, 1979, Vol. 1)—are easy enough to find in the literature.

Longitudinal Designs, Explicitly Targeted?

Three general conclusions apply to designs that incorporate longitudinal features. First, although the feature can hardly be said to characterize OD/QWL evaluative research, both early as well as recently, the literature contains more than a sprinkling of longitudinal exemplars. Thus, the granddaddy of early applications extended its testing of effects to about a decade and a half down the track (Marrow, Bowers, and Seashore, 1967). Second, public sector applications get reasonable representation among longitudinal designs (e.g., Golembiewski and Kiepper, 1976; Golembiewski, 1991, 1994; Lippitt, Langseth, and Mossop, 1985). Finally, longitudinal designs typically reveal persisting effects, consistent with the theory underlying OD/QWL applications. To illustrate, both regenerative and degenerative interaction are viewed as reinforcing and self-heightening. Hence, once a planned change application reverses an existing degenerative pattern, the ensuing regenerative system reminds one of the baseball Yankees: it tends not to lose for winning. These reinforcing effects can be both powerful and long-lasting, especially when existing work teams are the locus for learning and change.

An Illustration of Effects Over Time

Usefully, these three generalizations are reinforced by a specific case—from an OD application in the service sector whose reach extended more than a decade beyond the period of original interventions (Boss, McConkie, and Golembiewski,
Success Rates in North America

2002) in a 400-bed, for-profit hospital. A third-party intervention was the basic learning design (e.g., Walton, 1987; Lewicki, Weiss, and Lewin, 1992), and the specific first-order targets were the operating room director and a new chief of surgery. They, and their separate staffs, were embroiled in major and persisting conflict, and the third-party design was expected to ease interaction not only between the two principals but also—as second-order effects—between their two work units.

The effects of the design—details about which would burden readers here, but are available to interested parties (e.g., Boss, McConkie, and Golembiewski, 2002)—were substantial and persisting, as shown in Table 1.10. Effects on the interactions of the two principals as well as of their work teams were estimated over a 14-year span, with the major self-report criteria coming from Friedlander’s (1966, 1968). Profile of Organizational Characteristics, whose six major components are labeled in Table 1.10, in ways that should have sufficient face-validity for present purposes. Clearly, the design had effects consistent with expectations underlying OD and its theoretical bases, such as regenerative interaction. Thus, trust increased significantly after the early interventions, and stabilized (or better) at the observation point three months after the initial interventions. Between then and nine months later, GBI measures again escalated, and sustained these higher and intended levels at two later observational points—five years and fourteen years after the initial third-party interventions.

Such self-report data can be questioned from several points of view, of course, but other archival and “objective” measures reinforce the sharp picture of persisting change in Table 1.10. To illustrate the questions, for example, the table reports on only the observations of the two principals as they see the status of the relationships in and between the work units each principal manages. It is reasonable to question these two sets of self-reports over time, of course, but other changes imply that Table 1.10 does not reflect merely self-interested chimera. Consistent with the self-report data but extending far beyond them, the action researchers report that over various but extended periods of time (Boss, McConkie, Golembiewski, 2002, p.1):

Additional results include an increase in the availability of surgical supplies and equipment, a 95% decline in physician abuse of scheduling privileges, a decrease of verbal abuse of nurses by physicians, the reconvening of the OR Standards Committee, a decrease of nursing turnover from 40% the previous year to zero, and a decision by the surgeons to not build an outpatient surgical center.

Note also that such longitudinal results do not surprise on theoretical grounds. To illustrate, once induced, regenerative interaction has self-reinforcing features, and apparently robust ones, to judge from a substantial literature (e.g., Golembiewski, 1993, vol. 1)—in other words, high openness and owning charac-
### Table 1.10 Longitudinal Effects of A Third-Party Intervention

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Group effectiveness</th>
<th>Leader approachability</th>
<th>Mutual influence</th>
<th>Personal involvement</th>
<th>Intragroup trust</th>
<th>Worth of meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational Points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before—After Original Interventions</td>
<td>1</td>
<td>21.5*</td>
<td>54.8*</td>
<td>56.5*</td>
<td>288.0**</td>
<td>43.6*</td>
<td>145.8*</td>
</tr>
<tr>
<td>After—3 Months After</td>
<td>1</td>
<td>2.6</td>
<td>14.2</td>
<td>4.8</td>
<td>25.0*</td>
<td>5.6</td>
<td>112.5**</td>
</tr>
<tr>
<td>3 Months After—1 Year After</td>
<td>1</td>
<td>304.2**</td>
<td>169.0**</td>
<td>289.0**</td>
<td>No data</td>
<td>10.8</td>
<td>162.0**</td>
</tr>
<tr>
<td>1 Year After—5 Years After</td>
<td>1</td>
<td>0.9</td>
<td>No data</td>
<td>No data</td>
<td>1.0</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>5 Years After—14 Years After</td>
<td>1</td>
<td>1.0</td>
<td>No data</td>
<td>No data</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Success Rates in North America

CONCLUDING COMMENTS ABOUT TESTS OF THE COMMON WISDOMS

Responding with doubt and cynicism still has its motivators, but the success rates reported above substantially reduce the steam behind the pessimism. This is one of the virtues of assembling the several surveys of success rates in evaluative studies, here brought together fully for the first time, as far as this author knows. Let us grant the differences in operational definitions of success, the real concerns about methodological rigor, as well as the shortfalls in critical areas of research, inter alia. Nonetheless, the generally substantial estimates of success here constitute a powerful explanation for why OD and QWL are such growth industries. Directly, also, the multiple adoptions and raiding of planned change territory can be explained reasonably by a Polish proverb: Failure is an orphan, and success has a thousand parents.

Moreover, differences in approach do not seem to substantially distort or camouflage success rates. Perhaps the most conservative overview is provided by three researchers who used meta-analytic techniques and also gave attention to methodological rigor (Guzzo, Jette, and Katzell, 1985, esp. p. 290). As for traditional comparisons versus meta-analysis, these researchers conclude: “Overall, the two strategies of review did not produce conflicting results. The crucial general conclusion is the same for both.” And as for rigor, they conclude: “The better controlled the research design, the less impressive are the average results of the intervention, although they are positive even with the most stringent controls.”

At the same time, real caveats apply to overeager generalizations, and much attention should be directed at gaps and shortfalls in planned change. The litany of caveats will not be repeated here, for they have been forcefully and reasonably expressed, both early (e.g., Morrison, 1978) and in more recent days as well.

This summary reinforces two other significant points. Neither OD or QWL are an “it”: rather, they are a “them” of multiple designs that share values but often differ substantially in the degrees to which they require openness to interpersonal analysis, in the degree of initial trust they require, and in the degrees to which they are not applicable in organizations with authoritarian profiles (e.g., Golembiewski, 1991). OD and QWL values are quite clear, but they constitute goals to move toward rather than starting points (e.g., Skelley, 1989). Hence the inappropriateness of such questions: Does “it” (OD or QWL) “work” better, or only, in specific situational or cultural contexts? In contrast, the present results
suggest this generalization: some designs consistent with OD/QWL values can be tailored to many or most worksites and cultures most of the time while moving toward OD/QWL values from different starting points. As such, the success rate literature reviewed above is both incomplete and yet powerful enough to resist camouflage by many intervening and moderating conditions.

Second, a comprehensive treatment of success rates in planned change eventually will require several taxonomies. One taxonomy will detail the full characteristics of OD/QWL designs, and one or more taxonomies will characterize the sites of applications—situational features, specific hosts including (for example) their organizational cultures, developmental profiles of people and units, and environmental differences such as macro societies or cultures.

That time has not yet come, of course.

Several of the following chapters will deal with these great challenges. Numerous chapters relate to design features and their consequences (see especially Chapter 11 for progress on situational features and the success of OD applications).

The evaluative studies reviewed here deal largely (if not comprehensively) with OD/QWL design features, while their associated site-specific taxonomies remain challenges for future work. Moreover, the gross success rates above suggest not only a theoretical but also a practical applicability of OD and QWL in planned change. In addition, the different but noteworthy success rates will encourage the development of the taxonomies relevant to the environments within which hosts exist at the time of application. In turn, such taxonomies will permit finer tuning of design to site of application. This promises even higher success rates than the attractive ones reviewed here.

A final observation serves in closing. Overall, this chapter implies that “organization learning,” broadly defined, is reasonably facilitated by OD and QWL. This provides a solid foundation for such conceptual enhancements as organizational learning—this time defined more narrowly by recent emphasis (e.g., Senge, 1990; Watkins and Marsick, 1993). Consistently, OD and OD/QWL seem to share a substantial normative and conceptual range (e.g., Watkins and Golembiewski, 1995). Hence, the probability of mutual enrichment seems both real and substantial, if the products of this chapter are reasonable estimates of OD/QWL success rates.

REFERENCES

Boss, R. W. (1993). Team building and the problem of regression: The personal manage-


Golembiewski, R. T., and Bradbury, M. D. (1997). An add-on to success rates in less-developed countries. Working paper, University of Georgia, Athens, GA.


Success Rates in North America


Miller, G. J. (1979). The Laboratory approach to planned change in the public sector, Ph.D. diss. University of Georgia, Athens, GA.
Success Rates in North America


Further Weakening the Irony of Ironies
Success Rates in Global Settings

Not very long ago, the literature on OD success rates in non–North American settings also was enmeshed in the throes of negative speculation, but more deeply so. For one thing, OD values poorly fit much of the non-Western world, we were told; and OD was tailored to later stages of economic development characteristic only of patches of the world. And as for success rates, even if some successes existed in North American settings, the results overall, no doubt, would be pitiful and pitiable, even for those rare applications that could be found. Some observers allowed that things might change, and substantially so. But agreement was solid that now was not the time to declare victory over all that ails the theory and experience of OD in cross-national and especially cross-cultural applications. Perhaps most observers never even contemplated such a day. Ironically, no one looked very hard to test the common opinion. So, let’s try to do the obvious thing here.

No need to feign suspense. Success rates are notable, globally. And how can that be? The common view of pathetic worldwide OD effects relates to a stereotype of OD design as focusing on interaction only, and hence as fixated on conditions of high intimacy, as both goal and condition for a good fit to OD approaches. In the present sense, this view is myopic. One can move toward OD values by a specific model of interaction, for certain, but OD also encompasses structural as well as policy/procedure. This implies a much broader range of successful interventions than some observers allow.

We shall soon have the data to begin evaluating the credibility of this overview. Preliminarily, while major issues remain, data-gathering over the last de-
cade or so leaves us in decent shape to estimate OD success in non–North American settings.

Specifically, three separate lines of research on success rates about overseas OD applications are introduced below, and they fit quite comfortably with the largely North American literature reviewed in the previous chapter. In turn, attention here is directed to success rates in:

- OD applications in what may be best called “developmental settings”;
- in a Korean sub-sample of OD applications; and
- QWL applications from numerous settings, both North American and global.

A CONCEPTUAL CONTEXT FOR IRONY I

The thrust of this chapter is direct, in connection with Irony I. What is the record for success in non–North American applications? In general, expectations in the literature are modest: few such applications are said to exist, and that small panel usually is represented as having modest success rates, if not indifferent ones. The estimates are based largely on cross-cultural differences, and the contrary rooting of OD and QWL in Western problems and values.

The implied advice is not to waste time in assessing OD effects in different national and cultural settings, but this is not followed here, and largely because of aversion to easy forfeit. Forfeit may be necessary, but only after confronting the issues with research. This has not been done energetically outside the context of the three related lines of research to be sketched below, but we have a real base on which to build.

What may be gained by testing for cross-national and cross-cultural effects on OD or QWL applications? Such tests may confirm what most observers have believed—that OD (and QWL) is culturally bounded and hence will apply poorly at many worksites, even if perhaps well at others. Testing this view seems a no-lose situation, however. That knowledge could be used to raise success rates, obviously, by avoiding or finessing poor-fit applications. And if the OD applications prove broadly applicable? That knowledge also could be used to heighten success rates by adding applications with favorable probabilities.

As with Chapter 1, the purpose here is to review past OD applications as well as to motivate more of them in the future, if warranted. Hence, this chapter concludes with a sense of how success can lead to greater success. As does Chapter 1, then, this chapter also adds motivational impetus for exploiting Irony I in OD.

As the author sees it, the three surveys of applications in non–North American populations should prove especially useful in both kinds of motivation. Many observers may be surprised about the extent and character of global OD applica-
Success Rates in Global Settings

tions, and this may reinforce the push for further extensions of tests of the cross-cultural applications, if not universalism, of OD and QWL.

**OD APPLICATION IN DEVELOPMENTAL SETTINGS: A THIRD INTERIM REPORT OF WORK-IN-PROGRESS†**

Overall, this section discourages the often dour expectations about the possibility of successful planned change, and especially in those national settings that can be described as “developmental.” That term here refers to jurisdictions having 1980 Gross National Product per capita of $5000 U.S. or less, and this territory is often thought to be terra incognita when it comes to planned change.

Not that the basic issues were in any doubt, at least as far as most observers seem to have been concerned. Early on (e.g., Kirkbride, Duncan, and Tang, 1990), expectations about planned change tended to be low, with two basic kinds of prognoses dominating. Thus, some observers urged that planned change was simply unlikely, if not impossible, wherever and however (e.g., Kaufman, 1985). And other observers saw a bit more promise, but hedged their prognoses with very limiting conditions and constraints (e.g., Johnson, 1990). In some cases, commentators called for collections of culturally homogeneous approaches—for example, “North American OD,” where most applications of OD and the related QWL clearly have been made—as the best and perhaps only hope of building toward generalizations about planned change.

More recently (e.g., Head, 1994; Korotov, Makeskin, and Stepanova, 1995), this sense of pessimism, if not cynicism, is less common and more qualified, but it can still be found in both the managerial as well as the popular literatures. These twin assumptions remain: not much planned change is going on, say, India; and/or any ongoing efforts have little probability of achieving meaningful change in desired directions.

The opposed strategy here seeks to test these assumptions, under conditions unlikely to produce artifactual results. Thus, two framing decisions were made: to deal only with Organization Development (OD) applications, on the ground that this family constituted the most theoretically and practically advanced genre of planned change; and to focus on developing countries where, by reasonable rationale, one could expect the most severe tests of success.

Two earlier panels of evaluative studies of OD applications in developmental settings, 1950–1994, generated 121 cases with an attractive distribution of success rates in “developmental settings”—those national political jurisdictions with a per capita GNP of $5,000 U.S. or less in 1980. These early results provide a basis on which to build.

---

† An earlier and more extensive version of the rest of this section appeared as Golembiewski and Luo, 1994.
Extending the period of observation through mid-1998 does not require any substantive modifications of the conclusions associated with the earlier searches, and the additional applications make several contributions to theory and practice. The newly augmented panel now contains 201 cases, an increase of approximately 66 percent. Since several reports deal with multiple sites, \( N = 240 \) separate applications. In addition, the full panel now contains cases from 57 political jurisdictions, usually nation-states, which amounts to an increase in coverage of approximately 25 percent over the two panel-building efforts.

If anything, both the rate of applications and the estimates of success rates still seem to be increasing, and the end is nowhere in sight. These data may surprise some observers. Usefully, this study adds to a family of studies dealing with the substantial success of OD applications. Earlier work on assembling panels of applications dealt with Western or largely North American settings, with a specialized focus on the broad family of OD applications as well as on evaluative studies labeled Quality of Working Life, or QWL. See Chapter 1 for a review of this earliest work.

This section builds on other efforts that focus—in two earlier reports as well as here—on developmental national settings. Thus, an intensive search was first undertaken for OD applications in these settings through 1987 so as to determine the incidence of applications as well as to estimate success rates (Golembiewski, 1991). Then, this parent study was extended through 1993, for the obvious reasons (e.g., Golembiewski and Luo, 1994). In addition, various tests to check for the culture-boundedness of OD applications also were undertaken (e.g., Johnson and Golembiewski, 1992).

The present products of this strategy are quite direct, even as they rest on a mountain of experience and theory. Thus, the two earlier waves of observation in the present line provided substantial support for four conclusions about the two searches: numerous OD applications were found in many national settings; estimates of success rates imply much of use was accomplished; applications extend over a substantial interval of time, and may even be growing in incidence; and OD as a whole does not appear culturally bounded in any direct sense. In sum, although some OD designs can be more appropriate for one cultural setting than others, OD encompasses such a range of designs that some, at least, are applicable under most or all diagnosed conditions existing in different sites (Golembiewski, 1991).

Why this extension through mid-1998? Consider here only a few elements in a positive rationale. Thus, the significance of this effort grows in our increasingly global economy, now burgeoning in ways only dimly envisioned in the days of the original study. In addition, each additional year of observation helps test the easy interpretations of OD as faddish and culture-bounded. Finally, adding to the observational period in effect permits the analytic net of OD to be cast...
further—to additional nation-states as well as to different languages, and perhaps to more cultural settings.

**Some Details About 1998 Add-Ons**

Several details get us started. They relate to an extended rationale for this research; some conventions for search and measurement; and, finally, to results of this third round of search. Together, these details reflect the intention to reduce the provisional or speculative character of so much of the comparative OD literature, and thus to add momentum to a basic shift toward what does happen in planned change, and away from fixating on what might happen.

**Energizing Rationale**

The degree of “cultural fit” between OD values and various developmental settings is the key issue, and three points outline the rationale energizing the present study. First, diagnosis is primary in OD and, hence, the present concern is simply a specific manifestation of the general rule-of-thumb. Thus, given a specific diagnosis, wherever, the question is whether some design for change can be crafted consistent with OD values while also providing an “appropriate fit” for dealing with the diagnosed condition.

The general answer to the questions seems to be: Yes, such a fit seems possible in a large proportion of cases. Elsewhere, eight kinds of OD designs are distinguished, and they relate to a very broad range of diagnosed conditions (e.g., Golembiewski, 1991). They seem to provide useful leverage for change, and quite broadly.

Second, OD designs encompass such a large range that—within imprecisely known limits—an array of different local cultures and conditions can be accommodated by sub-sets of OD designs. As noted elsewhere (Golembiewski and Luo, 1994, p. 296): “Illustratively, interaction-centered designs are less appropriate to situations where differences in status and hierarchy preclude authentic relationships as well as require a high degree of formalism. Also, structural and policy changes often are more appropriate as start-up designs in autocratic or authoritarian settings and so also may be some survey/feedback variants.”

Very often, criticisms about OD shortfalls in different cultures reflect a pronounced myopia that fails to recognize these several classes of designs, thereby finessing the possibilities detailed above. This common view fixates on interaction-centered designs—and perhaps especially on the more dramatic varieties in T-groups or affective-oriented team-building. Some observers conclude, too broadly, that all of OD has a similarly poor fit to some local conditions such as social distance or high individualism. It seems more appropriate, given the present range of interventions and substantial success rates, overall, that some
OD designs seem applicable in all or virtually all work settings. Diagnosis is the key to exploiting this broad-gauged potential.

Third, OD in a normative sense intends to guide clients in moving from “here” to “there.” Hence, as we noted elsewhere (Golembiewski and Luo, 1994), OD shoptalk advises that one “should start from where the client is,” and then to move progressively toward OD values. As we extended the basic point:

ODers see themselves as working toward their values, rather than as insisting on them *ab initio* . . . For most practical purposes, then, the issues for ODers relate to the scope and speed of progress beyond the original condition and toward closer approximations of OD values. Slow and narrow progress may be seen as “copping out” or “cooling-out” a client so as to avoid changing local conditions favored by management but considered undesirable by employees. For some practical purposes, and for different ODers, various initial degrees of value incongruence may be “too great” (e.g., Golembiewski and Kiepper, 1988, pp. 205–226). In principle, however, value incongruencies typically will exist between all potential OD hosts and OD values. (p. 296)

To be sure, some here → there gaps can be so normatively great that an “appropriate fit” is not available for any OD effort. For example, Johnson (1990) makes just such a claim about Venezuela. Usually, however, this matter of normative gap is a concern to be accommodated rather than a *cul de sac* precluding movement.

Put another way, OD values do not have to be initially congruent with those of the host. Were that the case, clearly, OD interventions would have little to recommend them. Growing congruence with OD values is the goal, rather, no matter what the starting point and its distance from OD ideals.

**Conventions for Search and Measurement**

The search here for additional OD applications is basically the same as in the two earlier efforts. Thus, most applications come from a direct check of about 80 English-language periodicals published throughout the world during the period 1994–mid-1998. The list is available from the author.

Two other sources of applications deserve note. Experienced interveners also were solicited for contributions that were consigned to file cabinets and never published. Inventories of academic theses and dissertations also were searched for candidates.

Moreover, several literatures in languages other than English—especially Korean and Polish—have been searched in this quest for new OD applications, whatever the year.

In all cases, the target political jurisdictions come from Kurian’s list (1984, pp. 98-99), with rankings for 1980 per capita GNP under $5,000 US. Specifically,
Success Rates in Global Settings

this study relates to rankings 39 through 171, and includes a few cases from jurisdictions that had no Kurian ranking.

Results of 1994–98 Add-Ons

A baker’s dozen of generalizations highlight the results of the third search for OD applications in developmental settings. In general, the additional cases add credibility to earlier generalizations and further encourage a view of OD’s very extensive reach and its noteworthy grasp. These are especially prominent in the context of the low expectations most observers held about the fate of “North American OD” throughout the world, or very substantial parts of it.

General Literature on Change

As for the North American literature, the largest proportion of citations related to planned change worldwide have a non-empirical character, and even seem boosterish. Overall, they propose that OD has much to offer in (let us say) Ireland; those citations go on to detail aspects of Ireland’s cultural heritage or present conditions that are both needful of but unamenable to OD; and the typical article then notes that in Ireland, work going on in OD is either sparse and undocumented.

Such speculative pieces are sometimes done well (e.g., Scott, 1992); at times they are poorly and incompletely assembled, and perhaps most often are well-meaning but uninspiring or uninspired. Whatever their status, such studies get no further attention here.

Number of Empirical Evaluations

Empirical studies of OD applications continue to appear in prominent numbers in the literature, as Table 2.1 suggests. Perhaps the word “surprising” best fits these distributions of applications. And why surprising? Early on, most seasoned observers expected few applications in developmental settings, and especially those limited to empirical reports in the English language. And even now those reports still appear in substantial numbers, when faddishness presumably has diminished and when a past record of basic failure would have probably discouraged OD efforts.

The numbers remain small, of course, but the distribution of OD applications over time suggests, if not a tsunami of attention, at least substantial and clearly growing attention over the last 30 years or so. In short, mere enthusiasm for OD probably would have dissipated a long time ago.

Diffusion of Applications

OD applications through the end of the present observation frame (mid-1998) have been fielded in 57 “developing nations,” using the present definition. This
TABLE 2.1 Distribution of OD Applications Over Time

<table>
<thead>
<tr>
<th></th>
<th>Original panel</th>
<th>Two additions</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>No date</td>
<td>2</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>Prior to 1960</td>
<td>1</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>1960–64</td>
<td>0</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>1965–69</td>
<td>8</td>
<td>—</td>
<td>8</td>
</tr>
<tr>
<td>1970–74</td>
<td>26</td>
<td>—</td>
<td>26</td>
</tr>
<tr>
<td>1975–79</td>
<td>22</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>1980–84</td>
<td>31</td>
<td>5</td>
<td>36</td>
</tr>
<tr>
<td>1985–89</td>
<td>24</td>
<td>11</td>
<td>35</td>
</tr>
<tr>
<td>1990–94</td>
<td>—</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>1995 and later</td>
<td>—</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

* From Golembiewski (1991, p. 208). N = 114 separate citations concerning 112 applications for which success rates are estimated.

* From Golembiewski and Luo (1994, p. 298) and from Appendix B.

NB: In all cases, the dates refer to the published evaluation reports, on the obvious ground that (for most purposes) public knowledge is reasonably dated from that time, as contrasted with the date of intervention.

adds a dozen jurisdictions to the previous list. In alphabetical order, the new entrants come from Bulgaria, Egypt, Fiji, Gambia, Hong Kong, Lithuania, Mongolia, Morocco, Philippines, Uganda, Zimbabwe, and the “old USSR.” See Appendix 2.1.

There are a total of 132 listings of nations meeting the present GOP criteria in Kurian’s catalog (1984). Hence, the present data indicate that OD has penetrated into nearly half of what many consider this very unlikely population of possible hosts. At the same time, that penetration is clearly incomplete.

Note again that Tables 2.1 and 2.2 do not imply that all OD designs are applicable everywhere; but those data do constitute a strong case for the proposition that the full inventory of OD designs contains some exemplars applicable across that full range of settings that generated the present panel of 201 OD applications. As was noted elsewhere (Golembiewski and Luo, 1994, p. 299), “Most of these [applications] occur in ‘normal’ organizational sites, but a large proportion relate to out-back applications—as in dealing with ‘river blindness’ in a village setting” (p. 299), with community building in the hinterlands, or with attempts to build leadership resources among the poor.

Economic Rank Ordering of Nations Hosting OD

Other sources summarize the diffusion of OD through the “developing nations,” as currently defined. The additions to the present panel of applications begin with number 39 on Kurian’s list (1984), which is approximately half-way down his
category “Upper Middle—Under $5,000.” The last addition to the full panel is 169, or Bangladesh. Kurian’s list extends to entry 171—Bhutan, with a GNP per capita in 1980 of $80 U.S., the smallest GNP on his “Bottom 10.” That category is preceded by 74 countries in a “Lower Middle” category.

**Estimates of Success Rates**

The full record for success rates of OD applications overseas seems clear enough: the estimates in this and the preceding two searches suggest a noteworthy success rate in “developmental settings,” as well as in the panels of applications in North American settings reviewed in Chapter 1. However, the two profiles do differ. Table 2.2 contains relevant summary data for both generalizations. The sharpest differences clearly appear in two categories—the smaller proportion of the “best effects” and the greater proportion of cases of “no appreciable effects” in the applications in the developmental settings.

The determined reader can get details on the present panel of OD applications from several original publications (e.g., Golembiewski, 1991; Golembiewski and Luo, 1994; Golembiewski, 1998). Lengthy exhibits there briefly characterize the sites of all OD applications, the kind of designs employed, and the consequences observed.

Two major methodological notes are appropriate here, moreover. First, the assignments of success rates have acceptable reliabilities, but not unassailable ones. For the 574 applications in developed settings, for example, three independent estimates of success rates are made. Any differences were then discussed and, if not reconciled, the lowest estimate in question is chosen. Here, two raters are employed.

**Table 2.2 Two Estimates of OD Success Rates**

<table>
<thead>
<tr>
<th>Applications in full panel of developmental settings, N = 240</th>
<th>Applications in developed settings, N = 574</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success rate categories</td>
<td>Public sector</td>
</tr>
<tr>
<td>62 25.8 I. High positive intended effects</td>
<td>41% 40%</td>
</tr>
<tr>
<td>114 47.5 II. Definite balance of positive and intended effects</td>
<td>43% 49%</td>
</tr>
<tr>
<td>41 17.1 III. No appreciable effects</td>
<td>7% 5%</td>
</tr>
<tr>
<td>13 5.4 IV. Negative effects</td>
<td>9% 6%</td>
</tr>
<tr>
<td>10 4.2 ? No reported effects</td>
<td>— —</td>
</tr>
</tbody>
</table>

* See also Table 2.3.
Hence, if anything, the present success rates have a conservative bias. Typically, cases-in-contention involve I or II judgments.

Second, interpretations of Table 2.2 no doubt should be cautious. Directly, studies in the full panel do not constitute paragons of methodological rigor or subtlety. For example:

The validity of research designs still is suspect in many cases;
The modal design relies on a pre-test and a single post-test perhaps one to three months later; true longitudinal design remain rare;
Only a few variables estimate effects in the typical application, and these variables are dominantly self-reports as contrasted with "hard" or objective variables; and
In most applications, little attention goes to the situational features that might encourage the application of a particular design for change, or contraindicate it—e.g., union vs. non-union.

Favorable Trend in Methodological Concerns

The caveats above notwithstanding, the concern about rigor seems to be decreasing, and for several reasons. Four introductory items on the fuller bill of particulars will have to suffice here. Thus, some recent studies have utilized broad batteries of variables, both "hard" as well as "soft," and without any appreciable dilution of success rates (e.g., Basu and Das, 1998; Arthur, Doverspike, and Kathy, 1996). In addition, experimental and quasi-experimental studies lately seem more common (e.g., Oh, 1995; Park, 1990), and they still generate a similar pattern of effects.

Over the years, to add a second item to the present bill of particulars, success rates have remained both relatively positive and constant. Indeed, one can easily argue—based on Table 2.3—that success rates have improved over the period of 30-plus years during which applications were found. Not every comparison in Table 2.3 has the same character but, for example, the OD applications rated in success category I increase, over time, from 14.3 percent, to 19.1 percent, and then to 39.3 percent. To a similar effect, III and IV ratings also decrease in the third column vs. the average of columns 1 and 2.

Moreover, to move on to a third positive component of a trend that decreases methodological concerns, the panel of applications reflects a growing concern with situational features, as well as with different interventions that "fit" them better/worse, consistent with a trend in the cross-cultural literature (e.g., Hofstede, 1990). Finally, but only for present purposes, high success rates do not seem artifacts of low methodological rigor and validity (e.g., Golembiewski and Sun, 1990a).

Beyond these four preliminary features leading to decreasing concerns about methodological rigor, two broader-gauge issues also deserve note. To be-
**TABLE 2.3** OD Success Rates, 1960–1997, in Developing Countries

<table>
<thead>
<tr>
<th>Success rate estimates</th>
<th>Original batch&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Add-on batch&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Recent batch&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Highly positive and intended effects</td>
<td>16/14.3%</td>
<td>4/19.1%</td>
<td>42/39.3%</td>
<td>62/25.8%</td>
</tr>
<tr>
<td>2. Definite balance of positive and intended effects</td>
<td>56/50.0%</td>
<td>16/76.2%</td>
<td>42/39.3%</td>
<td>114/47.5%</td>
</tr>
<tr>
<td>3. No appreciable effects</td>
<td>24/21.4%</td>
<td>1/4.8%</td>
<td>16/15.0%</td>
<td>41/17.1%</td>
</tr>
<tr>
<td>4. Negative effects</td>
<td>10/8.9%</td>
<td>0/0.0%</td>
<td>3/2.8%</td>
<td>13/5.4%</td>
</tr>
<tr>
<td>? No effects reported (e.g., &quot;non-starters&quot;)</td>
<td>6/5.4%</td>
<td>0/0.0%</td>
<td>4/3.7%</td>
<td>10/4.2%</td>
</tr>
<tr>
<td>Total</td>
<td>112/100%</td>
<td>21/100.1%</td>
<td>107/100.1%</td>
<td>240/100%</td>
</tr>
</tbody>
</table>

<sup>a</sup> Differs from earlier estimates in a few particulars. Primarily, six cases were earlier reported as "nonstarters" and not considered further. Here, these six cases are included under the label "?” and are involved in calculations of success rate estimates. See Golembiewski, 1991, esp. pp. 210–222.

<sup>b</sup> Golembiewski and Luo, 1994, pp. 297–299.

<sup>c</sup> Calculated from Appendix below. These cases number 68, but multiple sites permit 107 estimates of success.
gin, and of more than passing interest in connection with rigor, the QWL literature implies an arresting but credible pattern concerning the association of methodology and validity, over time. Methodological rigor starts out low in QWL’s earliest days, then increases to, and peaks at, a much higher level for a decade or more, and then plummets to original levels, where it seems to have remained. Three estimates of rigor are utilized in this research (Golembiewski and Sun, 1990a), but details are finessed here.

No degree in rocket science is required to develop a credible interpretation. Early successes are viewed skeptically, to begin this interpretation, and can be seen as artifactual—as in reflecting a “positive-response bias” resting on weak methodology. Hence, an insistence about heightening methodological rigor. But success rates remained constant, if they do not increase, as rigor increases. So why bother with methodological exotica when managements hosting change efforts have a reinforced conviction that such efforts “work” in a pleasing proportion of cases?

A final note under this heading also decreases concern about methodological rigor. Specifically, growing evidence (e.g., Golembiewski and Sun, 1990b, 1998) implies that OD and QWL change efforts rest on a useful set of guides for praxis that, if followed, are associated with heightening success rates. Relevant research deals with twelve such guidelines for applications, which seem to be most closely associated with three of the four major categories of interventions—Human-Processual, Technostructural, and Combined—but are less associated with Sociotechnical Systems interventions. This evidence provides a growing source of confidence in OD and QWL applications. See also Chapter 11.

Success Estimates as Conservative: Conceptual Components

Not insignificantly, the estimates of success rates implied by Tables 2.2 and 2.3 are probably conservative, as a three-point argument implies. Let us begin with a conceptual perspective, even given interpretive cautions. That is to say, to begin, substantial estimates of success rates have existed over an interval of some 35 years, more or less (Golembiewski, 1991). During that interval, moreover, we have learned much about how to increase success rates; for example, concerning the “fit” of change designs to different diagnosed conditions (e.g., Golembiewski, 1991, esp. pp. 220–224); and about the relevance of large-organization designs as complements of the interaction-centered and small-group-oriented designs so characteristic of much OD and QWL. However, apparently due to success rates that are “high enough” even when learning designs are innocent of such elaborations, much of the practice of planned change finesses such fine-tuning. For those interested in the full argument, they are now reading the correct book to provide numerous perspectives on how success rates could be increased, even given already-high estimates.
Success Rates in Global Settings

Success Rates as Conservative: Practical Components

The probably conservative nature of the success rate estimates in Table 2.2 also is implied by several practical or operational features, with three proofs of this deliberate practical bias sufficing here. First, the conventions here were such as to deflate success rates. In one multisite report (Arbose, 1982), we are told that six of twelve cases “never got off the ground.” Consistent with the present conservative guidelines, the six aborted interventions are coded “III,” although it is not certain why they were nonstarters. This decision typifies all those made in this survey.

Second, several of the present measurement conventions also should serve to deflate success estimates, if anything. Recall that any differences between raters were resolved arbitrarily by adopting the lowest estimate in question, if raters could not agree after discussion.

Third, the full panel contains a growing proportion of large-organization changes (e.g., Arthur, Doverspike, and Kuthy, 1996; Basu and Das, 1998; Dugan and Prasai, 1990; DuToit, 1987; Rigg, 1992). This diffusion suggests the broadening application of increasingly powerful theories and techniques, as well as useful guidelines for their application in practice, not to mention the apparently burgeoning confidence levels among both ODers and their clients.

Cross-Cultural Features as Moderators

The issue of cross-cultural differences in values remains a central one, with Hofstede’s (1980a, 1980b) four behavioral dimensions usually being central in such analyses. Typically, to explain, the literature (e.g., Magalhaes, 1989) sees a disjoint between OD and analyses like Hofstede’s. Simply, OD and QWL are seen as valuing increased employee participation, as well as power-equalization principles. However, among other features, Hofstede’s research suggests that many countries in the present sample frame would score high on “power distance,” which implies a significant mis-fit between such a common cultural feature of “developing” societies and the values underlying OD and QWL.

The present success rate estimates do not give great currency to this common view; and our attempts to test the proposed regularities have not supported this view either (e.g., Johnson and Golembiewski, 1992). Elsewhere, this author speculates on this proposed disjoint (e.g., Golembiewski, 1990, esp. pp. 220–224) and, basically, urges two points. Thus, the common view of the reach-and-grasp of OD/QWL is oversimple; and the usual interpretations of Hofstede’s results are too exuberant.

In any case, the present success rates do not give strong support to the common notion of a fundamental disjoint between “developing” cultures and OD/QWL values. What can be said safely is that any culturally loaded dysjoints
serve only as moderating or intervening variables in the present panel of applications—real enough, but neither insurmountable nor so robust as to distort or camouflage the pattern of results reflected in Table 2.2.

The present view is consistent with De’s (1984) sense of the matter when he reflected on his own survey of work-system redesigns in India, broadly socio-technical in character. De observes that the skilled intervenor usually can adapt to cultural differences, even as certain generic situational features can influence the success/failure of interventions. De concludes (1984, my italics):

. . . It is difficult to identify any aspect of predominant organisation culture prevailing in India that can be upheld as critical for the success of a project. In negative terms, opposition or total apathy at the top level will harm any effort at change. In positive terms, management at one or two levels above the committed parties, if definitely supportive, is of significant help.

Associated with success were factors like a minimum continuity of the top management cadre and the availability of change agents. Frequent changes of committed personalities often acted as causes for delays. (p. 88)

De also provides a major hint about the acceptable fit of sociotechnical approaches to cultural features common in India—traditional concepts of authority as contrasted (for example) with the low social distance and participation characteristic of “human-processual approaches.” De adds (1984):

. . . practically all the external consultants and most of the internal consultants involved in the Indian cases have had an organisation development focus in their earlier work experiences. Some of them were active in the organisation development movement with a high degree of process orientation. However, OD work [defined as] dealing with the people system somewhat divorced from the task system, did not provide them with adequate satisfaction in their role and they steadily moved towards the socio-technical systems approach. (p. 87)

In sum, De’s scope of planned change encompassed both sociotechnical as well as human-processual approaches, and the former had the advantage of a closer fit with some Indian realities, even though these did not provide as comfortable a cultural nexus for sociotechnical applications as did/do the Scandinavian countries. As De concludes (1984):

The Indian experience, however limited in number, does indicate that it is possible to establish a culture of industrial democracy even in a large pluralistic society. . . . Even in a traditional Hindu pilgrimage centre like Hardwar, responsive behaviour was generated as a sequel to
Success Rates in Global Settings

a nurturing organisational milieu and not because of transfer of technology. This points to an inner urge the workers had, to get some meaning out of their own day-to-day work experience. (p. 99)

Like De, in my vocabulary (e.g., Golembiewski, 1979, 1993), I see “OD” as encompassing both human processual as well as sociotechnical approaches. These differ in terms of history as well as techniques and focus, with the latter being more congenial to the shop floor and the former more to managerial or executive levels; but both human-processual and sociotechnical approaches can share the same basic values and aspirations (e.g., Skelley, 1989). For cultural details, see also Chapters 7–9.

Alternative Range of Success Rates

Table 2.3’s estimate of success derives from a substantial panel of evaluations relying on empirical data, and that estimate also is higher than many other estimates that are anecdotally based.

Which estimate is more correct? From one perspective, both points of view may be correct in their own ways. Consider Babcock (1994, p. 58), who proposes that the “failure rate of O.D. programs [in Hong Kong] has been high but this rate has been camouflaged by an expanding economy.” He also notes that “… the quality of these programs [comes] in all quality ranges [high to low to ‘rip off’].” The low-to-“rip-off” OD programs probably would not get in the present population: the change agents would be unlikely to write-up their experiences, for one thing, and even if written-up, the products probably would not be accepted for publication.

Trade Unions and Worksite Redesign

Especially in India, the impact of the trade union movement was manifest in the augmented panel of applications (De, 1984). If the United States and United Kingdom, overall, represent a “distributive collective bargaining culture [and] confrontative” relationships, and if the Scandinavian countries, especially Sweden, provide strong support for the processes of worksite redesign and diffusion, then India falls somewhere in-between. A number of Indian applications demonstrate that worksite changes at operating levels are possible without labor unions as an active partner. No doubt, however, active trade union support could enhance both the success of individual applications as well as the diffusion of successful designs. The probable effects of union opposition also seem clear enough, in general.

Internal and External Consultants

The OD applications in the augmented panel provide some evidence about the desirable blend of internal and external consultants. The usual rationale for pairing both is direct: the internal consultant, or direct employee of the client, will
be the expert in the local culture and history; the external specializes in OD techniques and theory as well as in breadth of experience. That constitutes an often-useful combination.

Several cases in the panel bear on this blend. Thus, one failure clearly seems in good measure a function of the inadequacy of internal ↔ external ties (Bussom, Elsaid, Schermerhorn, and Wilson, 1984). And De (1984) sees another pattern, common but not general, in a covey of Indian applications. De notes (1984): “while the role of an external consultant is gradually becoming redundant after the initiation phase with some degree of stabilisation having taken place, [that role is] still considered relevant for the diffusion process. The Hardwar unit of Bharat Heavy Electricals and the post office cases—two comparatively successful cases—are examples of this dependence” (p. 88, my emphasis). The corresponding need for a training-of-trainers feature in consultations—to provide a supply of internal consultants—is noted with some frequency (e.g., Bussom et al., 1984).

REPRISE

The 1994–98 add-ons of OD applications in “developing nations,” as here defined, add bulk to the inventory of evaluations and contain only pleasant surprises. Thirteen generalizations develop this overall view, and three will suffice here to reframe the analysis. Thus, the extended period of observation suggests that the attention to OD has been maintained, if not increased. In addition, the success rates remain high. Evaluations of OD applications have been hosted in 57 nations; and the full panel now contains 201 evaluative studies, which permit 240 total estimates of success. No doubt, the cases that eluded this analysis still represent some large multiple of those OD applications that have not yet been cataloged because they are in various non-English languages, sit in some consultant’s desk drawer, or otherwise have eluded our three searches. This fact encourages continued search for additional OD applications in the broadest range of sources—from consultants’ files, academic theses and dissertations, and published sources in a growing range of languages. One Korean effort of this kind gets attention below.

Despite their partially cross-checking character, as well as their replication in part by several independent investigators, the results of this chapter and its predecessors have clear limitations. Patently, for example, multiple measures of success by independent estimators should be employed more frequently. The present research has attempted to build-in safeguards against bias via the composition of its several teams of raters, but these provisions can hardly be called independent. Several measures of change also have been used in the research sketched above, but replication is always appropriate and may be valuable. For details, consult Chapters 4 through 6.
In addition, the results above clearly require finer-tuning of effects. Directly, given the high overall success rates, what specific conditions (if any) will induce different levels of effects for specific kinds of designs? Existing work of this kind is both rare and inconclusive (Dunn and Swierczek, 1977; Golembiewski and Sun, 1990b). Work with a large batch of evaluative studies is under way, however, and holds promise of isolating quite regular effects for different classes of learning designs under a range of situational features (Golembiewski and Sun, 1998). See also Chapter 11.

IMPLICATIONS FOR MANAGEMENT

This study adds to an accumulating family of research that encourages management’s reliance on OD values and methods, to put the matter in its simplest form, with Chapter 1 providing much illustrative support with a Western flavor. Thus, the earliest work (e.g., Golembiewski, Proehl, and Sink, 1981, 1982) dealt with OD success rates, largely in Western or North American settings. Its findings of substantial success rates ran counter to prevailing beliefs (e.g., Burke, 1980, 429–430), but nonetheless received independent support (e.g., Nicholas, 1982). By the 1990s, indeed, nearly a dozen and a half independent surveys of evaluative studies were reported (Golembiewski, 1998), and these numerous panels reinforced the original findings, almost literally without qualification.

Subsequently, two major extensions of this basic orientation have been made, in addition to this chapter’s demonstration that “developing settings” need not be viewed as recalcitrant to OD application due to the latter’s allegedly pervasive cultural bias. Relatedly, high success rates have been reported for what can be viewed as a branch of OD—what its proponents call Quality of Working Life, or QWL (e.g., Golembiewski and Sun, 1990a). A later section illustrates this work by detailing the results of a recent follow-up to the 1990 results.

In addition, in response to the criticism that OD had a middle-management focus and QWL suffered from a “lower level bias,” a batch of applications at the “politics/administration interface” was assembled (Golembiewski, 1999; Golembiewski and Miller, 2002), and its findings also support OD’s efficacy.

All in all, then, this full line of research implies a substantial case for the broad managerial use of OD values and approaches—in numerous designs, at various levels of organization, and in different cultural and/or national settings.

DIRECTIONS FOR FUTURE RESEARCH

The end is nowhere in sight, of course, despite such useful reinforcement of evaluative studies. To suggest the possibilities, two kinds of further work seem reasonable. Obviously, to begin, additional languages should be explored to en-
large the total panel of OD applications. For example, an ongoing search of the Korean literature on planned changed has found about a score of OD applications (Yoon, 2002). The next section illustrates this useful work. In addition, learning approaches related to OD and sharing its values also deserve attention, for obvious reasons. The second section below takes such a tack in assessing Quality of Working Life applications. OWL has a close kinship to OD.

**OD In Korea: Assessing Incidence and Success On Confucian Turf**

The closer one looks, it seems, the more evidence of OD activity shows up in both numbers and successful effects. The most recent such assay of OD involves Korean applications reported in that language. The task was considered to be a search that was ordained to yield modest results: few applications were expected by the researcher, Yoon (2002), a Korean-speaking graduate student, on the basis of initial interviews; modest success was expected in whatever rare applications existed, for Confucian influences were in general said to provide poor humus for the growth of OD values and approaches (e.g., see Chapter 9). The graduate student had confidence that the common opinion was correct.

In addition, the odds were loaded against Yoon being surprised. The study period was a seminar under the 10-week quarter system, with no opportunity to travel to Korea, and with real constraints on developing a pool of informants. Basically, the Korean replication used the design conventions detailed in the previous section.

So what did Yoon find? Using methods very much like those described above, Yoon reports that the term “OD” came into Korean use in the late 1980s. Most informants contacted expected few OD applications in Korea, and especially in the context of labor union frictions that induce labor/management conflict rather than compensation. Given harsh time and distance limits, Yoon none-theless found 20 cases. Their success rates had this distribution:

<table>
<thead>
<tr>
<th>Applications</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Highly positive and intended effects</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>II. Positive and intended effects</td>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>III. No appreciable effects</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>IV. Negative effects</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>— Not available</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
In addition, Yoon reports on various practical and professional developments that suggest a healthy early condition for Korean OD. This situation also surprises most close observers, and Yoon had to penetrate beyond both the popular and academic literature to get direct indications of the state of OD activity.

**A Recent Survey of QWL Evaluations: Success Estimates of a Close-Kin Design**

A second important way of extending the assay for diffusion and success of OD values and approaches involves assessing close kin, such as Quality of Working Life. QWL rests on values virtually indistinguishable from OD; note that ODers and QWLers can work both sides of the street of planned change, as is true of this author. In due time, various distinguishing features of OD and QWL get brief attention, but the focus here is largely limited to success rates. In that objective, attention below goes in turn to methods of analysis and then to selected findings about QWL applications.

**Methods of Analysis**

This section reports on some substantial effort to isolate a panel of QWL applications. To be more specific by way of introduction, four sets of details provide a working familiarity with the search methods used to develop perspective on QWL success rates and diffusion. In preview, in turn, attention below goes to the search for QWL applications; certain measurement conventions; selected analytic procedures; and outcomes of QWL applications.

*Panel of QWL Applications.* A comprehensive collection of applications, 1965–1996, comes from five sources. Sun (1988) provides details but, for present purposes, the multi-pronged search identified applications from:

- 38 existing survey studies and bibliographies;
- A review of nearly 90 journals and periodicals published during the full period;
- An analysis of over 100 books;
- Mailings to about 100 QWL practitioners; and
- Various sources: e.g., *Dissertation Abstracts* and proceedings of professional meetings

Two goals dominate this search. It seeks comprehensive coverage of English language, published QWL sources over an extended period, so as to minimize the bias of any early misplaced exuberance. In addition, this search also targets representation of unpublished cases—dissertations, in-house consulting
reports, and so on. This seeks to reduce possible bias due to any tendency to publish “positive” results.

The results of the panel of OWL applications will be considered in three ways. Immediately, aggregate success rates will be determined. Later, and especially in Chapter 11, the aggregate panel of cases will be broken down to see if specific situational features of applications have distinct success rates. Moreover, later analysis will focus on the kinds of QWL applications, which Exhibit 11.1 later details.

The panel assignments basically recognize that many practitioners work in both QWL and OD, which reflects basic value agreement. Witness the commitment in both traditions to a “dual focus,” which in Goodman’s (1979) terms requires

\[ \ldots \text{improving both productivity and quality of work life dimensions.} \]

The latter includes physical and psychological safety at work as well as opportunities to learn new skills, to accept great responsibility, and for more satisfaction from work, and so on. The focus is not on either productivity or on the psychological outcomes of work, but rather on jointly improving both of these dimensions (pp. 7–8)

Nonetheless, the two traditions have their own developmental histories (e.g., Skelley, 1989). For example, one can distinguish these different central tendencies, among others, in QWL and OD:

<table>
<thead>
<tr>
<th>QWL applications</th>
<th>OD applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tend to emphasize</td>
<td>Tend to emphasize</td>
</tr>
<tr>
<td>Shop floor and operations</td>
<td>Executive suite and managerial realms</td>
</tr>
<tr>
<td>Unionized or unionizable grades</td>
<td>Managerial or exempt grades in policy networks</td>
</tr>
<tr>
<td>Jobs and work systems</td>
<td>Interpersonal and group relationships</td>
</tr>
</tbody>
</table>

\textbf{Outcome Measures.} Two classes of outcomes are assessed here: hard-criteria and global. Inclusion of a case in the panel \textit{requires} that it measures hard or objective outcomes—waste, absenteeism, and so on. But all included studies also provide soft or self-report outcomes, and these are factored into a “global” assessment, by conventions that will soon be detailed.

\textbf{Measurement Conventions and Reliabilities.} Each QWL case receives two ratings as to outcomes: one for hard criteria and a second “global” estimate
that encompasses hard as well as soft criteria. Four classes of ratings are possible for each of the two estimates: (1) highly positive and intended effects, (2) definite balance of positive and intended effects, (3) no appreciable effects, and (4) negative effects. Assignments to categories 1 and 2 require that many T2 vs. T1 differences—and especially on major target variables—fall in the initial direction, with special relevance to differences that attain the .05 level of significance on statistical tests or, alternatively, that show 10 percent or greater intended changes in T2 vs T1 comparisons. Category 1 assignments meet these requirements for two of every three variables measures, in general; and Category 2 assignments average a minimum of 20–25 percent large and intended differences, plus little or no evidence of negative effects, or changes in an unintended direction. In turn, category 4 estimates of effects include applications meeting one or both of the following two requirements: when a quarter or more of the target variables change in unintended ways, even if the differences are small; or when even a few changes fall in an unintended direction and attain statistical significance or are greater than 10 percent when no statistical tests get reported. All other applications are assigned to Category 3—no appreciable effects.

A high degree of confidence seems appropriate about the interobserver consistency in the scoring of outcomes as well as situational features. Three raters were employed, with only one of them (Sun) having a clear sense of the study’s focus on the situational variables related to QWL success.

Training and discussion resulted in substantial agreement between raters, as variously estimated. For the situational features to be emphasized in Chapter 11, interrater reliabilities range from 92 to 98 percent, with a mean of 94 percent. Reliability on the outcome assessments relied on a random selection of 120 cases from the total of 314. The mean reliability approximated 94 percent, with the paired-comparisons of the three raters falling in an attractive range of agreement—from 92 to 96 percent agreement.

Cases that were in dispute before calculating these initial inter-rater reliabilities, to continue the common convention throughout this book, and that remained so after discussion, are assigned to the lowest rating in contention. This is a conservative convention, patently. In an elemental sense, consequently, effective reliability is always 100 percent.

Findings of QWL Success

Most findings related to the QWL panel appear in Chapter 11, and coverage here is limited. Table 2.4 presents data about success rates. Note that two estimates are given: for an Original Panel of N = 231, for which search extended only to late 1988 (Sun, 1990; Golembiewski and Sun, 1990a); and for an Augmented Panel, which extended search for applications through 1996). Clearly, the findings are quite similar for the original and expanded panels, and quite positive.
Indeed, the summary data in Table 2.4 not only are impressive but leave little room for improvement. What accounts for that? No one can say, definitely. But QWL applications often take place at operational levels, and often involve structural designs of tasks and work. These may constitute more definite, and more easily controllable, targets for change than concern the typical OD project. No great concern about that important point is attempted here.

**ANOTHER CONCEPTUAL CONTEXT FOR IRONY I**

Here, again, note only that the irony of ironies operates with particular force. Success rates are formidable in the context of tentativeness if not pessimism about attaining the intended effects of planned change applications.

It would probably be presumptuous to attempt an explanation of why OD success rates—contrary to much commentary—are substantial both in North American as well as in a range of global settings. But we here entertain a suggestion or two.

Basically, observers are myopic when they see planned change as interaction-only, as noted. For example, although much OD practice has been interaction-centered, in a large majority of applications, OD conceptually has a strong foundation in a range of other types of designs. For convenience, we can identify two such additional categories: structure, and polices and procedures. QWL is often restricted, but in this case to structure as well as policies and procedures.

If planned change is seen as a three-legged stool, as it were—as I advised some time ago (Golembiewski, 1989)—the findings reviewed in Chapter 1 and directly above constitute no real problem. Directly, the values of OD and QWL
Success Rates in Global Settings

...can be applied to most (or even all) work settings via interaction, structure, or policies and procedures—alone or in combination.

Exhibit 2.1 provides an in-depth illustration of this crucial point. It shows how some kinds of OD designs, but not all, have a good fit with cultural conditions that are normally seen as recalcitrant to OD applications. These cultural conditions derive from Hofstede’s (1980a) leading work, although he does not put those cultural dimension to the same use as does Exhibit 2.1 (e.g., Hofstede, 1980b).

Details are available elsewhere about both Hofstede’s cultural dimension (Hofstede 1980a, 1980b) as well as on the several OD designs (Golembiewski, 1989), and hence are omitted here. The direct point is the same. To illustrate, interaction-centered designs might often be at odds with culture featuring high power distance, but they also might be exactly what such settings need! Again, for cultures scoring low on uncertainty avoidance and power distance, interaction-centered designs would tend to be good fits.

**EXHIBIT 2.1 Good Fit of OD Designs to a Range of Cultural Combinations**

<table>
<thead>
<tr>
<th>Cultural conditions*</th>
<th>Good fit OD designs †</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. High individualism</td>
<td>Job enrichment</td>
</tr>
<tr>
<td></td>
<td>Management by objectives</td>
</tr>
<tr>
<td></td>
<td>Task-oriented interventions</td>
</tr>
<tr>
<td>2. Low uncertainty avoidance, low power distance</td>
<td>Group confrontations</td>
</tr>
<tr>
<td></td>
<td>Interpersonally oriented team-building</td>
</tr>
<tr>
<td></td>
<td>T-Groups for intact work team</td>
</tr>
<tr>
<td></td>
<td>Showing 3-D Images</td>
</tr>
<tr>
<td>3. High uncertainty avoidance, medium to high power distance</td>
<td>Policy changes like flexible work hours</td>
</tr>
<tr>
<td></td>
<td>Role diagnosis</td>
</tr>
<tr>
<td></td>
<td>Role negotiation</td>
</tr>
<tr>
<td></td>
<td>Broad range structural changes</td>
</tr>
<tr>
<td></td>
<td>Survey/with aggregate feedback</td>
</tr>
<tr>
<td></td>
<td>Third-party confrontations</td>
</tr>
<tr>
<td>4. High individualism, low-masculinity, high power distance</td>
<td>1-to-1 consultation</td>
</tr>
<tr>
<td>5. Low individualism, low power distance</td>
<td>Demotion design</td>
</tr>
<tr>
<td></td>
<td>Role negotiation</td>
</tr>
<tr>
<td></td>
<td>High-intensity interpersonally oriented team-building</td>
</tr>
<tr>
<td></td>
<td>T-Groups for intact work teams</td>
</tr>
</tbody>
</table>

* See Hofstede, 1980a, 1980b.
† See Golembiewski, 1990, p. 47.
A WORKING SUMMARY: WHY SUCCESS CAN LEAD TO GREATER AND GROWING SUCCESS

Earlier discussion tentatively speculated on why such results as those above have not occasioned greater public whoops of celebration, but we can be bolder about why acknowledging the high success rates can increase success, as paradoxical as that may appear initially.

Consider only four perspectives on the major point. First, these favorable success rates do not mean that all planned change problems have been recognized, let alone solved to such a degree that designs and perspectives can be applied following a cookbook approach. Positively, rather, these results imply that whatever exists in the organizational world can be accommodated, most of the time, by the kind of OWL or OD intervenors who research and write up their experiences, and with an attractive batting average.

What are the helpful implications of this first positive interpretation? Basically, the success-rate literature suggests a kind of “floor” for OD practice. If an intervenor falls consistently below that floor, the major prescriptions seem direct. An intervenor should continually do something to improve his or her diagnosis or prescription, or try another line of work. And both of these prescriptions should contribute to enhanced success rates, obviously. The general effect will be heightened and accelerated when—notice I do not write if—various OD monitoring institutions begin using the success rates to set expectations among clients and intervenors alike. I do not refer to policing activities, either exclusively or even primarily. The focus should be on education—exclusively at first, for a long period, and always as a priority thrust. The institutions include accrediting agencies like the now-defunct Certified Consultants International, mass-membership collectivities like the Organization Development Network, and registering agencies such as the Organization Development Institute.

Second, the present results do not imply that public-sector OD is easier than “in business,” more difficult, or the same. The results here only imply that whatever the constellation of unique constraints existing in various organizations, whether governmental or business, they can be accommodated to by those who have written up their experiences of appropriate OD interventions.

This is no cute conclusion. For example, we know quite a bit about how to develop such accommodations to the specific characteristics of businesses or agencies in the public sector. This is not the place, however, to detail that experience and theory, which has been accomplished elsewhere (e.g., Golembiewski 1985, 231–366). Chapter 11 also will help. Overall, in general, to the degree that high success rates motivate reliance on such context-relevant guidelines, patently, so also will OD success probably increase.

Third, the success-rate literature—whether approximately correct, or even
if wildly off the mark—can improve OD success, if that literature is responded to appropriately. True believers or unreconstructed critics cannot provide such responses.

To provide some specifics on this third perspective, the available studies of OD success rates require further testing, and that will improve OD whether the tests are positive or negative. For example, greater specificity will be required of more finely tuned analyses than those relied on here. In general, future comparative analyses will require a more precise typology of interventions, as well as a more complex differentiation of the hosts or targets for specifically differentiated interventions. This consciousness-raising has been begun elsewhere (e.g., Bowers, Franklin, and Pecorella, 1975) as well as at numerous points throughout this book (see especially Chapters 3 and 7). However, much remains to be done. In the present case, interventions are distinguished only in gross terms, and targets/host are differentiated only as “public” and “business.” A more satisfactory typology of systems hosting OD efforts will eventually take into explicit account the full range of differences and similarities undifferentiated by in the short-hand “public versus private” distinction; and it seems just as clear that this typology also will need to encompass those equally significant differences and similarities within public and business sectors. Derivatively, success rates should increase, even though that boost will be bounded by the already substantial success rates. Nonetheless, the testing of the success-rate literature will help improve OD diagnosis and intervention, in the senses illustrated here as well as in others beyond the scope of this chapter. It is that simple, and that demanding.

Fourth, and finally only for present goals, this chapter may be faulted by a major contaminant. As some observers emphasize (Mirvis and Berg, 1977), published materials may be biased toward reporting positive results, at least in the short run. If this bias characterizes the present data base, that would obviously account for some part of the high success rate. Our procedures provide only partial protection against such a bias. Note the effort to solicit unpublished materials—consultant reports, in-house memos, theses and dissertations—that seeks to counterbalance any bias toward positive results in published work. Presumably, unpublished material would be less contaminated in this regard.

Again, we will not know until several investigators look intensively and with skill; the success-rate literature encourages—even demands—that kind of looking, by both critics and proponents of OD. In this fourth sense, then, the success-rate literature can serve to increase OD success, if approached in a concerted and balanced way. Chapter 3 helps us gauge the required degree of tentatives by estimating the “positive-response bias” in QWL evaluations. Readers will have to await the details but, broadly, only minor adjustments seem necessary to the highlighting here of “substantial, if not robust” success rates.
**APPENDIX 2.1** Distribution of Political Jurisdictions in OD Panels, Less than $5000 GNP Per Capita, in U.S. Dollars, 1980

<table>
<thead>
<tr>
<th>Ranking per capita GNP</th>
<th>Country</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upper Middle—Under $5,000</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Ireland</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>Greece</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>Israel</td>
<td>14</td>
</tr>
<tr>
<td>44</td>
<td>Singapore</td>
<td>1</td>
</tr>
<tr>
<td>47</td>
<td>Netherlands Antilles</td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>Hong Kong</td>
<td>3</td>
</tr>
<tr>
<td>50</td>
<td>Bulgaria</td>
<td>2</td>
</tr>
<tr>
<td>51</td>
<td>Poland</td>
<td>8</td>
</tr>
<tr>
<td>55</td>
<td>Venezuela</td>
<td>3</td>
</tr>
<tr>
<td>61</td>
<td>Puerto Rico</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>Argentina</td>
<td>1</td>
</tr>
<tr>
<td>69</td>
<td>South Africa</td>
<td>4</td>
</tr>
<tr>
<td>70</td>
<td>Chile</td>
<td>2</td>
</tr>
<tr>
<td>71</td>
<td>Mexico</td>
<td>4</td>
</tr>
<tr>
<td>72</td>
<td>Brazil</td>
<td>2</td>
</tr>
<tr>
<td>75</td>
<td>Fiji</td>
<td>2</td>
</tr>
<tr>
<td>80</td>
<td>South Korea</td>
<td>9</td>
</tr>
<tr>
<td>80</td>
<td>Turkey</td>
<td>2</td>
</tr>
<tr>
<td><strong>Lower Middle—Under $1,310</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>Ecuador</td>
<td>3</td>
</tr>
<tr>
<td>92</td>
<td>Dominican Republic</td>
<td>1</td>
</tr>
<tr>
<td>93</td>
<td>Guatemala</td>
<td>1</td>
</tr>
<tr>
<td>97</td>
<td>Jamaica</td>
<td>3</td>
</tr>
<tr>
<td>98</td>
<td>Nigeria</td>
<td>4</td>
</tr>
<tr>
<td>100</td>
<td>Peru</td>
<td>5</td>
</tr>
<tr>
<td>109</td>
<td>Philippines</td>
<td>5</td>
</tr>
<tr>
<td>115</td>
<td>Zimbabwe</td>
<td>2</td>
</tr>
<tr>
<td>116</td>
<td>Dominica</td>
<td>1</td>
</tr>
<tr>
<td>118</td>
<td>Egypt</td>
<td>4</td>
</tr>
<tr>
<td>124</td>
<td>Liberia</td>
<td>1</td>
</tr>
<tr>
<td>128</td>
<td>Djibouti</td>
<td>2</td>
</tr>
<tr>
<td>130</td>
<td>Sudan</td>
<td>1</td>
</tr>
<tr>
<td>133</td>
<td>Senegal</td>
<td>2</td>
</tr>
<tr>
<td>134</td>
<td>Ghana</td>
<td>1</td>
</tr>
<tr>
<td>135</td>
<td>Indonesia</td>
<td>1</td>
</tr>
<tr>
<td>136</td>
<td>Kenya</td>
<td>3</td>
</tr>
<tr>
<td>147</td>
<td>Pakistan</td>
<td>1</td>
</tr>
<tr>
<td>148</td>
<td>China</td>
<td>1</td>
</tr>
<tr>
<td>150</td>
<td>Uganda</td>
<td>2</td>
</tr>
</tbody>
</table>
## APPENDIX 2.1  Continued

<table>
<thead>
<tr>
<th>Ranking per capita GNP*</th>
<th>Country</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lower Middle—Under $1,310</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>Sri Lanka</td>
<td>1</td>
</tr>
<tr>
<td>156</td>
<td>Tanzania</td>
<td>7</td>
</tr>
<tr>
<td>157</td>
<td>Gambia</td>
<td>1</td>
</tr>
<tr>
<td>158</td>
<td>India</td>
<td>70</td>
</tr>
<tr>
<td>160</td>
<td>Zaire</td>
<td>1</td>
</tr>
<tr>
<td>163</td>
<td>Senegal</td>
<td>1</td>
</tr>
<tr>
<td><strong>Bottom 10—$200 and Under</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td>Ethiopia</td>
<td>1</td>
</tr>
<tr>
<td>168</td>
<td>Nepal</td>
<td>1</td>
</tr>
<tr>
<td>169</td>
<td>Bangladesh</td>
<td>1</td>
</tr>
<tr>
<td><strong>No GNP Ranking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Caribbean Countries”</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>“Spanish speaking”</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Iran</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Jordan</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Lithuania</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Mongolia</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Morocco</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>“Pacific Island”</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Somalia</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Western Australia</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>“Developing Country” (Anonymous)</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>“old USSR”</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Several evaluations deal with two or more loci.

Source: Based on Kurian (1984, pp. 98–99).

### REFERENCES


Success Rates in Global Settings


Success Rates in Global Settings


Samtel Corporate Training Group (1996b). Supervisory effectiveness programme at Sam-
tel. In U. Jain, U. Pareek, and M. G. Jormon (Eds.), Road to empowerment: HRD experiences with workers and support staff (pp. 270–290). Ahmedabad, India: Academy of Human Resources Development.


One subtle approach to corrupting the success rate estimates of the preceding two chapters involves a kind of sleight-of-mind. Yes, the estimates are substantial, goes this view but that is only because the methodology is shoddy. The better the methodological status of the Organization Development Quality of Working Life (OD) or (QWL) evaluations, one might argue consistently with this position, experimental effects tend to diminish, if not disappear.

One of our Georgia Ph.D.s (Sun, 1988) provides the best-available evidence of a findings bias in a large batch (231 cases)* of QWL applications. He finds a clear tendency, but not a dominant or unqualified tendency, for evaluative studies to show poorer outcomes as methodological rigor increases.

This conclusion about “positive-response bias” has a king-sized status, and deserves corresponding space, although there is no easy way to present the details of Sun’s analysis. Indeed, two indicators must suffice here (Golembiewski and Sun, 1988). Perfect support for the hypothesis of a positive-findings bias in Sun’s study involves 216 predictions, of which 69 percent are realized. In general, then, as rigor increases, outcomes deteriorate for the first three of four classes of outcomes. However, only a small percentage of the variance is accounted for—approximately 7 percent. Moreover, the association is not linear. Even more sub-

* A complete bibliography of these 231 applications is available from Robert T. Golembiewski, University of Georgia, Baldwin Hall, Athens, GA 30602. The complete data set also is available.
tly, the vast bulk of the deviant cases involve the least successful of four classes of outcomes. Contrary to the hypothesis of a positive-findings bias, in those cases the least rigorous studies generate the poorest effects.

A CONCEPTUAL CONTEXT FOR IRONY I

The contribution here is direct, and it is capable of summary in a few words. This study finds statistically significant support for a positive-findings bias hypothesis, but rigor explains less than 7 percent of the variance in outcomes. This implies only modest support for the position that attractive QWL results can be substantially accounted for by a positive-findings bias, which has been prattled about by many careless commentators.

This contribution plays a major role in regard to Irony I. It undercuts the major facile rejection of Chapters 1 and 2, and in an economic way. The easy explanation having been taken out of play, and early in the game, both proponent and opponent can go about making their best possible cases without caterwauling about a positive-findings bias.

Until the contents of this chapter appeared about a decade ago, far more heat than light had been shed. Today’s scholars also seem bent on reliving history, warts and all. Hence, the chapter below is reprinted in essentially unchanged form, in the hope that the contents can do for a today’s readers what it did for yesterday’s. The distinctions here still need to be made, and insistently.

POSITIVE-FINDINGS BIAS IN QWL STUDIES: AN INTRODUCTORY PREVIEW†

Chapters 1 through 3 summarize the recent attention to assessing the efficacy of attempts at planned change in organizations, especially via Organization Development (OD) and Quality of Working Life (QWL). Most of the work focuses on OD (e.g., Margulies, Wright, and Scholl, 1977; Morrison, 1978; Golembiewski, Proehl, and Sink, 1981, 1982; Nicholas, 1982; Terpstra, 1982; Mitchell, 1981; Woodman and Wayne, 1985), and that sub-literature isolates success rates that are substantial and, indeed, formidable. If anything, QWL applications show an even stronger record of intended outcomes than does OD, as over two dozen separate surveys indicate (Sun, 1988; Golembiewski and Sun, 1988, 1989).

This chapter builds on the available work about QWL applications while transcending it in three significant particulars. First, existing QWL surveys usually feature a small number of evaluations of applications, with \( N \) ranging from 30 to 50 (Sun, 1988, chapter 2). This small size precludes sensitive analysis, and results in sparsely-populated or empty cells even in bivariate comparisons.

† An earlier version of this material appeared in Golembiewski and Sun, 1994.
Second, survey studies of both QWL and OD applications almost always focus on narrow outcomes, for example, did applications succeed or fail, overall, and in what proportions? Only rare attention gets directed at the rigor of methodology and design of evaluative studies. Moreover, survey studies typically rely on a single measure of rigor (e.g., Terpstra, 1981; Woodman and Wayne, 1985), on those few occasions when that significant issue gets any attention.

Consequently, the change literature is open to suspicion about the validity of its findings, but appropriate tests in a large population of evaluative studies provide only spotty support for conventional views, as Chapters 1 and 2 demonstrate. Here, we go even further: the focus is on the positive-findings bias, which basically proposes that attractive results are artifacts of poor methods and designs. The essence of the notion is that an inverse relationship exists between the rigor of an evaluation and the success of the intervention being evaluated. As rigor increases, this view implies, studies will reveal fewer attractive effects, or perhaps no change or even negative effects.

Third, QWL evaluative studies usually deal with “soft” outcomes such as self-reports about productivity. This leaves results open to charges of being superficial feel-goodisms, of being “Hawthorne effects,” and so on (Bass, 1983). Survey studies that overcome one of these three limitations typically fall victim to the others. For example, one OD survey encompasses over 500 cases but differentiates neither degrees of methodological rigor nor studies providing “hard” data about outcomes from that majority offering only “soft” data (Golembiewski, Proehl, and Sink, 1981, 1982). Similarly, one OD survey admits only studies providing “hard” data but includes only a handful of studies and neglects most other aspects of methodological rigor (Nicholas, 1982).

This chapter reports on an effort to minimize all three limitations in assessing QWL interventions. It considers unpublished as well as published evaluative studies while testing whether or not attractive success rates are the products of faulty methodology and design. The present author knows of no similar survey of QWL applications that seeks to learn from as well as to augment the sparse and inconclusive attention devoted to the positive-findings bias in OD (e.g., Bass, 1983). Some observers report a positive-findings bias in OD applications (Terpstra, 1981); other studies fail to replicate this finding (Bullock and Svyanetek, 1983, 1985); and still others suspect a positive-findings bias in one class of OD interventions but not in other classes (Woodman and Wayne, 1985). This condition patently demands improvement for all except those who have a monumental capacity for tolerating ambiguity.

METHODS

Five sections carry the burden of the present approach to minimizing the three common limitations of QWL evaluative studies. These sections, in turn, briefly introduce the pool of QWL studies, detail three estimates of rigor, sketch ways
of estimating the success of individual QWL applications, provide estimates of
the reliability of scoring, and outline the present analytic approach.

Sample of QWL Studies

A multipronged search sought a comprehensive sample of QWL evaluative stud-
ies, both published and unpublished, during the interval 1965–1987. All accept-
able studies provide “hard” measures of outcomes—objectively measured output,
turnover, absenteeism, cost of raw materials, and so on. The search involved:

discovery of over three dozen bibliographies;

examination of 90 English-language periodicals;

review of *Dissertation Abstracts*, as well as of the proceedings of several

professional associations—Academy of Management, American Psycho-

logical Association; and

a mail solicitation of about 100 QWL practitioners.

The search isolates 231 studies, mostly from published sources but with 13
percent from unpublished sources such as in-house memos or consultant reports.
Although “hard” measures of outcomes are required for inclusion in the sample,
all 231 studies also report “soft” outcomes—self-reports about satisfaction, coop-
eration, and so on. Usefully, this permits testing the casual common wisdom
about the alleged inferiority of self-reports.

Two other features of the sample of QWL evaluative studies deserve note.
This sample includes cases from numerous collections—other surveys, bibliogra-
phies, and so on—and adds to their number by over a third. Moreover, domain
issues are not crucial here. Conceptually, most observers define OD broadly
enough so that it subsumes QWL as a major class of interventions. In practice,
OD and QWL applications share values centering around participation and
involvement, but the applications often differ in a range of particulars, to risk a
bit of repetition—QWL is more likely to deal with unionized employees and
OD with management, QWL emphasizes structure, OD has a dominant focus on
interaction, and so on (Shelley, 1989). Except in a few cases of egregious misla-
beling, the author’s description of a case as “QWL” is accepted.

More specifically, the present sample encompasses nearly twenty distinct
kinds of interventions. They include goal-setting; structuring work as well as
social relations in autonomous groups; work redesign and, especially, job enrich-
ment; Quality Circles; and other applications that combine two or more distinct
interventions of the kinds illustrated.

Estimates of Rigor

Three estimates of rigor are employed, so as to avoid the common reliance on
a single measure. An overview of how these measures assess rigor is helpful,
and can be brief.
Positive-Response Bias?

Terpstra’s M/D Score

This measure of methodology/design rigor is employed in an early analysis of positive-findings bias in OD research (Terpstra, 1981). Six dimensions are assigned a score (1 or 0), depending on whether or not a specific study is considered “rigorous” on each dimension. A study with a perfect score of 6: (1) utilizes an acceptable sampling strategy such as a full census, (2) has a sample size greater than 30, (3) employs a control or comparison group, (4) uses random assignment to treatment or control, (5) provides at least one pre-test and one post-test estimate of effects, and (6) achieves observed differences $p < .05$ on statistical tests of differences.

Woodman and Wayne’s M/D Score

The developers of this second measure propose to improve on Terpstra’s version by adding three criteria to his list of six, basically (Woodman and Wayne, 1985). Three additional points are assigned to studies that (7) show no reliabilities <0.60 and provide some evidence of validity, (8) include “hard” data relevant to objective criteria, and (9) employ an appropriate multivariate analytic procedure.

Morrison’s I/E Validity Score

Morrison takes a different tack. She focuses on Campbell and Stanley’s (1963) detailed tests of how a research design can eliminate threats to validity. Morrison’s (1978) I/E score ranges from 0 to 12, with the higher scores indicating greater effectiveness in eliminating threats to validity. That is, rigorous studies seek to minimize or eliminate these four threats to external validity: interaction of test and treatment; interaction of selection and treatment; reactive features that limit generalizability of results; and multiple-treatment interferences. In addition, a rigorous evaluative study also deals with these eight threats to internal validity:

- Changes due to factors other than the intervention;
- Maturation of subjects or groups;
- Testing effects;
- Instrument effects;
- Regression effects;
- Selection biases;
- Mortality of subjects; and
- Interaction effects of two or more of the seven threats above.

Although the three measures of “rigor” tap similar dimensions, each appears in the analysis below. Specifically, Terpstra’s M/D score and that of Woodman and Wayne correlate .91; and Morrison’s I/E Validity score correlates .60 and .67 with the two M/D measures, respectively. These coefficients imply some redundancy in subsequent ANOVA analysis, but given the lack of research in
this area, better more than less. Given substantially the same patterns of findings for the three measures of rigor, subsequent analyses will have strong justification for using either a composite rigor score or MANOVA.

**Conventions for Estimates of QWL Success Rates**

The criterion for inclusion in the present pool is that a study assesses QWL outcomes with “hard” or objective data, but all studies also provide “soft” data about outcomes. This analysis scores 16 objective outcomes—e.g., quantity of output, various costs such as those for materials and labor, and personnel turnover. Scoring also differentiates 18 self-report outcomes—for example, job involvement, various facets of satisfaction with work, and organizational commitment.

Here, for each QWL study, this analysis estimates both hard-criteria and global outcomes, with the latter combining objective and self-report data. In both cases, four categories of outcomes are distinguished:

I. Definite balance of highly positive and intended effects
II. Definite balance of positive and intended effects
III. No appreciable effects
IV. Negative effects

In general, this study hypothesizes that QWL applications will have a consistent set of intended effects. For example, following a QWL application, productivity should increase, employees will report greater satisfaction, and job commitment will grow.

A clean sense of the present scoring conventions requires a bit more detail. Thus, in $T_2$ versus $T_1$ changes, “highly positive and intended effects” include statistically significant changes ($P < .05$) or those with a magnitude of 10 percent or greater in the expected direction where no statistical tests are employed. Category I assignments require that more than half of all pre- vs. post-test comparisons meet one or both of these standards, with most other changes also falling in the expected direction. Category II assignments require that most comparisons are in the intended direction, and also include a substantial proportion of statistically significant changes. Small and random changes dominate in Category III. Category IV includes all applications with more than a sprinkling of unexpected effects and all cases that in a contrary direction attain statistical significance.

Both hard-criteria and global outcomes are governed by the same conventions for successful assignments. Global outcomes are based on all self-report data in a particular evaluation plus all objective variables. Hard-criteria outcomes only involve the latter variables.

**Inter-Observer Reliabilities**

The evidence implies that little variance can be accounted for by differences between coders. Three raters are utilized, and agreement between pairs of observ-
ers ranges from the low- to mid-.90s for the several outcome and rigor codings. The measure of agreement assigns either a 1 or 0 to each pair of codes, depending on whether they agree or disagree. The total points assigned are then divided by the total number of pairs, and multiplied by 100 to generate a percentage estimate of agreement. There are no missing data for either rigor or outcomes.

A conservative convention applies to all codes-in-disagreement, if discussion fails to resolve them after the reliabilities are calculated. Each such case is assigned the lower/lowest score in contention. For outcomes, the typical case of disagreement involves a judgment about whether a QWL application merits success category I or II. The final assignment is a II, if discussion does not lead to consensus. Similarly, disagreement about any discussion results in a 0 code. As noted, each rigor component is scored 1 or 0 depending on whether its treatment is seen as contributing to validity.

In sum, the codings for outcome and rigor involve many judgment calls, and this project reflects a mixed but generally positive record of dealing with them, as three points imply. First, the inter-rater reliabilities are substantial, as estimated before discussion leading to a final judgment. This implies effective training of coders.

Second, conservative conventions reinforced by training govern judgments about both outcomes and the multidimensional scales used to assess rigor. This no doubt helps to account for the substantial reliabilities before discussion, but it comes at the cost of introducing a systemic bias that may effect validity. The point applies with special force to Morrison’s scale, which requires difficult judgments as to whether particular threats to validity are addressed in any evaluative study. Here, as elsewhere, individual cases are assigned a lower code when doubt exists.

Third, this study does a substantial if incomplete job of meeting a comprehensive list of fourteen criteria for surveys of this kind (Bullock and Svyantek, 1985, pp. 114–115). A well-designed survey

1. Uses a theoretic model;
2. Identifies its study domain precisely;
3. Includes all publicly available studies;
4. Avoids selecting studies in terms of rigor, etc.;
5. Publishes or makes available the pool of studies;
6. Selects and codes variables on theoretic grounds;
7. Provides details about the coding scheme and resolution of problems in its applications;
8. Uses multiple raters and assesses inter-rater reliability;
9. Reports on all variables analyzed in order to avoid problems with chance relationships in a subset;
10. Publishes or makes available the full data set;
11. Considers alternative explanations for findings;
12. Limits generalization of results to the specified domain;
13. Reports study characteristics to shed light on domain analyzed; and
14. Reports studies in sufficient detail to permit direct replication.

Reflecting the limitations of the journal article genre, numbers 7 and 14 are the least substantially met criteria in this report.

A fourth issue about coding constitutes a possibly severe constraint on the findings: both rigor and outcome scores are assessed by the same raters. Independence could have been assured, as by using two separate panels of raters, each blind to the study’s basic purpose (Woodman and Wayne, 1985). Here, the major defenses against contamination are the guidelines for estimating outcomes and training.

Analytic Procedures

In sum, this study employs twelve separate tests of the association of rigor with QWL outcomes. Since rigor is estimated in three ways, and since both hard criteria as well as global assessments of outcomes are available, this requires six tests of the positive-findings bias. In addition, because of the small number of cases rated as having outcomes 3 and 4, this chapter also will test for associations of rigor with three categories of outcomes—1, 2, and 3 plus 4. The infrequent 3 and 4 assignments have several possible interpretations, and combining the two categories provides guidance in evaluating those interpretations while it also tests for effects of small sub-sample size. This useful exercise adds six tests for a positive-findings bias.

When one-way analysis of variance (ANOVA) isolates statistically significant variance in rigor and outcomes scores, analysis will be supplemented in two ways. The statistical significance of all possible paired-comparisons will be assessed via the Least Significant Difference test, or LSD, as modified for unequal sub-sample sizes. The direction of all possible paired-comparisons also will be assessed for consistency or contrariness with the positive-findings bias.

For this chapter, evidence supporting a positive-findings bias requires that the higher the rigor of a QWL evaluative study, the poorer its associated outcomes. For each of the twelve tests detailed above, then, perfect support for the hypothesis requires three conditions: that all ANOVA tests attain $<.05$; that all possible paired-comparisons show that rigor is inversely associated with the favorableness of QWL outcomes; and that all paired-all comparisons in the intended direction attain $P < .05$, but no pairs in the contrary direction do so.

RESULTS

Review of the results proceeds on two tracks. First, the form of analysis and of the results will be illustrated by focusing on one of twelve individual analyses detailed above. Second, the results of all 12 analyses will be summarized.
TABLE 3.1  Methodological Rigor and Hard-Criteria Outcomes

| Outcomes of QWL interventions | $N =$ | Means, Rigor scores |  |
|-------------------------------|-------|---------------------|  |
| I. Highly positive and intended effects | 160 | Terpstra's M/D scores | Woodman and Wayne's M/D score | Morrison's I/E score |
| II. Definite balance of positive and intended effects | 50 | 2.98 | 4.00 | 5.12 |
| III. No appreciable effects | 13 | 3.62 | 5.16 | 7.62 |
| IV. Negative effects | 8 | 2.75 | 3.75 | 2.75 |
| $F =$ | 3.63 | 6.80 | 5.81 |
| $P =$ | .0137 | .0002 | .0008 |
| $\eta^2 =$ | .05 | .08 | .07 |

An Illustration: Rigor and Hard-Criteria Assessment

Table 3.1 details analysis of three measures of rigor and hard-criteria assessment, this time rated in terms of four categories of hard criteria outcomes. Four points highlight trends in the data. First, all three F-scores in Table 3.1 indicate nonrandom variance, but $\eta^2$ indicates that a bit less than 6.7 percent of the variance is accounted for, on average.

Second, a third of all paired-comparisons in Table 3.1 achieve statistical significance as well as fall in a direction consistent with the positive-findings bias hypothesis, and the direction of an additional 26 percent of the comparisons is consistent with that hypothesis. Note that six paired-comparisons are possible for each measure of rigor and the four categories of outcomes. Support for the positive-findings bias requires that for each method, the rigor of evaluative studies rated 1 should be less for those rated II, I less than III, and so on. Since there are three measures of rigor, Table 3.1 involves $3^3$, or 18, paired-comparisons.

Third, nearly 40 percent of all paired-comparisons fall in a direction contrary to the hypothesis, and a bit over 11 percent attain statistical significance. The contrary cases all involve outcome category IV, or negative effects. All three measures of rigor tend to rise successively for outcomes I, II, and III, and then fall for outcome IV.

Fourth, the pattern for each of the three measures of rigor is quite similar. Only minor variations exist.

Summary: All Measures of Rigor for All Outcomes

The pattern for the illustration above—for four categories of hard-criteria assessments of QWL outcomes—also characterizes the three other sets of associations.
### Table 3.2  Summary, Tests for Positive-Findings Bias

<table>
<thead>
<tr>
<th>Paired-comparisons</th>
<th>ANOVA, outcomes and Rigor, $P &lt; .05$</th>
<th>In consistent direction and statistically significant</th>
<th>In consistent direction</th>
<th>In contrary direction</th>
<th>In contrary direction and statistically significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Global outcomes, 4 outcome categories</td>
<td>3 of 3</td>
<td>5 of 18</td>
<td>11 of 18</td>
<td>7 of 18</td>
<td>0 of 18</td>
</tr>
<tr>
<td>B. Hard-criteria outcomes, 4 outcome categories</td>
<td>3 of 3</td>
<td>6 of 18</td>
<td>11 of 18</td>
<td>7 of 18</td>
<td>2 of 18</td>
</tr>
<tr>
<td>C. Global outcomes, 3 outcome categories</td>
<td>3 of 3</td>
<td>5 of 9</td>
<td>7 of 9</td>
<td>2 of 9</td>
<td>0 of 9</td>
</tr>
<tr>
<td>D. Hard criteria outcomes, 3 outcome categories</td>
<td>2 of 3</td>
<td>6 of 9</td>
<td>9 of 9</td>
<td>0 of 9</td>
<td>0 of 9</td>
</tr>
<tr>
<td>Totals</td>
<td>11 of 12</td>
<td>19 of 54</td>
<td>38 of 54</td>
<td>16 of 54</td>
<td>2 of 54</td>
</tr>
<tr>
<td>Means, %</td>
<td>91.7%</td>
<td>35.2%</td>
<td>70.4%</td>
<td>29.6%</td>
<td>3.7%</td>
</tr>
</tbody>
</table>
summarized in Table 3.2 A–C. Table 3.2D summarizes the results discussed later in connection with Table 3.2E.

Five points characterize the overall pattern in Table 3.2. First, nearly 92 percent of the ANOVA for overall rigor and outcomes achieve statistical significance. The appropriate tables underlying the summary in Table 3.2 are not reprinted to conserve space, but one of them shows that the single exception to our first characterization approaches statistical significance ($F = 2.37, P = .09$).

Second, paramountly, a noteworthy proportion of the differences between the paired-comparisons are quite robust as well as in the direction supporting a positive-findings bias. Specifically, over 35 percent of these cases attain statistically significant differences as well as are directionally consistent with the hypothesis, whereas chance allows 5.0 percent of the cases to be nonrandom.

Third, an additional 70.4 percent of the paired-comparisons are consistent with a positive-findings bias, although they do not attain $P < .05$.

Fourth, outcome/rigor pairs inconsistent with the hypothesis of a positive-findings bias are in a clear minority and, moreover, such differences almost never achieve statistical significance. Specifically, over 29 percent of all paired-comparisons fall in a direction contrary to the hypothesis, and less than 4 percent achieve the 0.05 level.

As with Table 3.1, all of the cases in Table 3.2 falling in a contrary direction—16, to be exact—involves QWL outcomes rated 4. Only two of those sixteen attain statistical significance.

Fifth, $\eta^2$ averages .067 for the 11 statistically significant ANOVA runs. This is only marginally higher than for the illustration in Table 3.1.

DISCUSSION

This analysis permits six major conclusions that deserve highlighting. These conclusions suggest how this analysis transcends earlier work, how it supports the hypothesis of a positive-findings bias, how that support needs to be interpreted carefully, and how the present analysis can be extended to encompass issues beyond the present scope.

Although not ideal, to begin, the present test of the positive-findings bias has a variety of advantages over earlier research. The population of cases is much larger; and can make a stronger claim to representativeness. In addition, three accepted measures of rigor are used, and all generate similar patterns of results. Moreover, hard-criteria effects are required of all cases admitted to analysis, which precludes criticism of the results as based on “mere self-reports.” As noted, global outcomes include “soft” data to provide a check on the hard-criteria assessment.

Consequently, special weight seems appropriate for the present conclusion that the positive-findings bias accounts for only modest variance (6.7 percent) in
QWL outcomes. In contrast, prior evidence from OD studies is inconclusive. Terpstra (1981) finds evidence of an inverse relationship between rigor and outcomes in OD, but Bullock and Svyantek (1983) do not, although each study draws its population from the same journal during the same interval. Woodman and Wayne (1985) attempt a similar test, and it suggests that a positive-findings bias may exist for some classes of OD interventions but not others.

In addition, this analysis provides perspective on the issue of whether or not “positive results” have a better chance of publication, a common point in critical arguments. Any such effect in this case may be diluted by the long period over which cases in our panel were published, but analysis shows the success rates of published versus unpublished studies differ only randomly (chi-square = 5.58, \( P = 0.13 \)). To illustrate, nearly 85 percent of the twenty-six unpublished cases fall in “hard” outcome categories I and II. This is a lower score than for the published cases (91.7 percent), but not significantly so.

Relatedly, only eight QWL applications have a Category IV outcome—that is, negative effects. Are these cases somehow similar in other regards? Individual inspection suggests no obvious similarities in date, worksite, intervention, and so on. But the N is too small to justify detailed analysis and interpretation.

A final perspective on tentativeness about positive-findings bias is worthy of attention. A separate test of that explanation of success rates focuses on public-sector worksites (Golembiewski and Sun, 1992), and it permits less confidence than the discussions above. The smallish public population also encourages interpretive caution. Whatever else, however, positive-findings bias does not dominate in the analyses undertaken so far.

So evidence supporting a positive-findings bias in QWL studies not only must be appropriately tethered, but two additional points apply. First, conclusions apply least to outcomes rated IV—that is, negative effects. This may reflect attenuated distributions of methodological rigor, or of outcomes, or of both. Alternatively, it may simply be that negative outcomes plainly advertise themselves, whether or not research designs are rigorous. Moreover, although this is a bit of a stretch, a kind of incompetence identity may even exist. Poor evaluative designs may occur along with careless or inept interventions, which places the issue in the specific implementor/evaluator rather than in QWL.

Second, the size of the present QWL sample permits perspective on the consistency of these findings. Consider only two possibilities—consistency between the several classes of QWL interventions and consistency over time.

As for QWL classes, this population encompasses seventeen distinct varieties of QWL interventions, which can be aggregated into four broad classes. Consistency of the present findings within each of these four broad categories will provide a powerful test of QWL. The small populations underlying most other surveys of planned change preclude such a test, although some observers suspect that differences exist (e.g., Terpstra, 1982, p. 415). A detailed study of trends by
classes is under way, and preliminary indications are that the four classes require no major modifications of present conclusions.

Relatedly, Woodman and Wayne (1985) raise the possibility that misleading results may be generated by including studies from the 1960s as well as the 1980s. Methodological sophistication and practical know-how presumably have grown, and this may influence associations between rigor and outcomes. However, we know that QWL success rates vary only marginally over the 22-year period of observation (Golembiewski and Sun, 1989). In addition, preliminary analysis contrasting early QWL studies with later ones isolates no regular differences from present patterns.

Overall, then, these results do not take all the wind out of the sails of those proposing that attractive QWL outcomes imply puny research designs. At the same time, these results limit the confidence with which such dour proposals can be offered as the gospel, without qualifications.

REFERENCES


Toward Building Work Cultures to Order
Illustrating Replications of Generic Designs

If the special genius of each age gets reflected in distinctive ways of organizing work, historians can easily characterize the 1980s. The preceding age stressed stability and consistency, roughly; the emphasis today is on organizing for change, whether for growth or decline.

The general point overwhelms. John W. Gardner reflects both the certainty and the caution in distinctive fashion. “What may be most in need of innovation is the corporation itself,” he notes. “Perhaps what every corporation (and every other organization) needs is a department of continuous renewal that could view the whole organization as a system in need of continuing innovation” (Gardner, 1965, p. 1).

So these times qualify as The Age of Intended Transitions, and our success (or failure) will be powerfully influenced by our ability to induce cultural patterns that facilitate transitions. Ironically, much of the challenge derives from our past and persisting success in creating cultures for stability.

This trio of chapters—4 through 6—draws attention to an irony that stands in relative neglect: the substantial success of OD applications has been achieved without major attention to replication, especially as loosely defined. And that definitely is not the way to go about facilitating higher success rates.

Here “replication” has three basic meanings, given a certain verbal flexibility. First, Chapter 4 adopts the garden-variety definition: using the same design or designs for learning in a variety of organizational settings, or at different hierarchical levels in the same organization. Two other conceptual twists on “replication” will be introduced in later chapters.
A CONCEPTUAL NICHE FOR IRONY II

The focus here is on the renewal and expansion of a stodgy municipal bus system into a spritely, integrated urban rail and bus system. Essentially, one set of group-related intervention—usually labeled the confrontation design that relies on the sharing of 3-dimensional Images—is relied on to build knowledge about relationships, as well as skills for behavioral change in a front-load to possible innovation in structure as well as style and procedures. The replication of the same design occurs at four hierarchical levels, with modifications.

Similar results in this chapter imply that the power of the design is such as to overcome or negate the probable variances in different settings that would incline toward camouflaging or destroying effects. The savings are obvious. Skilled resource persons can be developed, incremental improvements can build efficiencies as well as effectiveness, and so on.

Chapter 5 focuses on a second variety of replication. There, different issue arenas become the sites at which the explanatory and learning potentials of OD can be suggested. In this second case, any similarities of effects suggest the power of the OD technology cum values in coping with differences if not diversities. Obviously, either case provides evidence about the robustness of the OD technology cum values for change.

The third chapter in this trio, Chapter 6, illustrates a third usage of replication. Specifically, the technology has the capacity to cast its normative profile through time, as it were. This longitudinal capacity suggests the power of the technology to project its hologram in a series of frames at several points in time that, when viewed appropriately, suggest a motion picture, as contrasted with individual exposures. The goal of such replicatory efforts is to isolate design features that reinforce one another, and cumulatively build on each other.

This first sense of replication is not troublesome, and it may pay major dividends. Perhaps the most basic way to test replication uses the same (or very similar) designs in different contexts. Interveners can gain powerful guidance for their design choices, and hence clarity about what design fits specific contexts and which designs do not.

Basically, Chapter 4 illustrates how several kinds of learning designs based on OD values can help in reinforcing the development of organizations built to the OD order of things. The designs are Three-Dimensional Image development and sharing; guidelines for giving effective feedback; and both in the context of a Confrontation Design highlighting dilemmas that motivate amelioration innovations.

Altogether, the design elements were essentially used at several levels of analysis: the MARTA Board; the General Manager; the Assistant General Managers; and the Director reporting to the AGMS. Similar effects, in effect, would support the reliability of OD theory and several specific designs. In various co-
Illustrating Replications

Combinations, these three design elements may be thought of contributing to a regenerative system of interaction, whereas most existing systems tend toward the degenerative. Many illustrative applications are available. In rough chronological order, the earliest applications include Golembiewski and Blumberg, 1967a; Golembiewski and Blumberg, 1967b; Golembiewski and Blumberg, 1968; Golembiewski and Blumberg, 1969; Golembiewski and Blumberg, 1970a; Golembiewski and Blumberg, 1970b). Overall, based on several years of refining of a “confrontation design,” very high success rates were attained, in ways this chapter illustrates. Indeed, the eventual confrontation design was offered to clients with the expectation that 85–90 percent of associations between specific variables would fall in specified directions (e.g., Golembiewski, 1979, vol. 1). Or no fee would be necessary!

CLEAR CULTURAL SIGNALS FACILITATE TRANSITIONS*

The need for planned organizational change gets powerful embodiment in the burgeoning emphasis on Organization Development, or OD: Basically, in its re-educative variety that technology cum values seeks to induce a specific cultural pattern at work that facilitates transitions and helps moderate the associated human shocks. This cultural pattern provides clear signals to organization members, in short. As Winn explains, (1968, p. 5):

The term organization development . . . implies a normative, reeducation strategy intended to affect systems of beliefs, values and attitudes within the organization so that it can adapt better to the accelerated rate of change in technology, in our industrial environment and society in general. It also includes formal organizational restructuring which is frequently initiated, facilitated and reinforced by the normative and behavioral changes.

Changing attitudes or values modifying behavior, and inducing change in structure and policies—these constitute the three core objectives of OD programs. This brief characterization can be reinforced by much available literature, (e.g., Golembiewski, 1995).

This chapter begins an extensive overview of OD consultancy in the public sector to generate a high-performing organization—at MARTA, or the Metropolitan Atlanta Rapid Transit Authority. Despite attention elsewhere, “one more time” here seems appropriate for at least three reasons. The present version is substantially more detailed than earlier reports, both with respect to learning designs as well as consequences. Moreover, the detail should well serve the quite

specialized reader, for whom this book is most intended. Finally, it is now possible to provide long-run counterpoint to this description, as Chapter 6 does.

In brief, this chapter illustrates how an organization can be infused with values, using the same/similar learning design at several loci in the same organization. As it were, such an OD design provides a clear cultural template for the organization of the future.

EASING TRANSITIONS BETWEEN THE SEVERAL MARTA

How do the values underlying OD get embodied in concrete organizational action? And how does their specific cultural pattern get energized? These are tough questions, and this chapter will only begin the task of formulating a useful answer.

The basic organizational dilemma at MARTA start-up may be stated simply: How to structure work and relationships adequate to today’s tasks without exacerbating the very real problems of transitioning to tomorrow’s demands. The point applies with particular relevance to high technology, but few organizations can afford to neglect that dilemma.

Almost from the start, MARTA sought to develop a management system to facilitate the inevitable flow of today into a different organizational tomorrow. Authorized by a 1971 referendum, MARTA sought to gear up quickly to launch a program whose initial estimated cost was $1.3 billion, and whose developmental phase was projected to cover the better part of a decade. That made MARTA “the biggest game in town” in 1971, and also the largest regional public project since the TVA days. Since then, the ravages of double-digit inflation and the vagaries of funding have soon exaggerated the following list of major challenges:

Some Awesome Thoughts

Even in the context of easier times, MARTA faced a future that was both intense and uncertain, with no prospect that things would get easier or more definite. In general, the challenges facing MARTA in 1973 reflected many aspects of both opportunity and danger, as well as of “hurry up and wait.” Some awesome thoughts had to be accomplished, whatever the case, and on tight schedules lest costs get out of hand. Illustratively—at times simultaneously, sometimes in rapid-fire sequence, and frequently in fits and starts—MARTA had to:

- Enlarge an existing bus company using technologies that are well established and straightforward, in general.
- Monitor and coordinate the design and construction of 69 miles of rapid-speed rail lines with associated stations, park-and-ride facilities, etc., involving technologies of sometimes substantial indeterminacy and
complexity, and dealing with volatile issues of taste, aesthetics, politics, community aspirations, and so on.

Develop a broad range of design, development, and operating capabilities as a “lean and mean” central staff that monitored an external consortium of general engineering contractors. This stands in contrast to a staff with a narrower mission as was the case at BART, San Francisco’s rapid transit agency, or in contrast with Washington’s effort making the agency the general engineering contractor.

Aggressively develop a transit system when funding was highly dependent on grants from federal agencies whose appropriations were uncertain.

Develop fluid working relationships among executives recruited nationwide only over the past few months, few of whom had experience with projects of the scale or pace of the MARTA program, and some of whom would learn relatively early that MARTA was not their cup of tea.

Respond constructively to multiple constituencies:
- as represented by a MARTA Board whose directors were appointed by political bodies from four counties and from the city of Atlanta, with two counties having representatives even though local elections there had rejected referenda to authorize a sales tax earmarked for MARTA
- as represented by the entire state legislature, which authorized a local sales tax increment to get MARTA started, as well as a blue-ribbon committee to oversee MARTA operations
- as represented by federal agencies that were variously regulators and dispensers of grants for mass transportation projects

Be open to a broad range of local inputs—as to design features and so on—within the context of the authorizing referendum.

The Need for Multiple MARTAs

The dilemmas should be obvious. Spritely shifts were critical, so MARTA figuratively needed one foot firmly planted in the past and present while never neglecting the demands of the future. It all amounts to energetic effort, in many senses and by many people, to work themselves out of jobs, to leap in timely ways from the known and relatively comfortable to the less known and disturbing, and (perhaps most difficult of all) to identify resolutely with that which soon will need to give way to a succeeding stage of development. Success required unusually motivated people operating within a social structure that would encourage them to transcend their immediate personal interests. Moreover, the development of those people and that social structure could not be entrusted to natural processes; they required definite steering and early nurturance. That much was clear from the start.
Toward an Organic Managerial System

General Manager Alan Kiepper, appointed in late 1972, reviewed these basic facts about MARTA and came to two early conclusions. First, no "standard operation" approach would suit MARTA. From the very beginning, then, he planned on considerable time and energy to prepare MARTA officials to cope with certain uncertainty and permanent temporariness—to accept the need for swift organizational transitioning at emotional as well as intellectual levels, and to build skills and a managerial support group to ease the building and unbuilding of the several MARTAs. The guiding model was an organic one, of MARTA as a dynamic and evolving organization with which its members and several publics could identify as an open and effective managerial system.

Second, the task would be complicated by the backgrounds of most of the MARTA management, especially the senior staff—the seven individuals around whom Kiepper hoped to build the initial MARTA management system. Most of the senior staff, including Kiepper, came from nontransit backgrounds. Most, again like Kiepper, had been in and around local government. Far more significantly, few of these executives had any direct experience with building an organization from the ground up, or with the successive and sudden transitioning required in MARTA. In very basic senses, MARTA was a very new game for most of its prominent players. The MARTA environment contrasted sharply with their past experience. For example, Kiepper had been a city manager and these officials basically inherit a workforce, a set of policies and traditions, as well as a governing body with more or less clear policy preferences. City managers likely will not want for lack of those offering firm directions, nor for traditions and precedents.

The dangers of these executive resources were dual. Thus MARTA might suffer at the top levels from the inflexible imposition of policies and procedures born of experience in different arenas but inappropriate to the unique features of the MARTA environment and mission. Alternatively, MARTA might flounder through developmental periods with too little direction, too late, with the consequence that it would miss its opportunity to be "something more than just become another government agency."

Early on, Kiepper saw both the opportunity and the danger. He and his senior staff had the chance to develop a model agency, with policies and procedures relatively unhindered by the past practices and complex legal and personnel restrictions so characteristic of Georgia public agencies. On the other hand, building an organization—largely from the ground up—proved to be even more challenging than Kiepper had anticipated. Today’s decisions, in effect, might constitute precedents in terms of which tomorrow’s problems would be solved. The major difficulties? It was not always clear what specific problems tomorrow’s MARTAs would bring. Moreover, solutions to problems appropriate to one MARTA developmental phase might be very awkward at other stages. One senior
## Exhibit 4.1 Dominant Characteristics of Two Opposed Ideal Managerial Systems

<table>
<thead>
<tr>
<th>Coercive-Compromise System</th>
<th>Collaborative-Consensual System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superordinate power is used to control behavior reinforced by suitable rewards and punishments.</td>
<td>Control is achieved through agreement on goals, reinforced by continuous feedback about results.</td>
</tr>
<tr>
<td>Emphasis on leadership by authoritarian control of the complaint and weak, obeisance to the more powerful, and compromise when contenders are equal in power.</td>
<td>Emphasis on leadership by direct confrontation on differences and working through any conflicts.</td>
</tr>
<tr>
<td>Disguising or suppression of real feelings or actions, especially when they refer to powerful figures.</td>
<td>Public sharing of real feelings and reactions.</td>
</tr>
<tr>
<td>Obedience to the attempts of superiors to influence them.</td>
<td>Openness to the attempts to exert influence by those who have requisite competence or information.</td>
</tr>
<tr>
<td>Authority/obedience is relied on to cement organization relationships.</td>
<td>Mutual confidence and trust are used to cement organization relationships.</td>
</tr>
<tr>
<td>Structure is power-based and hierarchy-oriented. Individual responsibility. One-to-one relationships between superior and subordinates. Structure is based on bureaucratic model and is intendedly stable over time.</td>
<td>Structure is task-based and solution-oriented. Shared responsibility. Multiple-group memberships with peers, superiors, and subordinates. Structure emerges out of problems faced as well as out of developing consensus among members and is intendedly temporary or at least changeable.</td>
</tr>
</tbody>
</table>


Staff member put it succinctly, and with tongue-in-cheek: “It was all a very simple matter of learning how to hold fast to things that you could quickly let go of, just as soon as necessary.”

Complications notwithstanding, Kiepper knew one of the basic requirements for smoothing the several transitions that MARTA would face. That was an “effective management team,” whose members would support one another, share information, and work toward objectives that the members owned. Kiepper had a positive experience in his last job with a technology for helping build such a team among supervisors and managers in Richmond, Virginia, and he quickly
decided on a similar strategy for MARTA. His primary consulting resource became Dr. Robert T. Golembiewski, Research Professor of Political Science and Management at the University of Georgia—your author. Kiepper and Golembiewski had a consulting relationship going back several years, which they quickly renewed when Kiepper took the MARTA job.

The two men decided on the basic thrust—"team-building" experiences for all members of the MARTA managerial system, including the Board of Directors. To help prepare for the necessary transitional shocks, learning designs based on the "laboratory approach to organization development" were used to accelerate the development of the MARTA managerial team, (e.g., Golembiewski, 1972, 1979) as well as to influence the style in which that team would conduct its public business. Broadly, the goal had aspects of both avoidance and approach. As much as possible, the start-up goal sought to avoid the closedness and ponderousness of large-scale bureaucracy of an organized system that is oriented toward problemsolving and emphasizes timely change of complex, temporary systems.

Exhibit 4.1 provides substantial illustrative detail about what the development of the MARTA managerial system sought to avoid, and what it sought to approach. The anti-goal was a coercive-compromise system of management, and the thrust was toward enhancing the collaborative-consensual features of the MARTA managerial team. Neither pure case ever existed, but the tug-of-war establishing the shifting balance between them constituted the essential dynamics at MARTA.

ESSENTIALS OF EARLY LEARNING DESIGNS FOR ACTION-RESEARCH

The overall planning of early MARTA learning designs involved three approaches in moving toward a collaborative-consensual system: These include:

A team-building experience for the general manager (GM) and seven aides who made up the senior staff, a linkage between the first and second tiers of management;

An interface experience between the assistant general managers and the third tier of management, a linking of the top managerial levels;

An interface experience between senior staff and the MARTA Board, that is, an emphasis on the linkage at the politics/administration interface.

Each design element was the focus of a separate three-day session held at the University of Georgia’s Center for Continuing Education, and each element also involved various follow-up activities. We will call the three basic learning experiences Athens I, II, and III, respectively, after the Georgia city in which planned learning experiences were held, approximately one month apart.
Overall, the learning designs sought to sensitize top level MARTA managers to the “process” as well as to the substance of their work, as both contribute to an “action-research” orientation. “Process” refers to the “how” of the relationships of people and their groups: the quality of their interpersonal relationships, their orientations to conflict, the ways in which groups interface, the culture that comes to define their accepted ways of dealing with one another, (e.g., Schein, 1987). In essence, a process-orientation was seen as providing important raw material for the delicate transitioning between the several MARTAs. In short, the basic method at MARTA was to be timely and trusting talk.

Various off-site designs sought to encourage the attitudes and skills appropriate for a process-orientation but, far more, the goal involves building that orientation into the very fiber of the organization. Numerous targets of opportunity permitted real-time experimentation with the orientation as, for example, in the case of a senior staffer who arrived at Athens I with a serious skin disorder. “Too much sun?” someone inquired politely. “No, my doctor says it’s nerves. I guess I’m uptight about this meeting, and getting things started.” The consultant encouraged sharing of this unexpected learning opportunity. The senior staffer’s discomfort, in effect, became a vehicle for the process orientation—for exploring feelings and reactions about the Athens meetings and the broader job as well, as they impacted on participants.

The emphasis on action research sought to illustrate the multiple ways in which relevant data can be gathered, and how they might be quickly brought to bear to make many mid-course corrections and adaptations. Overall, action research implies a philosophy and a family of related technologies for learning about, and responding to, experience. To exaggerate a bit, the approach recognizes no mistakes, only unanticipated learning opportunities. Put another way, there are no bad decisions, only decisions with inadequate underlying processes and insufficient personal commitment.

FOUR DESIGN ELEMENTS

Four specific features also characterize the early MARTA learning designs. These were designed by Dr. Golembiewski, deliberately without involvement of General Manager-Kiepper so that he would start out in the learning process on an equal footing with other MARTA officials. These features included:

- Team-building
- Confronting
- Contracting

Measuring effects of team-building, confronting, so as to develop skills and experience relevant for an action-research orientation (e.g., Dyer, 1977, 1988).
Team-Building

Primary reliance was placed on varieties of team-building. Just as the last member of the senior staff had been recruited, and while other managers were still settling into new jobs, homes, and offices, three interlocking tiers of team-building were held. The basic rationale for team-building includes these three emphases:

1. Any management group can improve its operations.
2. Such improvement can be critical even for a management group that is well satisfied with its present performance, as in preparing for unpredictable stress situations.
3. An audit of interpersonal and group processes is an important way of testing for existing effectiveness, as well as of inspiring improvement. Such an audit can:
   - aid in increasing mutual understanding and empathy
   - heighten awareness of interpersonal and group processes, and so generate more realistic and detailed perceptions of “what’s going on”
   - help build identification, mutual goodwill, and comradeship born of a sometimes intense experience
   - facilitate the development of shared perspectives and frameworks that facilitate communication
   - emphasize the importance of reality-testing based on the fullest possible expression of information, reactions, and feelings
   - build norms encouraging openness, candor, and face-to-face confrontation.

Basically, team-building sought to encourage MARTA members to acknowledge and deal with the products of their interactions as persons and officials. Most of the officials associated with MARTA—both staff and board members—had only a brief interactive past, so there was little unresolved socio-emotional “garbage.” The focus was on the present and the future-soon-to-be-present. Team-building encourages almost nonstop interaction between participants, and this seems to speed up psychological time. One MARTA official reflected both aspects of the impact of team-building designs on time:

I’ve learned more in three days about you guys, and more about my place on the Senior Staff, than I probably would have learned in three weeks back at the office, for sure, or even in three months.

I also feel like I’ve been here forever, even though my calendar tells me it’s only been some 50 hours spread over three days.

MARTA’s planned reliance on team-building was not casually chosen. Much evidence implies that such speeded-up psychological time is particularly
useful around start-up. Simply, start-up implies a set of issues having substantial potential for polluting the rational-technical performance of an executive team. Illustratively, these characteristics include:

- Substantial confusion about roles and relationships;
- Fairly clear understanding of immediate goals, but lack of clarity about longer-run operations, which cumulatively induce wicked double-binds: a strong desire to get on with the task, and yet a pervasive concern that precedents may be set which can mean trouble over the long run;
- Fixation on the immediate task, which means that group maintenance activities will receive inadequate attention and individual needs will be neglected; and
- A challenge to team members that will induce superior technical effort, but that may also have serious longer-run consequences for personal or family life, and which in case probably will generate an intensity in work relationships that requires careful monitoring.

**Confronting**

Several designs encouraged MARTA officials to confront their several realities—to deal with one another and their tasks directly and supportively. Generally, confronting refers to a complex of attitudes and behavioral skills that are seen as enhancing a management team’s attitudes and skills and that can make members more aware of their socio-emotional processes, as well as more effective in their management.

Confronting often has a colorful press, as in versions that advertise “telling it like it is” or “letting it all hang out.” As used here, however, confronting involves these multiple social exchanges (e.g., French and Bell, 1973, 1994):

- Team members become more aware of their own reactions and feelings, as well as those of other members.
- Team members become more aware of the stimuli inducing particular reactions and feelings in themselves and others.
- Team members accept and maintain a norm that sanctions the sharing of the full range of applicable information, reactions, and feelings.
- Team members develop skills to share their concerns in ways that encourage similar expression by other members (e.g., French and Bell, 1973, pp. 99–121; Golembiewski, 1979, vol. 1, pp. 29–79).

Guidance in respecting OD values in the context of team building and action research came from the major external consultant—your author—who worked with MARTA at many levels for nearly a decade. As appropriate, other facilitators and consultants were added for different learning designs. Generally,
also, intervenors sought to emphasize degenerative \rightarrow regenerative interaction, which was introduced early in this volume and gets emphasis throughout.

Multiple vehicles were used to encourage confronting in the early MARTA designs, but the development and sharing of 3-dimensional images illustrate the genre. For example, General Manager Kiepper prepared three lists in response to these questions:

- How do you see yourself in relation to the assistant general managers (AGMs)?
- How do you see your AGMs?
- How do you believe the AGMs see you?

As a group in Athens I, the other members of the senior staff collaboratively also developed lists in response to three similar questions directed at the general manager. The lists were prepared separately and then shared, as later discussion will show. The design was similar in Athens II and III, although many details differ.

Extended discussion of the two 3-D images, with the aid of a consultant, constituted the basic early experience with confronting, and also provided substantial skill-practice with appropriate attitudes and behaviors. Mechanically, the procedure is simple. The two sets of 3-D images, written on large sheets of newsprint, are taped to a wall, side-by-side. Participants survey the lists, and are urged to ask for examples when the meaning of some item is obscure or confusing. The basic ground rule is that participants seek to understand the image, and to acknowledge any feelings of defensiveness or resistance but not dwell on them until all lists have been reviewed. Such discussion and analysis can be both varied and intense, but are typically accompanied by periods of explosive laughter and friendly commiserating, as in a mutual reduction of tension.

The reliance on team-building at MARTA rested on confidence that such designs tend to work for five basic reasons. First, participants—especially in rapidly changing or high-technology organizations—need such information, discomforting or even initially hurtful though it may be. One cannot have effective transitions involving complex collections of specialists while enjoying the luxury of letting issues work themselves out in their own good time.

Second, participants typically understand that the best way—indeed, perhaps the only way—to raise the probability of receiving such needed information in the future is to be accepting of the 3-D images in the present. Acceptance does not necessarily mean agreement, be it noted. “I can understand how you see it that way,” expresses acceptance without necessary agreement, “but I hope you recognize this is one of the many things about which reasonable people can and do differ.”
Third, confronting 3-D images is a shared experience that can build mutual identification and understanding, which is what many participants are seeking. The design is accepted and valued, consequently, and that makes it “work.”

Fourth, most individuals are uncomfortable if their verbal or nonverbal behavior is at some substantial variance with what they really know, believe, or feel. This lack of comfort increases sharply if the person suspects that relevant others are being similarly incongruent themselves, as well as aware of the general lack of authenticity. This analysis of 3-D images usually helps to reduce such variance by encouraging a mutual escalation toward openness and owning. Participants typically are much concerned about what the other group or person is writing on their 3-D image in that other room, for example, for they realize that too much varnishing of the truth on their part will be painfully apparent when the 3-D Images are compared. Greater but still tentative openness and owning will free for productive use energies previously needed to repress information.

Hence confrontation designs usually leave participants with a sense that barriers are being lowered, and things are “really happening” with less effort. Consider the symbolism in these common reactions to a 3-D image exchange. “Well, that took the cork out of the bottle, and about at the right time,” reports one participant. Another MARTA official provides this insightful perspective on the exchange of images:

That’s quite a load off my mind, although I didn’t quite dare to put down on paper all that concerned me. I’ll look for an early opportunity to make some further mileage. It was a good start, and not as tense as I expected. I guess all of us really wanted to get over the hump of mannerly closedness, but none of us knew how or was willing to risk starting what we all were clearly eager to do. We were off-and-running on the 3-D Images almost before the instructions were completed.

Fifth, substantial agreement typically exists between pairs of 3-D images, and this process builds additional content into increasingly shared visions of reality and desirability. This agreement almost always increases the participants’ sense of mutual competence and acceptance, by confirming that one person or group shares perceptions with another as well as by signaling that a real process of exchange has begun. Especially for a managerial team, it is both critical and comforting that its members see some important issues in similar ways, and that they also characterize the same processes in similar terms. Moreover, the resulting mutual enhancement of self-esteem can provide powerful impetus toward, as well as a solid foundation for, future communication and collaboration.

In sum, the areas of major agreement on pairs of 3-D images signal that the confronted other is like the self in significant ways, which can build a sense of identification and empathy. In turn, this learning can encourage further at-
tempts to uncover those areas where agreement or even awareness does not exist. As the degree of agreement approaches 100 percent, to that degree are Image sharings useful only as validations of a healthy condition of congruent perceptions that permit confident action-taking. To the degree that the agreement approaches 0 percent, so will Image designs tend to “work” with difficulty, if at all.

**Contracting**

A variety of contracting exercises also were built into the early MARTA learning designs to ensure that the sharing of 3-D images does not merely dissipate, the Assistant General Managers each prepared “shopping lists” directed at the other. The three lists constituted responses to three questions:

1. What should you *keep doing* about as now?
2. What should you *stop doing* that you now do?
3. What should you *start doing* that you do not do now?

Participants were encouraged to be as specific as possible about the behaviors or attitudes in question. Not particularly helpful are such global injunctions as Be smarter!

The three lists become the bases for a complex exchange process, which follows such a form: If you want me to stop X behavior, are you willing to do more of Y, which is on my list of behaviors that I would like you to start performing?

The transparent intent of contracting seeks to model a process that can be used not only at the training location, but especially back home. More immediately, the goal is to build agreement among participants about a few exchanges for openers, as it were. The process of reaching this agreement often induces forces that can later reinforce any trade-offs made in the contracting period. The potential social power implicit in such group decision-making has been amply documented by much behavioral research, beginning with the classic experiments by Coch and French (1948).

**SOME MEASURES OF MAJOR EFFECTS**

The heart of the action-research approach involves determining the effects of such interventions as team-building. The derivative feedback can be used to continue or increase what seems to work, and to abandon or modify that which does not have the intended effects. Often, ordinary language provided the vehicle, as MARTA officials shared reactions, feelings, and ideas. Specific exercises—such as one centering on guidelines for giving effective feedback—sought to improve the quality of such critical exchanges (e.g., Golembiewski, 2000, pp. 423–426).
Illustrating Replications

Sometimes, instruments of various degrees of structure and sophistication were used to solicit and share data. Later, Exhibits 4.2 and 4.3 reflect two examples of such approaches to determining if effects had a similar purpose.

SENIOR STAFF LEARNING EXPERIENCES

Most time and effort went into enhancing the relationships among members of the MARTA senior staff—toward highlighting relevant attitudes and for building appropriate skills to ease the transitions between the several MARTAs. That emphasis had a simple rationale. For good or ill, the initial senior staff would substantially influence the character of the unfolding MARTA. Some overall sense of this time and effort involving senior staff can be conveyed briefly. Members of the senior staff spent approximately 40 hours at off-site learning experiences being exposed to exercises/concepts relevant to group dynamics, as well as to appropriate skill-practices. Most of this learning time was scheduled in one 3-day period, Athens I. The exercises/concepts dealt with such topics as decision-making, interpersonal openness, giving and receiving feedback, functional roles, and interpersonal orientations of senior staff.

The senior staff also participated in Athens II and III, to be described below. In addition, a process observer attended approximately 10 worksite meetings of the senior staff during the 3-month interval following Athens I to help reinforce real-time effort consistent with the off-site experiences.

Four illustrations provide useful detail about the early learning experiences of the MARTA senior staff. Specifically, the examples deal with:

- The introductory experience, which directs attention to how well the senior staff operated as a group;
- The sharing of 3-D images, which directs attention to the perceptions of one another held by the general manager and the senior staff, as well as to how this means of confrontation led to useful contracting;
- How the personal needs of specific Senior Staff members conflicted with one another, or complemented each other; and
- How the effects of the learning interventions were tracked to suggest how such knowledge about results was used in an action-research mode.

Early Senior Staff as a Decision-Making Group

GM Kiepper early and insistently signaled his strong preference that the senior staff often would function as a group or team, as a collection of senior advisors who would discuss and debate in the process of moving toward decisions that all could own and implement effectively and (if possible) enthusiastically. This did not imply that all or many decisions would be made by the senior staff group. Sometimes that would be the case; but a decision also might reflect a general
manager’s “thou shalt.” In most or all cases, senior staff members should try to influence a decision, to “put their oar in,” so as to help determine MARTA’s directions.

 Appropriately, then, the first learning element scheduled for the senior staff focused on their effectiveness as a decision-making group. Newly assembled and with relatively little common experience, the senior staff gathered for an opening dinner at Athens I. The initial talk was mostly light and good natured, and senior staff clearly tended to see themselves as a special collection of individuals, the products of national searches. But some striking evidence suggested the undercurrent of concern about how the newly assembled individuals would function as a team.

 It did not take long at Athens I to elicit a deeper sense of senior staff dynamics. After dinner and cordial welcomes by University officials, the new executives assembled for three days of work to be directed by Dr. Golembiewski. He drew immediate attention to the model underlying the experience, which can be summarized in terms of two opposed models for interaction (see Figure 4.1).

 The MARTA executives acknowledged the role of the model in their challenges and inventiveness. The goal was to approach “regenerative systems” and to avoid “degenerative systems.” Golembiewski provided only a little detail. The two systems of interaction were products of four interacting variables, two of which needed special distinguishing. “Openness” was possible without “owning,” that is; one could own a lie and hence not be open. The senior staff got the point quickly. “If I told you that some people say you’ll probably never be a ballet dancer when I really had that opinion,” one senior staff member told Dr. Golembiewski, who is built like a fire plug, “that would be an open comment, but I would not be owning it.” We all enjoyed the laughter and tension-release, and

**Figure 4.1** Two systems of interaction.
the example made the basic point forcefully. Senior staff also discussed some of the typical consequences of degenerative systems, which include:

- Communication and decision-making processes become increasingly burdened.
- Organization members become less effective in isolating and resolving substantive issues.
- The amount of unfinished business sharply increases.
- Members feel diminished interpersonal competence and psychological failure as they fail to solve problems that stay solved without creating other problems.
- Members become more dependent and overcautious, and respond by “tattling” activities, by preoccupation with being “safe,” or by “don’t rock the boat” attitudes.
- Organization norms restricting owning and openness develop.
- Tendencies toward fragmentation or organization units and officials are enhanced.

Senior staff agreed that those consequences were well worth avoiding, and regenerative interaction would help do the job.

What was the early balance of regenerative versus degenerative tendencies in the new MARTA? And how did the senior staff go about making decisions or solving problems? Some indications came quickly. Senior staff were asked to work on the Desert Problem, a well-known training device that asks group members first for individual decisions about the relative value of 15 items to the survivors of an airplane crash in a desert, and then requests a single, collective rank-order. The importance of the items—a parachute, matches, water, etc.—had previously been ranked-ordered by many experts in survival. So decisions by individuals or a group about relative value can be reasonably scored by calculating differences between actual rankings and the “ideal.” For example, assume that an individual ranked water 4 in importance to survivability on the list of 15 items. The expert rank-order is 1. The differences in scores for all items can be added together, and the higher that sum, the “worse” the decision. A perfect decision would have a total difference score of 0—in this case, each of the 15 rank-orderings of some individual or a group would be exactly the same as each of the expert rankings. (e.g., Slevin, 1978).

Typically, the group difference score is lower than the average of individual difference scores. This reflects the power of groups—bringing together diverse resources, combining partial knowledge, and so on. Dr. Golembiewski expected some such result. He was also prepared to detail some guidelines that could help to consciously engage group resources so as to improve MARTA decision-making and problem-solving. The goal was clearly multiple, then: to provide
some experience with a facet of reality relevant to the senior staff; to evaluate collective performance; and to provide some conceptual understanding that could be useful in the future to help the developing team function effectively. This reflects an action-research model, if a sharply limited one.

The MARTA executives provided two surprises. The group difference score was unexpectedly higher than the average of the individual scores. Moreover, the group difference was higher than each and every one of the individual scores! Directly, that is to say, 30 minutes of group problem-solving had generated a collective decision “worse” than each of the individual decisions. The surprise on Golembiewski’s face was obvious as he listed the individual scores, determined their average, and then calculated the group difference score.

The “surprise” provided a valuable learning opportunity, which involved the senior staff in several hours of close analysis of what they had done to one another, or how their interpersonal and group processes had gone so awry. Those analytical dynamics cannot be captured here, but they can be illustrated by a few dominant themes. They show how a game can bring serious issues to the surface.

“We didn’t win any prize on that one,” one senior staffer noted. “Have you ever seen a sorrier outcome, Dr. Bob?”

“No,” he responded. “The outcome is almost unique—uniquely bad.”

The executives agreed, and began to work through the complex “whys” of that uniqueness. For example, the person dominating the group discussion had also been very influential back at the office. That person’s training and knowledge were basically legal, however, and ill-suited the desert problem. “I really led them down the primrose path,” that individual agreed. “What intrigues me is how easily they followed me, when I was winging it.” The laughter was uproarious, but suffused with grim realizations about how much the senior staff had to learn about how much progress had to be made toward developing “quality controls” for the information admitted to problem-solving and decision-making.

The Desert Problem also pointed up one of the major ironies of life in organizations. The appropriate resources were available in the case of the survivor’s game, the senior staff also learned. Revealingly, that is, one senior staffer had an almost-perfect individual score, but he could not sufficiently influence the group decision-making. He had tried, but he did not make it sufficiently clear that he had very special knowledge: not only was he a pilot, but he had some survival training, and even had worked on the same problem a year or so earlier.

“Nothing new in what just happened,” he noted. “It’s just like the office. I have trouble being heard there, too.” The situation had even more curious wrinkles. “I didn’t pound the table,” the person with special knowledge noted, “but I knew we were going to hell in a wheelbarrow. I was intrigued by how I reacted once the ball got obviously rolling on the bad assumptions. You guys almost convinced me I was wrong, so much so that I didn’t do something drastic to stop the way things were going. But I knew that you were wrong.”
Illustrating Replications

Patently, the learning design had unearthed precisely the kind of dynamics that might be devastating in real-time settings. OD values and skills were intended to reduce the probability of such outcomes when MARTA team members were back in their offices.

Sharing of 3-D Images and Contracting

Toward the end of Athens I, senior staff developed and shared 3-D Images, which are reprinted in Exhibit 4.2 with only minor editorial changes to permit understanding by those without specific knowledge of MARTA operations. The exchange of images was tense, as might be expected. Outbursts of laughter signaled the useful release of tension, as both the GM and AGMs explained their 3-D images, provided examples or illustrations, and experimented with how regenerative they could make their interaction.

The 3-D Images provided much grist for contracting, patently, and hours were spent on developing the details of stop, start, and continue associated with action planning. The specific contracts entered into by the MARTA Senior Staff will not be reported here. Recall that contracting involved the preparation of three lists—stop, start, and continue—about whose specific items GM Kiepper and his senior staffers could bargain. Generally, those contracts focused around more substantial freedom of action for the assistant general managers, a major item on the AGM’s start list which was exchanged for several items on the GM’s start list. Overall, senior staff became more aware of how their past experiences in more stable and structured local government situations could limit early and resourceful responses to the novelty and quick-silveredness of the MARTA challenges.

Assessing Personal Needs

As experience with the senior staff accumulated, and as mutual trust grew, learning designs emphasized deeper levels. For example, the chafing between GM Kiepper and one of his assistant general managers (AGMs) became increasingly clear to all. In part, discussion revealed, the conflict rested on opposing ambitions. The AGM not only wanted to be general manager, but believed he would do as well as or better than Kiepper. Reasonably, the GM did not share the AGM’s point of view. Once exposed, such crossed ambitions and their consequences had to be dealt with over time.

Dealing with such ambitions-in-conflict is never easy, but one approach isolated specific contributions to conflict between two men, in the hope of providing a general model applicable to such situations. For example, Dr. Golembiewski suspected the conflict inhered in substantial part in different personal predispositions of the two men. In general, it would be useful to segregate the conflict
### Exhibit 4.2 3-D Images That Facilitated Confrontation

<table>
<thead>
<tr>
<th>3-D image prepared by AGMs</th>
<th>3-D image prepared by general manager</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. How AGMs See General Manager</strong></td>
<td><strong>I. How General Manager Sees AGMs</strong></td>
</tr>
<tr>
<td>1. Unapproachable (closed door)</td>
<td>11. Sensitive</td>
</tr>
<tr>
<td>2. Not open</td>
<td>12. Organization above</td>
</tr>
<tr>
<td>3. Dedicated</td>
<td>13. Tough</td>
</tr>
<tr>
<td>4. Determined</td>
<td>14. Poor delegator</td>
</tr>
<tr>
<td>5. Cool under fire</td>
<td>15. Aloof</td>
</tr>
<tr>
<td>6. Hard working</td>
<td>16. Formal</td>
</tr>
<tr>
<td>7. Priorities (not ordered)</td>
<td>17. Highly structured personality</td>
</tr>
<tr>
<td>8. Meticulous</td>
<td>18. Programmed</td>
</tr>
<tr>
<td>10. Too detailed</td>
<td>20. Violates chain of command</td>
</tr>
<tr>
<td><strong>II. How AGMs See Selves in Relation to GM</strong></td>
<td><strong>II. How GM Sees Self in Relation to AGMs</strong></td>
</tr>
<tr>
<td>1. Insecure</td>
<td>1. Spread too thin personally</td>
</tr>
<tr>
<td>2. Ineffective</td>
<td>2. As spending too little time with AGMs as group</td>
</tr>
<tr>
<td>3. Frustrated</td>
<td>3. As emphasizing brushfires, without breathing room to focus on key issues</td>
</tr>
<tr>
<td>4. Unable to perform effectively</td>
<td>4. As too lenient with AGMs as to assignments, deadlines</td>
</tr>
<tr>
<td>5. Wasted and unimportant</td>
<td></td>
</tr>
<tr>
<td>6. Inhibited</td>
<td></td>
</tr>
<tr>
<td>7. Too willing to please</td>
<td></td>
</tr>
<tr>
<td><strong>III. How AGMs Believe GM Sees Them</strong></td>
<td><strong>III. How GM Believes AGMs See Him</strong></td>
</tr>
<tr>
<td>1. Relies on group (reluctantly)</td>
<td>1. As cautious, indecisive</td>
</tr>
<tr>
<td>2. Trusts us, with reservations</td>
<td>2. Not trusting, due to newness of relationships</td>
</tr>
<tr>
<td>3. Does not see us as team</td>
<td>3. Expressed lack of trust via deadlines, detailed reviews</td>
</tr>
<tr>
<td>4. Naïve (an sometimes we are!)</td>
<td>4. Busy-busy, not having time for at least some AGMs</td>
</tr>
<tr>
<td>5. Sees potential in us</td>
<td></td>
</tr>
</tbody>
</table>
deriving from the two sources—as much as that is possible—so as to treat differently the conflict deriving from different sources. Why? Strong forces encouraged trying to make a success of the new MARTA team, so laboriously assembled. Moreover, the AGM in question had strong technical qualifications and other strengths, and could contribute significantly to the Authority if his ambition for position could be somehow kept to manageable scale, or at least kept in clear perspective.

A learning design at Athens II in which all senior staff participated—but one chosen with a special eye toward the GM and his ambitious AGM—drew attention to the conflictful and complementary potential in individual differences. A lecturette introduced senior staff to the essentials of Schutz’s notions of “fundamental interpersonal relations orientations,” or FIRO (Schutz, 1958). Each senior staff then filled out a form like that duplicated in Exhibit 4.3—a quick-and-dirty but useful approach—which essentially asks an individual to rate self on three orientations, each taken from two perspectives. Senior staffers were asked to consult with fellow executives with whom they interacted well or poorly, so as to match their perspectives. The GM and his conflictful AGM got together and, as Exhibit 4.3 shows, their profiles were at extreme odds. Both needed to exercise power over others, for example, and resisted power that others exercised over them. Similarly, the GM and his AGM had contrasting orientations toward personal intimacy and need for inclusion. Relatedly, the two executives had contrary predispositions about expressing warmth toward others.

The likely results? Chafing or abrasion due to their make-up as people, which exacerbated any conflict due to conflicting ambitions.

Such designs need to be kept in perspective, discussion emphasized. Knowledge about differences in personal orientations did not eliminate all contention between the men, but it did provide one alternative way of interpreting some of the very real differences between them. And it helped avoid confusing quite different contributions to those differences. Consider that, without such a distinction, any difference between the men might be interpreted in these terms: “There he goes again. His ambitions got the better of him, and he is trying to undercut me.” It makes an enormous difference to say or think: “Our differing personality predispositions may be at work again.” The first alternative encourages win/lose defenses; and the latter implies the possibility of recognizing an individual uniqueness as motivating behavior that could be interpreted as reflecting some animosity directed at the other.

Tracking Effects and Action Research

All MARTA learning experiences sought two goals: to uncover timely and valid data on which action could be based; and, more basically, to practice with a model, along with supporting attitudes or skills, that could later be applied to
EXHIBIT 4.3 Fundamental Interpersonal Relations Orientations of two MARTA Executives

Locate yourself on each item below at a position that represents the way you see your interpersonal needs.

1. As I see my needs to set direction for other people, they are:
   
   ___ low  ___ medium  GM  AGM high

2. As I see my needs to have direction set for me by other people, they are:
   
   GM  ___ medium  ___ high
   AGM low

3. As I see my needs to include others in my activities, they are:
   
   GM low  ___ medium  AGM high

4. As I see my needs to have others include me in their activities, they are:
   
   ___ low  ___ medium  GM  AGM high

5. As I see my needs to express warmth toward other people, they are:
   
   GM low  ___ medium  AGM high

6. As I see my need to have other people express warmth toward me, they are:
   
   ___ low  ___ medium  GM  AGM high


ease the transitioning between the several MARTAs that would inevitably come, slowly or surely, for good or ill.

Many approaches to action-research were taken. For example, the effects of the confronting and contracting among the senior staff were judged by a series of semi-structured interviews as well as by periodic administrations of a paper-and-pencil test—the Group Behavior Inventory, or GBI (Friedlander, 1970). Data derived from these two approaches were fed back to the senior staff, to establish
effects of past designs as well as to permit a wide range of action-planning for the future.

The GBI, whose basic dimensions are sketched in Exhibit 4.4, proved useful in two basic ways. First, several administrations of the instrument permitted a test of whether the expected consequences of the team building did occur—for specific members of the senior staff as well as for the aggregate. Exhibit 4.4 details the effects anticipated from a successful team-building effort and also summarizes the actual results for the senior staff. Note that “short-run effects” compare GBI responses immediately before Athens I to responses obtained two weeks later; and the “longer-run effects” involve comparisons of the benchmark GBI scores with an administration of the instrument after Athens III, beyond which time the effects of the team-building became increasingly confounded by the rush of workday activities.

The GBI results in MARTA proved generally consistent with expectations about the effects of a successful team-building experience. Overall, that is, approximately two-thirds of the changes reported by participants are in the expected direction, and seven of the twelve aggregate comparisons achieve usually accepted levels of statistical significance. Interviews corroborate this pattern. One participant noted, to illustrate the dominant and all-but-universal reaction:

You didn’t promise us a rose-garden, I know. But it was quite a shock to confront an array of issues so early in the game. It violated all my governmental experience, where such confronting was done at greater leisure, and probably not at all. But I’m increasingly glad we did the team-building. MARTA cannot afford to let nature take its course. There is too much to be done in a short period to risk being sand-bagged, or just hoping that issues will go away if they are neglected long enough.

Note that the GBI results also would have been useful—perhaps even more useful—if they had indicated the designs had not worked. Such data would signal a dilemma, and no doubt would have motivated efforts to discover resolutions of that dilemma.

Second, GBI results also were used to indicate where follow-on activity might be appropriate for individual participants. This permitted another kind of action research. To explain, all senior staff identified their completed GBI forms, which were returned to the consultant at his university address. Only aggregate results would be discussed publicly. In the two cases in which senior staff indicated by their GBI responses that the short-run effects of the team-building were for them ineffectual or negative, however, the consultant contacted the respondents and sought to verify their “deviant” GBI scores.

The GBI proved sensitive in both cases. In one case, the respondent saw the initial experience as more hopeful than impactful, but certainly not harmful
### Exhibit 4.4  Target Dimensions, Expected Effects, and Actual Effects of Basic Team-Building Design

<table>
<thead>
<tr>
<th>Group behavior inventory dimensions</th>
<th>Predicted effects</th>
<th>Observed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group Effectiveness:</strong> solving problems and formulating policy through a creative, realistic, team effort.</td>
<td>Increase in effectiveness is expected but may be slow to build</td>
<td>5 of 8 cases</td>
</tr>
<tr>
<td><strong>Approach to vs. withdrawal from leader:</strong> at the positive pole are groups in which members can establish an unconstrained and comfortable relationship with their leader.</td>
<td>Increase in approachability is probable, but may be slow to build</td>
<td>6 of 8 cases</td>
</tr>
<tr>
<td><strong>Mutual influence:</strong> members see themselves and others as having influence with members and the leader.</td>
<td>Substantial increase even in the short run, which should persist; note that GM may perceive inroads on personal authority and power</td>
<td>6 of 8 cases</td>
</tr>
<tr>
<td><strong>Personal involvement and participation:</strong> members want, expect, and achieve active participation in group meetings.</td>
<td>Substantial increase even in the short run, which should persist</td>
<td>7 of 8 cases</td>
</tr>
<tr>
<td><strong>Intragroup trust vs. intragroup competitiveness:</strong> members have trust and confidence in each other.</td>
<td>Increases is expected, but may be slow to build</td>
<td>3 of 8 cases</td>
</tr>
<tr>
<td><strong>General evaluation of meetings:</strong> a generalized feeling about the meeting of one's group as good, valuable, strong, pleasant, as contrasted with bad worthless, weak, and unpleasant.</td>
<td>Increase is expected, but may be slow to build</td>
<td>4 of 8 cases</td>
</tr>
</tbody>
</table>

*Source: Based on Friedlander, 1978, p. 295.*
to his relationship with other senior staff or with the general manager. He rejected the consultant’s suggestion about a follow-on meeting with Kiepper.

The second respondent reported a significant worsening of his relationship with the GM over the interval between the first and second GBI administrations. The team-building session at Athens I reinforced the AGM’s perception of this rift and did nothing to resolve what was for him a one-on-one issue not amenable to discussion in that group setting. The consultant suggested a third-party design (e.g., Walton, 1969) to explore these issues. This suggestion the respondent accepted. Subsequently, the two parties got together at an early date to deal with their relationship—successfully, as it turned out.

Such transactions imply significant ethical issues even as they contribute to action research. In this case, basically, consultant Golembiewski sought to create the sense and reality of his independence. His commitment was to facilitate MARTA’s effectiveness, not to serve as an agent of the general manager.

One can fall into self-delusion in such matters, of course. Were the subordinates under duress to accept the consultant’s proposal? It is at least a good sign that one AGM felt free to reject that proposal, and GM Kiepper still is unaware of the identity of that AGM. Finally, as to the GM and the AGM whose performance until then was unsatisfactory to himself, as well as to the GM: both men had perceived growing issues between them. That AGM became a solid performer on the senior staff.

TWO LEVELS OF INTERFACE WITH DIRECTORS

Approximately one month after the session for senior staff alone, another three-day experience concentrated on two other sets of interfaces crucial to the fluid transitioning between the several MARTAs. The interfaces explored at Athens II include those:

- Between some 25 department directors and the senior staff
- Between each Assistant General Manager and the cluster of directors he directly supervised.

Senior staff sometimes met alone at Athens II, to continue working on their own relationships. And some time was spent working with their individual cluster of department directors. But most of Athens II involved working directly on the two interfaces above, with four themes suggesting the complex dynamics.

Confronting and Contracting With Senior Staff

3-D images at Athens II were used to explore the interface between the senior staff and the directors. Six sets of images were prepared: one by each of the five clusters of directors reporting to individual AGMs, and one by the senior staff.
These were prepared in private, as in Athens I, and then publicly shared in a large common meeting.

The sharing of 3-D images was an intense experience, in large part due to several major issues that had been generated in the early days of assembling a workforce and of developing personnel policies and procedures. Needs sometimes differed, feelings ran strong and deep, and the overall tone was a tense seriousness as MARTA officials sought to exploit the opportunity to influence the development of the specific kind of organization they preferred. Illustratively, these items were included by one group of directors in their list of perceptions of the senior staff:

- Some autocratic elements—dictatorial.
- Secretiveness—lack of communication.
- They have a tough job, in a rough environment, and are making good progress.
- Question competence level of some—admire competence of some.
- They earn their money.

Examples illustrating these descriptors were emphasized in an extensive public period, and the associated discussion centered around several major substantive issues, mostly issues introduced by the directors.

Contracting took place at two levels. The senior staff agreed to study a long list of issues, many of which were criticisms of newly instituted personnel policies and procedures. Parenthetically, some quick changes in policies and procedures were made soon after; other issues were studied over longer periods, with additional changes being made later. This responsiveness no doubt reinforced the impact of this second design element.

**Directors and Individual AGMs Build Teams**

Moreover, substantial time also was provided so that each of the five AGMs with directors reporting to them could begin some rudimentary team-building in the organization cluster each supervised. Again, mutual 3-D images were developed and shared in each of the five clusters of AGM-cum-directors. Consistently, also, contracting was encouraged between the several clusters of directors and the individual AGMs to whom each cluster reported. In addition, each cluster was mandated to provide any additional detail about the substantive issues raised in the large public meeting.

**Tracking Effects and Action-Research**

An action-research orientation was a critical complement of these complex activities—both to evidence a seriousness about really reflecting on experience so as
Illustrating Replications

to learn from it, as well as to provide participants with a sense of the degree to which their investments of emotional energy had begun to pay off. The effects of Athens II to develop an effective and open system in MARTA were estimated by two administrations of the Likert Profile of Organizational Characteristics (Likert, 1967, 1977), a simple and useful instrument. The “benchmark” administration was immediately before the organization-building session at Athens II, and the post-experience administration came approximately two weeks later. The form of the instrument used contained eighteen items.

The profile has several interesting features. First, its several items can be scored along a continuum of twenty equal-appearing intervals, which are differentiated into four major systems of management:

<table>
<thead>
<tr>
<th>Scores</th>
<th>Likert systems</th>
<th>System descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–5</td>
<td>System I</td>
<td>Exploitative-Authoritative</td>
</tr>
<tr>
<td>6–10</td>
<td>System II</td>
<td>Benevolent-Authoritative</td>
</tr>
<tr>
<td>11–15</td>
<td>System III</td>
<td>Consultative</td>
</tr>
<tr>
<td>16–20</td>
<td>System IV</td>
<td>Participative Group</td>
</tr>
</tbody>
</table>

Second, each item is anchored by four brief descriptive statements, one statement for each system. For example, one of the eighteen items deals with the locus of decision-making. The System I statement is “mostly at the top”; and the System IV statement is “throughout [the organization] but well integrated.” Intermediate statements anchor Systems II and III, emphasizing “some delegation” and “more delegation,” respectively. Broadly, the assumption was that a System IV organization would be better suited than Systems I and II to making the quick and complex transitions required in moving between the several MARTAs.

Third, the eighteen profile items are intended to tap six broad phenomenal areas of organizational relevance. The instrument includes items referring to leadership, motivation, communication, decisions, goals, and control.

Fourth, MARTA respondents were asked to score each profile item twice. A *Now* score reflects an estimate of the *existing* level of each item, and an *Ideal* score on each item provides data about the *preference* of the respondent. In addition to providing useful data, the exercise alerts organization members to any gaps between their preferences and the interpersonal and group relationships that actually exist in their organization. This can help motivate early remedial action and provides yet another illustration of how the early MARTA learning-designs sought to encourage an action-research orientation.

Overall, a successful experience should move respondents’ scores toward
System IV, with the Now scores for almost all respondents on most items being in the range from high System I to low System III. The ideal scores clustered tightly in mid–System IV.

Two points about the impact of Athens II were highlighted when the Likert data were fed back to MARTA senior staff and directors so that they could be built into plans for action. Existing interpersonal and intergroup relationships fell substantially short of where the directors preferred. All Ideal scores are much greater than the Now scores, the feedback of data emphasized. Much room for progress and growth existed, obviously. Roughly, substantial improvements occurred on the quality of leadership, motivation, communication, and decisions. MARTA respondents reflected a less positive pattern on goals and control. This seems reasonable, given the longer-range considerations implied by goals and controls. Communication improved, to put it briefly and oversimply. But it would take some time to verify that such improvements would be reflected broadly in goals and controls over the long run.

The pattern is understandable, discussion emphasized, even as it suggests ample room for improvements via effective action-planning. Most directors had been hired only recently; some had been on the job only a matter of days; and a few had just been hired. These were very early days, indeed, and hectic ones.

Moreover, the feedback of data also emphasized that Athens II generally succeeded in moving interpersonal and group relations in the direction preferred by the directors. This is not the place to review reams of data but, on the whole, Now scores shifted to the range from high System II to high System III. In the least impactful case, to put the point in another way, the scores for an AGM and his directors more closely approach Ideal scores in 14 of 18 cases, comparing Athens II to I.

Both the challenge and motivation for an action-research orientation to these data were highlighted. “We came some distance toward our ideal in a short time,” one MARTA participant noted. “And we have a long way to go.”

An Unexpected Learning Opportunity

Progress was not restricted to off-site locations, of course, nor was it always planned. Indeed, the intent obviously was to build appropriate attitudes and skills that would permit transfer of any out-of-building learning into the office, where it could influence a broad range of MARTA activities that were both unplanned and unexpected.

One such unexpected learning opportunity occurred shortly after Athens II; it involved maternity benefits for unmarried MARTA employees and their partners. The item had occasioned much discussion at Athens II and had become a major issue in the contracting between the department directors and the senior staff, especially GM Kiepper. Mutual agreement was reached on the coverage.
Illustrating Replications

Briefly, GM Kiepper had second thoughts about that item of insurance coverage about a week after Athens II; indeed, he was leaning toward an opposed position, and expressed that point during the small talk at a convivial dinner for the senior staff, consultant Golembiewski, and their wives.

The intensity of Athens II suddenly became manifest again, much to the puzzlement of the wives. Golembiewski reacted strongly, and countered that Kiepper’s reconsideration would mean his resignation. The contract, mutually made, was not unilaterally revocable, Golembiewski argued. A general discussion ensued, with the sudden senior staff animation contrasting starkly with the desire of most wives to “keep things pleasant” and the curiosity of some about “what did go on at Athens?”

The regenerative attitudes and values worked on at Athens I and II soon got tested, however, and publicly. Discussion reemphasized the crucial role of the item in the contracting with directors, and stressed the point that the integrity of the learning experience demanded that a reconsideration of that insurance provision—if any were made—had to be made in the same arena in which the original contracting had occurred. The general manager eventually concluded: “Bob Golembiewski raises a point we neglected. Day-to-day pressures encourage one to neglect the fine points. Good gurus are hard to find, you know.”

LINKING THE BOARD AND SENIOR STAFF

The final design element in MARTA’s first pass at institution-building involved the board of directors and the senior staff. Like its predecessors, Athens III took approximately three working days. The meetings constituted the first extensive interaction of the board with many of the senior staff, and they also sought to provide some direct experience by board members with the attitudes and skills whose development was being encouraged among other MARTA officials.

Athens III came at a significant point in the development of the board, whose members were appointed by elected officials in four counties and the City of Atlanta. The board of nine members was clearly in transition. Earlier boards had been peopled by macro-prominents, often but not exclusively from the economic elites, with independent power-bases, and a collective regional orientation. Subsequently, appointments were made from other tiers of community leadership, with different profiles and backgrounds. “New” board members were generally less prominent than their predecessors, and their support-base was more likely to be in local politics than in large-scale business or commerce. Consequently, there was over time a growing responsiveness by Board members to more local needs and aspirations. At the time of Athens III, the board had quite substantial independence from local politics, but that relative autonomy clearly was being tested and would be substantially reduced as MARTA signed agreements with local governments and otherwise moved toward actual construction. Specifically,
MARTA had to deal with local communities on a wide variety of issues looming just over the horizon. For example, only local governments could close streets needed for MARTA construction; similarly, only governments could condemn properties that MARTA required; and so on.

The board’s style was to become increasingly active in MARTA business as staff were selected and policies developed, discussions between GM Kiepper and consultant Golembiewski emphasized. Hence the special importance of directly exposing board members to the kind of developmental experiences to which MARTA management had devoted some time and effort, with board knowledge. Moreover, the board had until then uneven but typically brief and sporadic exposure to the MARTA senior staff other than the GM. Board members desired far more contact with AGMs as a prime way of developing first-hand information about those AGMs whose policy recommendations and detailed design and construction proposals would increasingly come before the board for its action as MARTA moved toward construction. The importance of an early and mutual getting-to-know-one-another provided significant motivation for Athens III.

Three Design Features

Athens III’s contribution to institution-building in MARTA had three elements, each lasting perhaps two-thirds of a day. Sequentially, the design emphasized:

1. Separate meetings for Board members and the Senior Staff, during which data about their respective internal dynamics were fed back by consultants who summarized interview and questionnaire information to serve as stimuli to encourage the two groups to evaluate their ways of relating to one another.

2. Two integrative experiences, the first relatively gentle and the second far more threatening:
   - Board members and the Senior Staff independently developed verbal statements describing their concepts of MARTA’s mission in some detail, which statements became objects for public sharing and comparative discussion.
   - Board members and the senior staff independently developed 3-D images, and these became central stimuli for public confronting activities.

3. A work session, in which matters to be publicly presented in the immediate future were discussed by the board and the senior staff in common session.

The flow of the design involving the Board is direct, then. The first element stresses internal dynamics; the second encourages limited integration, consistent
with the offering roles and the board and the senior staff; and the third seeks to test the usefulness of the outcomes of the first two elements in a more or less normal work context. There were major elements of risk and threat in dealing with the board in the confrontive spirit with which MARTA managers sought to deal with one another. But there seemed no viable alternative to the risk and threat, given the active board role and given the style which earlier design elements had sought to foster among MARTA management.

One example illustrates how the board approached its own processes in an action-research mode. Interviews conducted by Dr. Golembiewski just prior to Athens III provided the basic grist for the mill. Basically, Golembiewski interviewed board members about their perceptions and reactions, their sense of progress and problems. Summaries of these data were reported back to the Board at Athens III, for Board analysis and planning in what amounted to a variety of confrontation design. For example, board members were asked three questions, among others:

On a nine-point scale, how do you rate the effectiveness of the MARTA Board?
Which specific Board members do you believe give the highest ratings?
Which specific Board members do you believe give the lowest ratings?

Avoiding details, the consultant coded all responses and entered them on a single visual. At a glance—although individuals were not identified—board members got an overall view of one another’s perceptions. Board members spent several hours contemplating that visual. The range of discussion was broad. For example, all agreed that the ratings could be very much higher, even though the board was new at its work. And hopes were expressed that those members giving the “bad marks” could be open with their reactions, tethering them to specific examples and suggesting improvements. Special concern was directed at those several unidentified board members who were seen by others as most positive about board efficiency, but who reported low ratings for themselves to Dr. Golembiewski.

Tracking Effects

No major effort was made to measure the effects of Athens III, but three conclusions are safe enough. First, the experience was a far more critical one for the senior staff than for the Board, for obvious reasons. Basically, staff was still concerned that its concept of its job might conflict with the board’s view of its responsibility; and staff was unclear as to board reactions to its overall performance.

Second, reactions of participants to the design were uniformly positive, and typically emphasized the useful beginning or acceleration of processes that
required nurturing over the long run. The responses of the board chairman and vice-chairman are typical. They were initially positive, and remained so after nearly a year. One official observed: “Unquestionably, I considered the experience worthwhile. It afforded us an opportunity to know one another better.” The other board official had a similar reaction, but stressed the need for determined follow-up after the team-building experience:

In general, because of the difficulty in getting people to listen, much less understand, I feel that such sessions are constructive, beneficial, and desirable . . .

Specifically, I feel that this particular meeting gave me an opportunity to know our Board and staff members better and to more fully appreciate the special relationships between the two groups.

I’d very much like to see these sessions repeated on an annual basis for two reasons: one, there are almost always personnel changes each year; and two, it appears that some participants have a tendency to forget the vital issues discussed and generally agreed upon.

Third, the board has met for a number of follow-on experiences in subsequent years, based on the same values that motivated the Athens experiences.

CONCLUDING OBSERVATIONS

No complex reality such as that sketched above can be summarized easily, but three points provide perspective on those early efforts toward making MARTA an open and effective giant, toward facilitating the transitions between the several MARTAs required for a successful program. First, the goal was to help build more effective teams at several levels of organization: in the board, on the senior staff, and among directors reporting to the same AGM. This team-building has an internal thrust and seeks to make members of small, formal groups more cohesive and more aware of their own dynamics. This encourages members of these groups to confront the differences that will inevitably exist among themselves, encouragement deriving from the similarities of experience and identification that are highlighted by successful team-building.

Broadly, several similar types of learning designs had similar effects. This suggests the reliability of the several levels of organization. Or to put the point in another way, OD values can be approached at different levels of organization by similar designs.

Second, each of the three major design emphases has a major relational thrust. Directly, team-building can be pernicious if it merely creates strong bonds of experience, identification, and affection among the members of some small
Illustrating Replications

Team-building in this sense can develop an intense sense of we-ness only at the expense of highlighting and perhaps manufacturing a they-ness to be distinguished from, or cross-walk experiences into the design for a more open and effective MARTA; the desirable outcome of team-building is to utilize any forces deriving from a successful experience to help bridge social and psychological and hierarchical distance rather than to merely exaggerate that distance that is variously normal or necessary or convenient.

The relational thrust is vital in MARTA for a very practical reason. Given its lack of taxing and independent condemnation power, MARTA had no reasonable alternative but getting and staying in the frame of mind that doggedly seeks the elusive reality of a complex sense of us-ness, not only inside MARTA but also (perhaps, especially) with various units of government and a wide range of interests.

Third, the three elements of the early learning designs also reflect an institutional or contextual thrust, in the form of a set of values that condition both the internal and relational thrusts distinguished above. In Selznick’s terminology, the team-building experiences and the cross-walks between them sought to infuse MARTA with a specific set of values. These values characterize MARTA’s “cultural pattern,” and imply partial answers to this critical question: To what social or moral purposes does the team-building and cross-walking contribute?

This set of values is reflected at several levels, with the Collaborative-Consensual system sketched in Exhibit 4.1 reflecting the basic normative thrust of the entire experience. To a similar point, the expectations about relatively specific consequences—as determined especially by the GBI and the Likert Profile—reflect a confidence about moving closer to preferred conditions. Those expectations at once constitute predications about what can be encouraged to exist; and those expectations also imply value judgments as to those conditions in MARTA that are desirable in both instrumental and ethical senses.

These basic issues of value often get overlooked, but they definitely were front-and-center in the three designs sketched above. Consider only a brief illustration of how the designs imply reinforcing sets of values that can serve the MARTA project. For example, movement toward System IV is desirable in at least three distinct senses. Thus, participants overwhelmingly prefer System IV, as their Ideal scores unanimously testify, probably because respondents conclude that System IV will meet their needs more than System I. Moreover, evidence seems to indicate that successful large-scale projects in fact require substantial doses of System IV philosophy and relationships (e.g., Sayles and Chandler, 1971; Walton, 1969). Finally, System IV more closely approaches a variety of ethical guides associated with openness, participation, and employee freedom than do Systems I, II, or III (Golembiewski, 1965, 1993). See also the discussions of regenerative interaction (e.g., pp. 110–111).
Care is necessary to avoid overrepresenting the narrative above. Consider the “success” of the efforts above. Several indicators imply that expected things did happen.

Changes in the Group Behavior Inventory, overall, were not only expected but imply that a range of more effective relationships among senior staff were developed and maintained over time.

Changes in the Likert Profile, overall, were not only expected but also moved in the direction of presumptively greater effectiveness of the organization level reporting to senior staff. (e.g., Likert, 1977)

Interviews with both directors and managers, overall, indicated that most personal definitions of “success” were met.

But the efforts sketched above are tethered short in significant senses. They do not report a “complete change”; rather, they seek to sketch the character and some of the consequences of initial steps to develop a team consciousness and to imprint a “cultural pattern.” Relatedly, the results of the interventions cannot be expressed as direct management or administrative results. These results must become manifest in the bottom line—the efficient and effective construction of the MARTA system.

NOTE

For those interested in such matters, an important technical issue inheres in how these difference-scores are calculated. On the point, see Dennis P. Slevin, “Observations on the invalid scoring algorithm of “NASA” and similar consent tasks,” Group and Organization Studies, 3 (4): pp. 497–507.

REFERENCES


Illustrating Replications


Saving Conceptual Shortfalls from Themselves
Enriching New Public Management (NPM) as Exemplar

OD values and methods also can be easily tested for usefulness in multiple ways that amount to augmenting shortfalls in popular managerial concepts. This chapter and its two companions focus on replication as a case in point of how OD can do better via replication.

The focus here is on a second kind of replication, loosely conceived. Consider Total Quality Management (TQM), for example. It is long on attractive ideals, but short on ways-and-means to implement those ideals (e.g., Golembiewski, 2000). Illustratively, Deming (1982) proudly announced that if you gave him a cooperative group, he could give you a program of continuous attention to quality. One of his shortfalls is, of course, that the issue might just be the creation of cooperative groups at work. About this crucial point, Deming is only incompletely helpful, as in rightly urging the elimination of fear from the workplace, against which workers would reasonably seek to protect themselves via defensive interpersonal and group relationships—the major consequences being alienation from management and an associated lack of cooperation with management.

The focus here is on another widely accepted managerial approach, which has both substantial prominence these days but also major shortfalls—New Public Management, or NPM. For those interested, a substantial extension of the present approach is available elsewhere (Golembiewski, 2000).

Conceptually, this chapter and its associated pair—Chapters 4 through 6—pose no problems. Given substantial OD success rates, the present burden of proof is light. If we can in effect replicate OD by adding depth or detail to other
major systems of thought or application, so also will we contribute to greater success rates in OD, at least in two senses. That is, such efforts will clarify and extend OD motives; and such extensions in turn will also help tie to OD theory and experience the bodies of thought that have developed with costly independence from OD. Directly, OD could help remedy such shortfalls in specific areas, and vice versa.

CACOPHONIES IN THE CONTEMPORARY CHORUS:
ABOUT CHANGE AT PUBLIC WORKSITES, AS CONTRASTED WITH SOME STRAIGHT TALK FROM A PLANNED CHANGE PERSPECTIVE*

New Public Management (henceforth NPM) has coalesced into a movement in a short period of time, virtually worldwide. Thus, inter alia, we hear about the allegedly new focus on the “customers” of public services, which are to be provided by “public intrapreneurs” as well as by cadres of employees at all levels who are “empowered.” And so on and on—through the conventional organizational litany including cross-training, total quality management, and performance measurement, and eventuating in strategic planning. These emphases make for a pleasing, even convincing, organizational libretto.

If the “chorus” proclaiming the NPM libretto is ubiquitous as well as insistent, however, their chanting is at times often loosely coupled, curiously directed, and even contradictory—at times so much so as to alert one’s native suspicion about what forces are really at work. Hence, the reference here to the chorus and also the cacophonies this chapter detects in NPM’s ardent vocalizing. This reflects our judgment that, in equal measure, NPM combines ubiquity, makes too much of some useful things, fails to reconcile flagrant diversities, as well as raises issues at sixes-and-sevens characteristic of any view of NPM beyond the superficial.

This set of points at cross-purposes implies serious inadequacies and shortfalls in advertised effects, which this chapter will begin to remedy. How? Going beyond pessimism that NPM can “walk its talk,” several emphases will at once help explain how NPM was all but predestined to experience serious inadequacies or shortfalls, as well as prescribe in part how NPM might rise above these limitations. And how? These several reconstructive emphases include some sense of how NPM can realize its fuller potential by pursuing “learning organization” features, as well as in related particulars giving attention to appropriate interaction guidelines and structural arrangements.

* The original version of this chapter was prepared for presentation at 1999 Annual Conference, American Society for Public Administration, Orlando, Florida, April 11, by Robert T. Golembiewski, Eran Vigoda, and Ben-chu Sun.
Four emphases relate to these critical-cum-constructive ambitions. In preview, NPM

- Seldom even attempted detailing a useful approach to applications;
- Typically neglected systemic or millieu characteristics within which applications can occur;
- Usually did not specify a useful front-load in designs: that is, training in values, attitudes, and interaction skills that would facilitate developing a “cultural preparedness” for appropriate applications; and
- Seldom specified supportive structural/managerial arrangements.

This chapter takes a direct approach to describing New Public Management, both chorus and cacophonies. Introductory attention goes to NPM as a “liberation” of theory and practice from the classic conservatisms of Public Administration. Then, four limitations of this NPM chorus will be detailed, and this quartet of cacophonies also imply ways to enhance NPM applications. More narrowly, these limitations also justify a stark warning against overselling.

THE CHORUS: NEW PUBLIC MANAGEMENT AS “LIBERATION”

Since the early 1980s, much work conducted in public administration theory and practice expresses dissatisfaction with conventional conservatisms. This “liberation” is today often packaged as the New Public Management, or NPM. Often loosely drawing on the experience of the business/industrialized/private sectors, a more “demanding” attitude has developed toward the activities, responsibilities, and productivity of public organizations. NPM reflects a remarkable degree of consensus, given anybody’s criteria, and perhaps because of the common sponsorship of political leadership and opinion makers of various countries (Maor, 1999, p. 15). This suggests widespread problems—indeed, all but worldwide problems—that were both pressing and obvious. However, supporting technical resources and literatures are nowhere near as definite.

What are the roots of NPM, and in what ways (if any) are they actually new? Several theoretical fundaments, as well as practical issues, help answer this compound question. The first, and probably the most basic, stimulus for NPM emerges from the distinction between two central terms—administration and management. Since the late 1880s, the monopoly on the term administration has been in the hands of political scientists. Scholars like Goodnow and Wilson perceived public administration as a separate and unique discipline that should consist of independent theory as well as practical skills and methods, even if the founding fathers were all but completely unhelpful about the latter. Conservative “administration” tended to analyze the operation of large public bureaucratic systems as well as other governmental processes aimed at policy, with an emphasis
on narrow “implementation” of legislatively made and judicially informed “policy.” On the other hand, and often loosely defined, “management” refers to the more or less specific practice of empowering people and groups in various social environments, as well as to the exercise of handling manifold resources to maximize efficiency and effectiveness in the process of producing goods and services. In sum, the term “management” refers to all arenas for “getting things done through people,” and was used widely by researchers and practitioners—in organizational psychology, in business studies, and so on—with less concern about jurisdictional issues concerning policy versus administration or implementation.

A few researchers did straddle administration and management—for example, Simon, March, and your author, among others. In different ways, such students intended to build on the established public administration, but also to extend and enrich it, if in different ways. Simon’s critique of classic public administration begins over 50 years ago (e.g., Simon, 1946), and focused on decision-making and its premises; March in various ways enriched this basic work while resisting the facile extrapolations that came into favor as “rational voluntary action” models favored by “free market” theorists (e.g., March, 1992). And Golembiewski, here as well as elsewhere, sought to tie classic concerns in public administration into value-loaded applications of planned change (e.g., Golembiewski, 1977a, 1977b, 1985, 1995a, 1995b). Such approaches had their impacts, greater or lesser, but they tended to be seen as beside the point of a basic polarization, often polemical, between adherents of administration versus policy-making, as broadly defined.

The growing dominance of “management”—as in NPM—rests on a kind of analytic and practical sex appeal, among other features. That is to say, relying on an extensive survey of public sector research in America, Garson and Overman (1983, p. 275) argue that this increasing popularity of NPM was due to the more virile and expansionist connotations of the term management as compared with administration. Over the years, a growing cadre of observers perceived Public Administration as an old and declining discipline, as increasingly unable to provide practical or theoretical guidance concerning burgeoning problems. This was especially the case for those who, and most did, see administration in sharp contrast with “policy.” Given other similarly oriented features, the effects were clear, though eclectic in cases. For example, many schools of public administration during the 1980s and 1990s were transmuted into schools of public management. In the quest for alternative ideas, management philosophy (even mythology!) and research were proposed as a source of new and refreshing perspectives.

Consistently, Perry and Kraemer (1983) urged that an influx of new ideas and methods from the field of public management into the venerable administrative science is nothing but essential and natural. This builds on Rainey’s (1990, p.157) claim that such a reorientation rests not only on management’s attributed successes but also on the growing unpopularity of government during the 1960s.
and 1970s. And Ott, Hyde and Shafritz (1991, p.1) add that “public management” holds out more promise than “public administration” of an affinity with powerful analytic and implementation tools, techniques, knowledge, and skills that can be used to turn ideas and policy into successful programs of action.

However, the realities were not so simple. Basically, in fact, “the managerialism” of figures like Barnard and Drucker was already beginning to fade as seen by a growing cadre of observers (e.g., Scott, 1992). In obvious senses, indeed, that managerialism is limited enough, even on initial analysis: for example, the claims of benefits were exaggerated, the doctrine rested on a pastiche of values that gave more color than guidance or support, and managerialism was short on theoretic specifics as well as on tools and techniques (e.g., Golembiewski, 1995b).

Given both common motivation and inadequate closure, it comes as no surprise that, during the past two decades, many definitions for NPM were suggested, and they broadly compare with this older view of Garson and Overman (1983, p.278): for these careful observers, NPM is “an interdisciplinary study of the generic aspects of administration . . . a blend of the planning, organizing, and controlling functions of management with the management of human, financial, physical, information and political resources.” To expand on this general view, Lynn proposes (1996, pp. 38–39) that six growing differences exist between public administration and public management that especially recommend the latter as a new field of study and practice. These differences include:

1. The inclusion of general management functions such as planning, organizing, control, and evaluation in lieu of discussion of social values and conflicts of bureaucracy and democracy;
2. An instrumental orientation favoring criteria of economy and efficiency in lieu of equity or responsiveness associated with political salience;
3. A pragmatic focus on mid-level managers in lieu of the focus on political or policy elites;
4. A tendency to consider management as generic, or at least to minimize the differences between public and private sectors in order to learn from each, and especially from the latter;
5. A singular focus on the organization with external relations treated in the same rational manner as internal operations, as contrasted with a fixation on laws, institutions and political bureaucratic processes; and
6. A strong philosophical link with scientific management traditions and business research, as distinguished from close ties to political science or sociology.

While (to a far lesser degree) the emergence of NPM is also frequently associated with the increasing penetration of positivist behavioral science on the study of politics and government (e.g., Lynn, 1996, pp. 5-6), several practical
features also deserve attention (Golembiewski, 1995b). Difficulties in policy making and policy implementation faced many countries beginning with the 1970s—more and more to do, and a probably shorter supply of resources to do it with. These practical difficulties are viewed today as an important trigger for the evolution of NPM. Reviewing two recent books on NPM (Aucoin, 1995; and Boston, Martin, Pailot, and Walsh, 1996), Khademian (1998, p. 269) argues that American and Westminster advocates often find considerable common ground in explaining why NPM reforms are necessary. Elegantly, for example, Aucoin summarizes a trinity of broadly based challenges with which Western democracies have coped, and with which they will probably continue to struggle in the future, partly through the management reform that this trinity encompasses:

1. Growing demands for restraint in public sector spending;
2. Increasing cynicism regarding government bureaucracies’ responsiveness to citizen concerns and political authority, as well as dissatisfaction with program effectiveness; and
3. An international, market-driven economy that often dominates in domestic policy.

In sum, these challenges led many Western governments—in America, Britain, New Zealand, Canada, and elsewhere—to the view that firm reforms and changes in the public service should be made. How and why did this diffusion occur? The impact on states of international banking, as aided and abetted by the globalization of the consulting firm played two dominant roles. The banks gained leverage via making loans, and the consultants supplied guidance and counsel in burgeoning measure indicated by huge rates of growth throughout the world, rates that approximated hundreds of percent during recent decades.

**A Double “Liberation”**

Viewed from another perspective, help seemed available for the required transformation. Scholars agree today that at least some of the accumulated wisdom of the private sector is transferable to the public sector (Golembiewski, 1985, 1995b; Pollitt, 1988; Smith, 1993), even if the claims are at times more hyperbolic than justified (e.g., Golembiewski, 1995b, pp. 23–26). In an attempt to “liberate” the public sector from its old conservative image and tedious practices, as well as to “liberate” energies for motivating associated changes, NPM was advanced as not only a relevant but also a promising banner under which to assemble the new assault on growing challenges. For example, NPM has strongly advocated the implementation of specific performance indicators used in private organizations to create performance-based cultures, reinforced by matching compensation strategies. NPM has recommended that similar indicators be applied in the public
Enriching NPM as Exemplar

sector (e.g., Smith, 1993; Carter, 1989) since they can function as milestones on the way to the greater efficiency and effectiveness of public agencies. Broadly, the goal is to apply market-like forces in the public sector (e.g., Golembiewski, 1995b).

In a reinforcing way, heightening citizens’ attention to the performance of public services was suggested as a core element of NPM, since that model proposes to increase the political pressure placed on elected and appointed public officials, thereby enhancing both managerial and allocative efficiency. Many scholars who advocate NPM associate this process of public accountability with energized citizen stakeholders, on the general order of shareholders in the business model (Smith, 1993). As in the private sector, increasing external-related forces can have a profound impact on control mechanisms internal to organizations, as public servants become more sensitive to how they are perceived in fulfilling their duties as well as responding to a new consciousness about how even more highly committed they have to become to serve their public customers/stakeholders.

In view of the past and looking toward the future, Lynn (1996, p. 231) suggests that NPM of the late 1990s had three constructive legacies for the field of public administration, for democratic theory as well as practice. These include:

1. A stronger emphasis on performance-motivated administration and inclusion in the administrative canon of performance-oriented institutional arrangements, structural forms, and managerial doctrines fitted to particular contexts;
2. An international dialogue on, and a stronger comparative dimension of, state design and administrative reform; and
3. The integrated use of economic, sociological, social-psychological, and other advanced conceptual models as well as heuristics in the study of public institutions and management, with the potential to strengthen the field’s scholarship and the possibilities for theory-grounded practice.

This chapter goes only a little way in assessing the solidity of the foundations that NPM provides for this trio of legacies. The analysis below begins with a sampler of the best/good practices “liberated” by NPM; that analysis moves on to discuss four categories of shortfalls, what NPM prescribes to amalgamate into the theory and practice of the workings of the public sector.

This analytic journey leaves your author underwhelmed. In effect, the best/good practices deal with the themes favored by the NPM chorus; and the four shortfalls introduce some ways to moderate the cacophonies in that libretto, in a manner of speaking. As usual, we can do a lot better in planned change than NPM does.
Sampler of Best/Good Practices

Major forces—not always oriented in the same or even consistent directions—birthed a new emphasis in the public sector on similar ways and means, virtually worldwide and in a short period of time (e.g., Pollitt, 1988; Boston, et al., 1996; Charih and Rouillard, 1997). While an earlier age spoke of “principles” of broad or even universal application, this new age urged a catalog of variously preferred “practices.” At first, these tended to be labeled “best” practices but, gradually, most proponents seemed to settle for the merely “good.” In the present view, neither descriptor applies.

Table 5.1 presents a selective inventory of these best/good practices, which tended to be accepted—usually in bits and pieces rather than as a whole—by political elites, worldwide. Broadly, these practices embodied new or novel “institutional controls,” and their general acceptance by policy-makers raised real issues with and for students of government (e.g., Radin, 1997).

Several caveats apply to Table 5.1. First, although the several techniques listed had substantial histories in the management sciences, the hook-line-and-sinker acceptance characteristic of their political reception overdoes it. Typically, the technical literature was full of contingencies and qualifications, and tests of the whole NPM package plus supportive/contraindicating situational features simply does not exist. This chapter tries to do better in this regard.

Second, relevant applications in the management literature were typically bounded by specific contexts and technologies-cum-values for application. Differences in such significant fundamentals were steamrollered in NPM’s broad political acceptance, as the summary perspectives below imply. For example, see Management by Objectives, and especially in connection with situational features that predispose (or contraindicate) successful application (Golembiewski, 1979, vol. 2, pp. 166–179). The NPM literature contains little of those bounding conditions.

### Table 5.1 A Selection of Major NPM Techniques and Approaches

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job rotation</td>
<td>Planned movement among activities to meet many human and organizational purposes; to build some variety into the job, to permit building-up to and then down-from the most intense activities—as for air traffic controllers—and so on (e.g., Golembiewski, 1995a, 1995b)</td>
</tr>
<tr>
<td>Cross-training</td>
<td>Preparing individuals to perform activities so as to enhance variety and skills utilized, to permit managerial flexibilities, and so on (e.g., National Performance Review, 1993)</td>
</tr>
<tr>
<td>Job enrichment</td>
<td>To build content into jobs so as to increase employee mastery at work, to simplify organizational arrangements associated with inspecting work and measuring performance (e.g., Carter, 1989; Golembiewski, 1962); this “vertical loading” contrasted sharply with mere job enlargement, which merely adds activities to a job profile</td>
</tr>
</tbody>
</table>
### TABLE 5.1 (continued)

**Revisions in position classification:** e.g., via “broad band” approaches to facilitate job enrichment and job rotation or, looked at another way, to counteract the impetus in bureaucratic methods to fragment jobs and work into narrow packages

**Performance measurement and appraisal:** include a wide variety of policies, procedures, and techniques to induce motivational and competitive forces to stimulate production and quality e.g., Management by Objectives, or MBO (Golembiewski, 1979, pp. 168–179; Smith, 1993; Pollitt, 1988), with a focus on the customer

**Total Quality Management:** values-with-technology to continuously improve quality of output, at all levels of organization but especially at operating levels: comes in various forms, and is often associated with a “close to the customer” emphasis, considering both external customers (e.g., Pollitt, 1988) as well as internal customers of, say, staff groups (e.g., Barzelay, 1992; Golembiewski, 1995b, pp. 76–83).

**Project teams:** an overlay of bureaucratic structures to provide alternative memberships/loyalties, with detailed arrangements usually favoring the dominance of bureaucratic vs. project forces (Golembiewski, 1967)

**Managerial credos or culture statements:** to provide normative guidance or “templates” for action-taking that can facilitate loose-tight management— “loose” concerning discretion about how to meet “tight” objectives set centrally (e.g., Golembiewski and Carrigan, 1970)

**Strategic planning:** the complex set of activities to create overall agreement about major objectives, contingencies, and operational measures of performance to provide detailed but still “loose” templates for decision-making and action-taking, an approach that dominated business management since post–World War II days, and much later became prominent in public and voluntary sectors (e.g., Bryson, 1988; Golembiewski, 1989, 1995b)

**Autonomous groups:** units at operating levels containing all/most of the activities required for a complete flow of work (e.g., Hanna, 1988), as contrasted with the monospecialization prescribed by the bureaucratic model

**Mission statements:** in part due to problems with strategic planning, coupled with the often-powerful need to counteract divisive forces generated by bureaucratic structures and policies, such normative overlays became common at many levels of organization (e.g., Golembiewski, 1999b)

**Divisional structural models:** see the section below, which focuses on this post-bureaucratic variant at executive or mid-management levels

**Matrix structural models:** a complex approach to organizing work that combines multiple authority structures such as those related to product and geographic area (e.g., Davis and Lawrence, 1977; Teasley and Reddy, 1981)
Third, supporters of several of the entries in Table 5.1 were beginning to lose steam, as via significant reinterpretation, even as their acceptance in NPM was gaining momentum. This was true of strategic planning, for example, which started to get powerful support from PA mainliners (e.g., Bryson, 1988), even as major reevaluations of the approach were being proposed by business observers (e.g., Mintzberg, 1994) as well as a few PAers (e.g., Golembiewski, 1989).

Fourth, NPM applications did not uniformly occur across the full range of items in Table 5.1. More or less, applications were most common concerning the items leading that list, while divisional and matrix models got little notice, especially in the United States and with the major exception of Great Britain (Pollitt, 1990). This is unfortunate, as the last section below proposes. Only rare exceptions to this generalization exist (e.g., Posner and Rothstein, 1994; Miller and Golembiewski, 1999). Looked at in another way, NPM as reflected in National Performance Review sources emphasizes the first several themes in Table 5.1, almost to the exclusion of themes introduced later in the list.

FOUR CACOPHONIES IN THE NPM CHORUS

Elaborating the dissonances in this NPM “chorus” could take several routes, but here four related emphases receive attention. Major cacophonies derive from NPM’s lack of a model for guiding applications; inadequate situational awareness concerning good/poor fits of specific situations to specific “good practices”; incomplete information about how to increase an organization’s “cultural preparedness” for NPM; and undercutting orientations to structural features congruent with NPM.

Best/Good Practices Without Model for Applications

With few exceptions, NPM focuses on practices or policies that are deemed “better” or “best” without providing a road map of how to get there, from whatever “here” characterizes a specific worksite. The numerous publications of the National Performance Review clearly indict themselves by their all-but-complete silence on this elemental but significant point. The common assumption seems to be a kind of tacit equifinality—you can get there, wherever you start from and however you proceed.

Perhaps the most egregious example of this kind goes back several decades to the ill-fated prescription of Programming-Planning-Budgeting System, or PPBS. The literature is all but barren on how to get there, and even on how PPBS came to be, with that management history having powerful but neglected implications for implementation (Golembiewski and Scott, 1989).

What helps explain this curious unconcern about how to “get there”? Several central political and technical issues provide some perspective. By blending
aspects of aging memories with some historic facts, for example, Golembiewski and Scott (1989) provide a “conjectural footnote” on the slow development and sudden as well as ill-fated diffusion of PPBS throughout the U.S. federal government. By a process that includes serendipity and the need to find an application for an early conceptualization of what was then advanced data processing, President Johnson came to learn of a substantial technical achievement: the existence of a matrix for a Latin nation of total spending by all federal agencies for all projects/items/costs. The president was delighted by this “best/good practice” for it gave him what he wanted: control. And it appeared this would be low-cost and uncomplicated. The early PPBS form provided information in a manipulable form never before available which, if the practice were diffused, would provide the central controls that Johnson saw as useful reins on the accelerated guns-and-butter pressures on spending that he saw in our immediate future. In effect, PPBS promised to provide a basis for new and comprehensive institutional controls (see Radin, 1997).

In short, this was another case of a failure-in-linkage between hierarchy and specialty. With no check of the facts to hedge enthusiasm—neither the president nor the cabinet member from State who delivered the output of this early PPBS variant had even a vague idea about how the matrices were assembled—the innovation was mandated for diffusion throughout the entire federal government. This lack of concern about the model for application generally doomed PPBS variants at the federal level, if only because the model of change that had worked for State was not quickly applicable in most federal locations, if at all. Several of the situational features that should have delayed President Johnson, or perhaps even deterred several presidents, are prominent (e.g., Golembiewski and Scott, 1989):

The relevant financial data were tightly held by each separate federal agency, and sharing them usually would have had career-damaging consequences for agency employees, and perhaps especially such sharing with the president’s office;
The Latin country in question was not of main-line concern for most federal agencies, and employees had unusual latitude to meet one another and even to collaborate;
State Department personnel were increasingly influenced by Organization Development relying on T-grouping and its emphasis on trust-building and collaboration (e.g., Harmon, 1988), as rooted in a clear analog of regenerative interaction;
The State “desk” in this case was ecumenical in spreading that OD gospel to personnel from other federal agencies in that Latin country; and
The financial matrices were a tangible result of 2 to 3 years of “cultural preparation” to share in the local PPBS-like effort, aided and abetted by little or no knowledge/oversight by Washington.
Such conditions would have been difficult to recreate on a federalwide basis, if not impossible, but awareness of them did not even touch the consciousness of federal decision-makers at the level of the office of the president.

As with NPM, action-taking missed the point: that a “good” or “best” technique could not be considered independent of the context of application or of a model of theories/techniques to support learning about adopting that innovation.

The general point has all-but-universal applicability. Absent such cultural pre-work for PPBS, its life-chances were all but doomed systemwide and would not have been very favorable in any locus.

To be sure, the present example goes back about 30 years in time. But it is reasonable to think of this case both as an early exemplar of the NPM orientation and as sharing basic limitations with later variants.

This common NPM neglect of a model for application—as in NPR—is a double-dip deficit, as it were. For example, the success rates of OD applications like those utilized in State are substantial, even formidable, and about the same in public as well as in business contexts. Relevant data have been widely reported—for example, a survey of 16 separate batches of OD evaluations places success rates at about the 70–75 percent level (Golembiewski, 1998). Here, the full range of interventions is involved: from OD applications to individuals as well as to technostructural designs for large systems. Relatedly, Quality of Working Life interventions also report substantial success rates (e.g., Golembiewski and Sun, 1990); with QWL designs emphasizing operating level interventions whereas OD has a bias toward managerial or executive levels. Of special relevance—for both OD as well as QWL success rates—are these three factors:

- Applications in the public sector compare favorably with business, and in cases surpass the latter success rates;
- Self-reports as well as objective criteria generate similar patterns of success rates (e.g., Nicholas, 1982; Golembiewski and Sun, 1999); and
- Thousands of applications are involved.

More broadly, see Chapters 1 and 2 above.

**Best/Good Practices Without Milieu-Specificity**

A related shortfall further blunts the NPM thrust: the literature typically provides little or an inadequate sense of the milieu or context within which an application occurred. To put it positively, a practice is best/good in reference only to some specific “where” concerning which the “degree of fit” is hugely consequential. The point is obvious in many cases, as when subsistence farmers in the well-known Puebla experiment were taught new tricks with novel seeds for planting maize as well as about new technologies for cultivating that corn. Production skyrocketed, but no ways were provided to get the unprecedented crop to markets:
no improvements were made in transit and distribution. This describes a change effort that failed the more it succeeded (e.g., Golembiewski, 1991, pp. 210-211).

That “there” features are crucial to successful applications at a specific “here” of good/best practices is also usefully illustrated in detail. Many examples are available, beginning early (e.g., Fleischmann, 1953) and including contemporary cases, but one case must suffice here. The immediate context of the intervention is an operating locus in a business plant: broadly, the locus suffered from systemic inertia and personal disaffection, if not alienation. A best/good practices prescription was applied. Operating jobs were “enriched” and cross-training was instituted, with a consequent “empowerment” of the employees and a reorientation of the job of the immediate supervisor from direct oversight to facilitation and trouble-shooting. The predicted outcomes are direct: individual needs and systemic concerns would be better provided for, as in the job rotation and enrichment available to individuals through cross-training, as well as in the derivative flexibility available for systemic purposes.

Those predicted outcomes did occur at a specific site, but they did not last. Indeed, the worksite situation deteriorated from both individual and systemic perspectives. In short, that “where” provided an inhospitable locus for the best/good practices permitted by the well-intentioned interventionists. What happened? Full details cannot be recited, but a few bullets provide an instructive summary:

The structural change sharply reduced the numbers of first-line supervisors and middle managers who were needed, but no forethought was given to the specifics of how the required cut-backs would be made;

The new supervisors lost aspects of the old job associated with monitoring compliance, long ago built into a system yielding point totals for rating status and performance; however, no changes were made in that rating system to reflect monitoring by inducing commitment (e.g., Gortner, 1997) required by the new structure; and

Various attractive personnel upgrades were keyed to these point totals, but paradoxical consequences followed the structural change—for example, a supervisor doing a mediocre job under the old system was better off than a supervisor doing an excellent job under the new structure, absent changes in the rating system, which were not made.

In sum, the best/good practices were a poor fit with the larger organizational context. No wonder about the worsening conditions in the status quo ante, in the absence of a reform of traditional institutional controls like the supervisory rating system. Even initially successful applications can be expected to experience fade-out effects which, from important perspectives, can be more frustrating and even deflating than flat-out failures.
Best/Good Practices Without Front-Load Training

All but universally, further, NPM pays no attention to what might be called the “cultural preparedness” of host agencies and the personnel. This organizational equivalent of spontaneous conception seems too hopeful, at least in the vast majority of cases. Some kind of facilitative training seems useful, if not necessary, in most cases to facilitate transformations. In about equal measure, such pre-work would relate to inducing an appropriate new culture as well as to developing necessary skills and attitudes. For a typical OD example, see Chapter 4.

In OD applications, for example, this “cultural preparedness” is often approached via planned changes in the interaction between individuals and in groups. Table 5.2 sketches a typical schema underlaying such pre-work on cultural preparedness: certain macro-level values are emphasized; they are reinforced by micro-level practice with associated values/skills; and the goal is to serve major practical purposes that facilitate many best/good practices. For example, the “regenerative interaction” illustrated in Table 5.2 clearly would facilitate a program of cross-training or job rotation (Golembiewski, 1995a). Readers can easily work out the complications for NPM techniques such as cross-training or job rotation when interaction is “degenerative” (i.e., openness, owning, and trust are low, and when risk is high). Good practice or not, a degenerative interaction system will provide little support for it.

Two basic reasons particularly encourage a front-loading of designs for change at worksites with degenerative interaction. Directly, major aspects of degenerative interaction exist in many worksites, and they often will undercut applications of best/good practices. Attractively, success rates with designs inducing regenerative interaction are in the 75-plus percent range (e.g., Golembiewski and Sun, 1990, esp. pp. 11–29; 1998b).

However, not all useful planned change must be preceded by specific pre-work just before each interaction. Thus, “work out” and “future search” (e.g., Ashkenas, et al., 1995) can be useful designs, and neither places any overt emphasis on pretraining in interaction. The same can be said of “appreciative inquiry” (e.g., Golembiewski, 1999a), but this class of designs raises issues far beyond the present scope. For example, the required skills in GE may have existed applications of work-out. Some change programs explicitly propose that all required skills are typically available in the target class. This is alleged for future conferences, for example Emery (1976), and appreciative inquiry.

Best/Good Practices Along With Undercutting Bureaucratic Structure

With few exceptions, NPM variants are typically presented as add-ons to a basic bureaucratic structure, as in all of the National Performance Review literature that your author has reviewed, which is a formidable mass of paper (e.g., National
### Table 5.2: Components of A Value-Loaded Enhancement of Cultural Preparedness for NPM Applications

<table>
<thead>
<tr>
<th>Macro-Level Values of OD</th>
<th>Associated Micro-Level Values/Skills: Regenerative Interaction</th>
<th>Probable Consequences of Regenerative Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>An attitude of inquiry, including a hypothetical spirit and experimentation</td>
<td>High openness, or “telling it like it is” High owning, or psychological acceptance of one’s ideal/feelings, and acceptance of responsibility for them and their consequences</td>
<td>Real issues are likely to be isolated and resolved Communication and decision-making processes are likely to be effective</td>
</tr>
<tr>
<td>Expanded consciousness and recognition of choice</td>
<td>Low risk, or objective threat in the environment High trust, or confidence that “things will work out”</td>
<td>The amount of “unfinished business” between individuals and within groups probably will not increase, or may even decrease, leaving a greater proportion of energies available for task</td>
</tr>
<tr>
<td>A collaborative concept of authority including the open resolution of conflict and a problem-solving orientation</td>
<td></td>
<td>Participants are less likely to be dependent and overcautious Organization folkways and cultures are less likely to reinforce features of degenerative interaction, which are present in all systems</td>
</tr>
<tr>
<td>An emphasis on mutual helping relationships to reflect our social nature and connectedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>An emphasis on authenticity in expressing feelings and dealing with their effects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a For supporting citations, see Golembiewski, 1993, pp. 55–72, 92, 163–169.*
This persisting coupling has long bemused the present author, and even baffled him, both early and late (e.g., Golembiewski, 1962; 1995a). Nonetheless, this curious coupling largely remains, even as major inroads on it have been made lately, and especially but not exclusively in business (e.g., Golembiewski, 1995b). Whatever the case, it has long been clear that the bureaucratic model suffers from serious disabilities and deficiencies, over the generality of cases. To illustrate, conventional notions about structuring work rest on inadequate supports—logical (e.g., Simon, 1946), normative and methodological (e.g., Coker, 1992), as well as behavioral (e.g., Golembiewski, 1965, 1993), among other classes of deficiencies. Perhaps paramountly, the bureaucratic model is non-democratic, or even antidemocratic, which does not seem to trouble most observers, even well-placed ones such as Dahl (1973), who urges that democratic political systems require autocratic or non-democratic administrative/managerial systems. One could put their view as follows: Autocracy at work is the price we pay for democracy after hours.

In any case, the acceptance of the bureaucratic model has multiple roots, no doubt (e.g., Golembiewski, 1985, pp. 143–225, 1995b), and it has often been challenged but never substantially supplanted. Indeed, new and major support now exists for that model (e.g., Goodsell, 1983, 1985, 1994), which tradition of support also goes back to the earliest days in the history of Public Administration. Recall this doggerel:

> Over forms of structure,
> let fools contest;
> Whatever is Best Administered, is best.

The specific source escapes the author. Could it have been Woodrow Wilson himself? Whoever the author, he or she has much company. Better said, without a doubt, only rare PAers argue cogently that “organization structure matters,” (Moon, 1999), while providing empirical research about the details for both bureaucratic and alternative structures. Your author has tried (e.g., Golembiewski, 1965, 1993, 1988, 1995).

How Structures Can Make Differences

From your author’s point of view, the couplet above is bad managerial advice as well as pathetic poetry; and even brief illustrations will highlight the serious cost of NPM’s failure—and especially in American variants like National Performance Review—to sharply separate itself from bureaucratic structures and their associated policies and procedures. Figure 5.1 will be used in two reinforcing ways to illustrate the managerial point. Thus, some major differentiating characteristics of two alternative structural forms will be detailed there. And then, discussion will illustrate how the techniques and approaches in Table 5.1 are better approached under the post-bureaucratic structure in Figure 5.1, while also being
made less necessary by that model. This approach is harshly selective, of course. Most commentators require a larger number of ideal types, hybrids, or archetypes (e.g., Miller and Friesen, 1984; Mintzberg, 1994) for comprehensive comparisons. The two models chosen here relate to most of the cases usually observed, which provides support enough for the present approach.

As a preliminary, note only that the GMU in Figure 5.1 refers, in a whimsical turn-of-label, to the Golembiewski managerial unit. The GMU is defined broadly as that portion of an organization presided over by an authoritative decision-maker who can make reasonable decisions about a total flow of work.

**Different Characteristics of Two Structures.** Figure 5.1 distinguishes two alternative structural forms with significantly different characteristics that have major practical relevance. If this were a fuller analysis, other structural models would be useful—networks (e.g., Gooden, 1998), matrix (e.g., Davis and Lawrence, 1977), and so on.

Carew and his associates (1977) provide a dynamic view of comprehensive differences, but here a summary list of contrasts about Figure 5.1 must suffice:

- **Structure A** departmentalizes around like or similar activities, whereas **B** focuses on a relatively complete flow of work involved in caring for clients in difficulty, while reserving Facility and Personnel to units outside each Youth Service Team (YST).
- Structure A serves the needs of the separate functions, whereas B seeks to accommodate a specific cluster of clients and to be “customer-friendly” by moving toward “one-stop shopping.”
Structure A must have a narrow span of control, or risk disorder as well as jurisdictional and control conflicts; B can have a very broad “span of control,” largely because of conveniences it offers in assessing the comparative performance of the several Youth Service Teams, as by comparing the percentages of recidivist clients. “Span of control” refers to the number of subordinates reporting to any integrative authority. That is to say, A structures will have a larger GMU, whereas B structures can have a sharply smaller GMU, as in Figure 5.1.

Thus, A structures will have many levels of control and communication and, hence, are “tall,” whereas B structures will be “flat.” A structures encompass probably fragmenting departments with no immediate incentive to resolve conflicts among them, whereas B encompasses several organizing around several comparable departments (YST), each of which has a continuing incentive to resolve internal differences lest a specific YST be disadvantaged in comparisons with the other YSTs.

Detailed comparisons of probable distinguishing characteristics of the two structures are conveniently available (Golembiewski, 1987, 1995b).

**Bureaucracy As Barrier to NPM.** Even as NPM tends to retain the bureaucratic model, paradoxically—as in the National Performance Review—that structure complicates and confounds approaching most of the techniques and approaches illustrated in Table 5.1. Or more precisely written, NPM applications seldom penetrate to the last two or three items in Table 5.1, which are here seen as the infrastructure most capable of supporting the approaches listed earlier in Table 5.1. Hence, the fade-out of NPM effects is probable even in those cases in which applications tend “to work” in the short run.

An alternative formulation of the generalization underlaying this subsection is also revealing. In general, A structures at once are in greater need of NPM variants and also make such adoptions difficult. This is an awkward combination. For many of the same reasons, B structures simplify adoptions of Table 5.1’s early-listed ways-and-means as well as stand in less need of them.

These two basic generalizations underlying this subsection can be supported by two related streams of analysis. In turn, these streams emphasize NPM prescriptions at cross-purposes; and brief illustrations of how bureaucratic structures have features that undercut NPM objectives.

**Your Author’s View**

This much having been written, what is your author’s take on the broad acceptance of NPM? I can be brief. NPM promised better consequences while retaining bureaucratic approaches and preserving traditional hierarchic controls and prerogatives. This must have seemed a very good deal to many authority figures.
NPM Prescriptions at Cross-Purposes

The clearest case in which NPM trips itself involves the simultaneous emphasis on performance and responsibility for results, while also retaining the bureaucratic structuring of work. As reference to Figure 5.1A will suggest, it is not possible to easily measure performance at the S-level in bureaucratic structures, basically because that effort involves complex technical/political issues associated with how many of A = how many of B = how many of C. Relatedly, restriction of output in bureaucratic models is easier than raising output, for an obvious reason: any single S-unit can generate major momentum to restrict output, while only all S-units acting collaboratively can increase output.

For quite-straightforward rationales (e.g., Golembiewski, 1995b), the alternative model in Figure 5.1B presents more managerially felicitous probabilities. Any S-unit there can raise output; and the performance of all S-units is directly comparable. To be sure, Figure 5.1 deals with a simple case, A + B + C yielding some product or service. But the general principle applies quite broadly.

Bureaucratic Structures as Directly Undercutting NPM Objectives

This second stream is clearly paradoxical, if not bizarre. Take job rotation as the simplest example. This straightforward effort toward greater managerial flexibility also can meet human needs for progressive mastery and variety, but Figure 5.1A structures pose major barriers to such rotation.

These are awkward dynamics. Directly, all or many functions will have to be involved to support a substantial rotation effort, and each of these functions can be a veto center. Certainly, no single S unit can undertake a work enriching program. Why? For openers, trust between the several Figure 5.1A departments is likely to be low, with we/they relationships being variously reinforced by the basic departmentation organized vertically around separate functions, which encourage competition for resources between the separate functions. And the large GMU usually will dilute the loyalties necessary to positively engage contributors to the...
full flow of work. As one consequence, In-Take in Figure 5.1A (for example) might be conflicted about rotating their better performers into Assessment or Placement, and solid employees might be hesitant about “moving in with an enemy.” The potential for required coercion in even such a modest application seems clear enough. Awkwardly, to put it another way, each department would likely claim credit for any apparent success; and all departments also most likely would strive to avoid responsibility for any failure. This sketches some volatile potentialities.

For Figure 5.1B structures, in contrast, any YST can independently field a job rotation program; employees would retain their organization location; and each YST would pay the costs as well as profit from any benefits of a program for which each team is fully responsible. Here, the potential is substantial for generating self-interested commitment, as well as for profiting from successful marshalling of that commitment to improve performance on a total flow of work. Any real success will show up quite directly in the comparative performance of any specific YST rotating jobs, as contrasted with a YST not doing so.

Similar contrasts suggest themselves when later-listed themes from Table 5.1 are considered—for example, strategic planning. Figure 5.1A structures suggest major barriers to strategic planning (e.g., Golembiewski, 1989,1999c). To illustrate, the vertical fragmentation characteristic of bureaucratic structures often will complicate strategic planning, and hence might increase reliance on what can be called Control I—control by command or coercion. Relatedly, Figure 5.1B structures have a higher potential for Control II—control by commitment within each YST, which would reinforce loose-tight executive control. “Looseness” could exist in how a YST accomplished its work, and “tightness” could characterize the missions and objectives to be accomplished. See Gortner et al. (1997) for further contrasts of these two concepts of control.

Finally, it is anybody’s guess what motivates NPM’s schizoid character. Elsewhere, I try to present the full range of possible rationales (e.g., Golembiewski, 1985); here, only the simplest form of an explanation must suffice. NPM supporters at once seemed to wish the benefits of good/best practices, but they were even more interested in preserving the elite-serving features of the bureaucratic model. That model was at once their major enemy and also their strongest commitment via self-interest.

**SUMMARY**

This chapter engages four basic tasks. First, it characterizes the political contexts in which New Public Management developed, virtually worldwide.

Second, the related urgencies and conveniences reflected in the NPM development neither rested on nor encouraged a satisfactory technical development of the several best/good practices associated with NPM.

This evaluation of NPM success rates does not encourage but neither does it exaggerate. Consider only the summary in Exhibit 5.1, which skims the essentials of perhaps the most ambitious evaluation of NPM applications.
EXHIBIT 5.1 Schema of summary of testing NPM applications (based on Thompson, 1999).
Interested readers can slog through the details, ideally with strong comparative attention to Chapters 1 through 3. It suffices to note here, in sum, that the present pointed critique of the NPM approach could be consistent with only such a paltry record of success, as is illustrated by Exhibit 5.1. The discussion above provides a catalog of reasons for the overall shortfall schematized in Exhibit 5.1.

Hence, third, the four cacophonies above intrude on the NPM chorus—no model for applications of best/good practices; the neglect of milieu- or situational-specificities; the general failure to enhance the “cultural preparedness” of hosts for NPM applications; and the curious retention of the bureaucratic model in most NPM variants, and by most NPM proponents. Exceptions do not occur frequently (e.g., Golembiewski, 1998a; Posner and Rothstein, 1994), although no generalization holds all the time and everywhere.

Fourth, the four cacophonies are not mere carping. Each “squawk,” if you will—in various ways, and typically with the support of substantial literatures—is presented along with a primer on how to better engage the full array of NPM ways-and-means, as characterized in Table 5.1.

To conclude, although our present purposes do not include comparing alternative models for change, a few analytic lines can be drawn in the sand concerning the realism of moving toward public-sector change in structures and interaction. Paramountly, the available literature (e.g., NPR, 1993; Light, 1997; Peters, 1996) pays little or no attention to technologies-cum-values for change. This is not only a glaring inadequacy but, in our view, also unnecessary. We support one such technology-cum-values—what is usually called Organization Development (OD) or Organization Development and Change (ODC). Associated designs have been applied broadly; their public or business applications are roughly proportional to the sizes of public versus private employment (e.g., Golembiewski, Proehl, and Sink, 1981, 1982); OD or ODC success rates are substantial, even formidable (e.g., Nicholas, 1982; Golembiewski, Proehl, and Sink, 1981, 1982; Golembiewski and Sun, 1990; Golembiewski, 1998b); and success rates in government are comparable to those in business (e.g., Golembiewski, 1998). These constitute a catalog of attractive features. Paramountly, this chapter illustrates two important ways in which OD can deal with the irony of ironies. Thus, OD can be tested by replications to other issues arenas such as NPM. Relatedly, success rates in OD proper can be enhanced by a growing knowledge of what potentials exist in other issue arenas to extend or enhance OD.

REFERENCES


Enriching NPM as Exemplar


Checking Downstream Progress, Years Later

Replication as a Stream of Events over Time

A direct if often anecdotal and impressionistic approach to replication in its third present meaning also is too seldom employed in OD. This involves going back to an application, or several of them, substantially after the fact to see what (if anything) remains of the wit and will once lavished on a site.

This approach to replication is neither the only, or the most desirable, approach: indeed, it has serious limitations. Paramountly, what are reasonable expectations about the staying power of even major changes? No one knows for sure. In the most energetic form of this kind of replication (Seashore and Bowers, 1979), revisits made over a decade and a half clearly reflected the robust signs of the change program that had been there originally (Marrow, Bowers, and Seashore, 1967).

Despite this overall record of general neglect only occasionally relieved by spectacular exceptions, replication in the form of longitudinal research designs has much to recommend it. Here, a generic research design permits testing for specific dimensions to be tested at several points in time, and T₁ versus Tₙ comparisons thus can provide a firm basis for judgments about persistence, fade-out, or even accelerated learning of the generic design employed.

Such replication-by-revisiting can be useful, if expectations are reasonably tethered. Thus, it is neither necessary nor convenient to plan a longitudinal design for each OD application, and revisits to sites of applications of generic designs provide a convenient way of assessing long-term effects of the genre. Moreover, revisits can be a way of working on still-unfinished business, or even nostalgic concerns; and revisits can help signal when reinforcement is useful or perhaps necessary.
Hence, given the dictum that nothing lasts forever, and that little of even the best survives beyond (let us say) intermediate terms like 3 to 5 years, this chapter reflects one such revisit to a site already made a part of this text in Chapter 4. The revisit postdates the early interventions by the better part of a decade during which major political and demographic changes beyond the original plans and assumptions not only had been experienced but were intensifying.

A CONCEPTUAL CONTEXT FOR IRONY II

Several senses in which this second meaning of replication occupies a useful niche within which success rates may be raised deserve note. First, such replications can help determine the “staying power” of particular approaches to change. This seems obvious enough, even though it is foolish to think anything lasts forever. Moreover, not every lasting change is a boon. Indeed, some changes can outlive their usefulness, and we would be fortunate if the staying power of some specific changes were limited.

Second, the persistence of a change provides one indication of the relevance and salience of the underlying model of change and its associated theory and experience. This feature should not be overstated, although it seems real enough. Thus, this second position implies this major caveat: persistence is revealing if, but only if, it can be shown that major conditions have not changed. Persistence with changes in basic conditions leaves us with a puzzle of being correct, but probably for the wrong reasons. That may be sufficient comfort for some but will not do here.

Third, probably in most cases, change is inevitably a blend of values and approaches for controlling events in nature. Here, as Machiavelli would put it, the intervenors will profit from having *buona fortuna* on their side. No doubt, consequently, the more replications under similar conditions, the more certain the expectations and interpretations. Most likely, we will be only partially able to define all or most relevant conditions, at either original intervention or during the intervals of replication. Both tentativeness and modesty are appropriate, then, but they should encourage replications rather than inhibit such efforts. Ideally, replications will result in growing sensitivity to conditions of application.

MARTA IN THE 1990s*: THE CHALLENGE CONTINUES

Overall, this chapter revisits the Metropolitan Atlanta Rapid Transit Authority, or MARTA, for the purpose of assessing the degree to which OD-induced values

and relationships persisted serviceably over the years. Interviews plus limited observation and survey data suggest two conclusions. Qualitatively, many features consistent with the early MARTA still existed, and the Authority remained an effective and attractive employer. However, analysis suggests some degradation of the interpersonal as well as group processes and relationships that OD efforts sought to induce a decade earlier as MARTA’s infrastructure.

Broadly, although the original balance nonetheless remained attractive and satisfying to employees, this balance probably may be inadequate for the choice-points that seem in MARTA’s short-term future. The overall hope is that regenerative interaction and OD values, if they persist, will be sufficient to induce an appropriately moving balance. However, this is by no means certain. MARTA faces an already shrinking passenger-base within its historic jurisdiction, and this implies a key choice-point for strategic planning and its subsequent implementation. So persistence of the OD normative structure seems to exist, but its conscious reinforcement seem prudent as MARTA enters a future that encourages new as well as enhanced managerial techniques and approaches.

Why revisit MARTA? In many senses, MARTA was a model OD application; some observers urged for a long time its OD application might be *sui generis* because of changes in federal spending for urban transport and because of numerous other idiosyncratic features over the intervening decade or so. Revisiting MARTA after a decade or so had strong technical support; in addition, revisiting MARTA was like going home for this author because of its significance to him and to the OD theory/experience he represented in it over a decade as external consultant.

Beginning in the mid-1970s, MARTA hosted a systemic application of Organization Development values-cum-technology, and this chapter constitutes another expression of the interest in that effort as it approximates a decade after the still-expanding rail/bus system went operational in the 1980s. See Chapter 4 for some institution-building details. Specifically, this chapter targets the persistence of that early shaping of MARTA by OD values (Golembiewski, 1979a, esp. vol. 1). In focusing on persistence of OD effects, this piece joins a slim literature (e.g., beginning with Seashore and Bowers, 1970), and especially as related to public-sector OD efforts (e.g., Harman, 1975; Bruce, 1992; Golembiewski, 1979b, 1985; and see Chapter 17 below).

Updating the MARTA experience—to see what has lasted, and what has decayed, as far as that is possible—faces real limitations. Much time had passed since my last formal observations in 1982–83, and this chasm requires real interpretive tentativeness. Moreover, time pressures largely limit this update (1991) to interviews with MARTA executives, which need reinforcing detail from all levels. Nevertheless, countervailing attractions encourage seeing what can be usefully done. Part of the motivation is personal, in that this author not only devoted much time to the development of the early MARTA, but also had anticipated,
along with the first general manager, that MARTA would experience major challenges as it moved beyond start-up (Golembiewski and Kiepper, 1983a, 1983b).

Four emphases detail this updating. In turn, attention is directed at:

- MARTA’s approach to using OD values and approaches to achieve its tough objectives and budget constraints;
- A brief introduction to the literature describing and evaluating MARTA;
- Some mid-1980s concerns about MARTA’s post-start-up future; and
- A major focus of attention to an overview of MARTA in the early 1990s, relying on interviews and survey items with top executives.

**“NOT JUST ANOTHER PUBLIC AGENCY”**

The aspirations reflected in the heading encapsulate the spirit of MARTA’s founding leadership, and it legitimated doing new things in novel ways, with OD being a primary vehicle. Conventionally, the initial approach was to focus on interaction—to build at start-up in MARTA a regenerative cultural context for individual and group relationships, plus associated skills. As the notion is developed at several points above, “regenerative” here refers to a combination of four variables: high levels of openness, owning, and trust, along with low levels of risk. “Degenerative” relationships often exist in organizations and, while they often merely bedevil everyday management, they would herniate a project like MARTA. In OD parlance, regenerative interaction “empowers” by directing timely and powerful attention to real problems that can be dealt with in terms of valid data, while not only building agreement about solutions but also contributing to psychological success (Golembiewski, 1979a, vol. 1, esp. pp. 1–103). Among other features, strict dollar and time targets in MARTA required spritely surefootedness to avoid the unforgiving dynamics of degenerative interaction.

Conventionally, over time, increasing but incomplete attention also was devoted in MARTA to developing structures as well as to policies and procedures, with both being compatible with regenerative systems (Golembiewski and Kiepper, 1988, pp. 95–123). These would at once reinforce regenerative interaction while being supported by it. The total sense of it is a self-heightening and self-reinforcing spiral of individual and collective competence (Golembiewski, 1979a, vol. 1). This spiral gets especial attention in Chapters 12 through 14.

Some make strong normative arguments for the values underlying OD approaches and techniques (e.g., Sashkin, 1984), and no real alternative existed for MARTA. Certainly, bureaucratic arrangements did not fit well with MARTA’s environment. Additional details appear elsewhere (Golembiewski and Kiepper, 1988, pp. 203–205), but a limited list of broadly environmental constraints suffices to make the point. In outline, among other requirements and in rapid succession, MARTA had to
Assemble a top-level team and several levels of management—from scarce labor pools, from diverse sources, often via national searches, while reflecting an unprecedented commitment to affirmative action;

Do much on-the-job development and training, since few recruits would have experience in either urban transit or in big-project management, or both;

Assimilate into MARTA an existing bus property with its often-recalcitrant culture, a segregationist view that was hugely out-of-sorts with the future MARTA, and with a carry-over of top-level personnel who had just acquired major capital gains in the sale of the bus property to MARTA;

Transition between several overlapping modes, with little waste motion—e.g., from design of a bus/rail system, through construction, and then to operations;

Respect very tight time-frames: MARTA would complete in a decade a project estimated as requiring other local public agencies 3–5 times as long to complete;

Operate under the pressure that the project likely never would be completed unless both time and dollar targets were met;

Discipline itself to uncertain as well as limited local and federal funding;

Operate without the powers to tax or of eminent domain, which required building excellent working relationships with numerous state and local governments and agencies; and

Adjust to daunting local cross-currents involving race and class, which had complex social, political, and economic ramifications.

These determinants suggest that MARTA had much on its plate. Technically, the challenge inhered in MARTA being the largest regional public works project since TVA days, which had to be designed and brought on-stream by an ad hoc organization built largely from scratch by cadres of managers and professionals often lacking urban transit experience. MARTA largely had to “grow its own,” as it were, because urban transit had been in grievous national decline at least since World War II, and its managerial cadres had been aging and shrinking.

At the same time, MARTA’s political context was turbulent. For example, a new black political leadership was coming into power in Atlanta, while the burgeoning suburbs included in the Authority’s coverage remained white and upper income. The opportunities and dangers escaped no one’s attention. Both proponents and opponents alike often saw MARTA as the vehicle increasing metropolitan mobility between local governments that differed widely in racial composition. MARTA had to balance diverse interests, then, while developing an organization that soon approximated and then exceeded a 50/50 distribution of white/black employees, including substantial representations of minorities and
females at all levels—on the board, among executives and managers, and throughout MARTA.

**MARTA AS EXEMPLAR OF PUBLIC MANAGEMENT’S FUTURE**

MARTA’s successful start-up—especially in the context of its multiple challenges, and given the common taunts that predicted an organizational pratfall—drew considerable attention. Two basic reasons undergird this interest:

- MARTA deliberately set out to build its system in nontraditional ways—indeed, it could succeed only in such ways due to strict time and budget constraints.
- MARTA stands unique among large transit projects, worldwide, in meeting both time and dollar targets on a major design and building program.

Much of the attention to MARTA can be labeled “popular” or “political,” and sees the Authority as an exemplar of a management approach that deserves to be influential in the public sector’s future. For example, outlets with all kinds of readership expressed this theme—general readers, public-interest target groups, and professional readers (e.g., Sweet, 1983, pp. 937–949). And representatives of several public agencies—and especially UMTA, or the Urban Mass Transit Administration, which provided vital funding for MARTA—all but universally portrayed the new rail-bus system as “class transportation, not mass transportation.”

In academic circles, much of the interest in MARTA was generated by “insiders”—both individuals and teams, eager and perhaps overanxious to verbalize the good news and deeds. Several milestones distinguish this insider interest. The mid-1970s’ efforts at system-building were summarized for a specialized audience (Golembiewski and Kiepper, 1976); a set of 17 case studies appeared in the early 1980s; two major retrospective articles appeared following the early openings of parts of the rail system (Golembiewski, 1983a, 1983b); a synthesis reflecting on programs like MARTA appeared in the mid-1980s (Golembiewski, 1985); and a summary book on MARTA was written by the Authority’s first general manager and his prime external consultant (Golembiewski and Kiepper, 1988).

Various “externals” have monitored this attention by those closely associated with MARTA. At times, this critical focus has been on the OD approaches relied on in MARTA (e.g., Denhardt, 1984, esp. pp. 101–107). Among those scholars focusing on the MARTA system, most consistent and constructive has been Gabris, who both early (1983) and more recently (1991) has focused on the MARTA literature.
MARTA CIRCA MID-1980s AND EARLY 1990s

In the aftermath of a decade-long effort that met both time and dollar targets in bringing on-line the core of the bus/rail system, two close observers surveyed the effects. One (the author) had resigned from MARTA over a basic difference with the other (General Manager Kiepper), who himself would soon leave to direct a much larger urban transit system, which the author also would serve in a consulting role. For compelling reasons in addition to scientific curiosity, then, both men sought to gain perspective from MARTA on how to do it better the next time (Golembiewski and Kiepper, 1983b), based on the experiences and reactions of MARTA executives (Golembiewski and Kiepper, 1983a). The longer-run product was a coauthored book (Golembiewski and Kiepper, 1988).

Their shared concerns centered around whether the OD infrastructure built over the preceding decade would last. Both men were conscious that their growing concerns might be influenced by their respective detachments from MARTA—already definite in one case, and growing into a relocation in the other. The GM and consultant sought to avoid a sour-grapes orientation, but nonetheless worried that a different MARTA might emerge after only a small (if significant) part of the regional transportation problem had been addressed.

In 1982–83, the prospects seemed daunting that MARTA would survive much in the original mold, as eight summary points suggest. In no particular order of relevance, but with cumulative impact, MARTA faced

A sharp reduction in federal funding, which would inhibit if not preclude planned extensions beyond the system’s core facilities;

Limits on local funding from a sales tax and fare-box receipts;

A possible flattening-out of effort by employees after a very energetic great-leap-forward;

Major personnel losses at the top two or three levels of management—people who had gained very high visibility in urban transit as a result of MARTA’s successes and who were assiduously (and successfully, as it often turned out) courted by other properties (Miami, Houston, Los Angeles, Denver, and so on);

Replacements of any managerial losses would probably be lower-profile appointees with system-maintaining tendencies, if anything, as contrasted with system-building proclivities;

A tendency toward a kind of creeping bureaucratization, as a result of the growing relevance of bus/rail operations versus development along with a corresponding shift from an entrepreneurial mode to more micromanagement;

The growing prominence of internal staff units—e.g., audit and personnel—whose roles had been variously constrained by the several pushes to “build the railroad”; the new balance was seen as shifting toward...
greater procedural regularity and “going by the book,” in part due to external pressures from professional standards or federal authorities, and in part also due to internal pressures to fine-tune the distribution of power within MARTA; and

For various reasons, the MARTA Board was seen as interested in “pulling-in the reins,” especially in connection with the general manager, who had gained considerable leeway in the rush toward milestones involved in building the core of the system.

All in all, then, the 1982–83 probabilities implied a greater precision in getting things done and a refocusing of energies toward micro-management, as well as a likely lowering of intensity. The new MARTA might have more regularity and predictability in a steady-state mode, in sum, and less of the ebullience if not feistiness and rambunctiousness of start-up.

Depending upon where one sat, these tendencies were differently viewed. To some, they augured the threat of the very bureaucratization that the original MARTA had sought to avoid. To others, the expectations suggested a welcome stage in MARTA’s maturation.

REALITIES IN THE 1990s, TENTATIVELY PROPOSED

These tendencies seem variously overstated and underspecified, given a set of interviews held in May 1991. Overall, the subtitle of this chapter catches the sense of that update: The Challenge Remains. So read a home-made hanging on the wall of one of MARTA’s back-hall offices; and interviews suggest that mission survived in the 1990s.

Three emphases detail this conclusion and the search underlying it. In turn, attention is directed at the motivation for a 1990–91 check on the earlier predic-tions; a sketch of the diagnostic approach; and a summary of the dominant realities of MARTA circa 1991.

A Dual Serendipity

Two factors combined to motivate an updated description of MARTA. One was that demand grew. Thus, the ex-consultant completed several major assignments elsewhere, and his personal need surfaced to have another close look at the MARTA for which he had worked over most of its founding decade. Moreover, an invitation to present a convention paper came at about the same time that the Organization Development Journal proposed reprinting the 1976 summary of the MARTA start-up, along with a critical reaction to it. In sum, the scheduling gods aided and abetted my growing felt-need.

On the supply side, MARTA officials proved responsive to a suggested update. In part, this recognized days past, and current officials not only warmly
welcomed the return of a prodigal son but also extended numerous courtesies to him. In addition, the wheels at MARTA also were turning: an internal strategic planning exercise had just begun and, as fate would have it, officials saw themselves approaching an important—even critical—choice-point about MARTA’s future. An external view might contribute to MARTA’s just-beginning evaluation of how the Authority would position itself to face its fast-approaching future.

Diagnostic Approaches

The basic vehicle for diagnosis involves structured interviews with the MARTA Senior Staff—encompassing the new general manager and the assistant GMs reporting to him. Twelve interviews were held—eleven with members of the senior staff, and the other with a just-resigned official having close ties to MARTA over nearly 20 years. The interviews were scheduled for 1.5 or 2 hours, but they often extended far beyond that. Indeed, one lasted for 5 hours, including dinner.

Rapport was often prebuilt, as it were. Three of the interviewed officials had served as executives during the ex-consultant’s MARTA tenure, with two having had extensive earlier involvement with him; and a fourth 1991 executive was a former student. Three other 1991 executives had various vantage points on the interviewer as MARTA consultant. Beyond this, it appears that the consultant’s credibility and motives had been recent topics of several internal conversations between new executives and those with substantial tenure.

The interviews were semi-structured, and recorded for later transcription. The flow and format emphasized:

The ex-consultant’s role as a “visiting fireman” working on a convention paper rather than as a representative of MARTA or anyone else;
A gentle opener: a request for the several “best things” about today’s MARTA management, as well as the “things most needing improvement”;
What defined “success” or “failure” in today’s MARTA, and especially as reflected in the organization’s epic myths and in their underlying ideals or values;
Assessments of the degrees to which these ideals or values got approached in practice;
MARTA myths or legends about the “early days,” and especially about the first general manager;
Comparisons of today’s MARTA with an article (Golembiewski and Kiepper, 1976) describing the efforts to induce an intended normative infrastructure via OD: the article had been supplied several weeks earlier to all those interviewed, and had been read or reread by most of them in preparation for their interviews; and
Reflections on the 1991 MARTA as compared with the interviewee’s earlier experiences in the Authority, which were extensive except in the case of one recent hire.

In addition, MARTA officials responded to some standard questionnaire items. These included:

- Items from the Likert Profile of Organizational Characteristics, which had been applied in MARTA several times over the interval 1975–82;
- Items estimating the degree to which the present MARTA system was regenerative or degenerative—high vs. low on owning, openness, risk, and trust

Various other inputs added to this database. Some officials volunteered written materials, for example; and the ex-consultant’s meanderings through the corridors of MARTA’s headquarters occasioned unexpected opportunities to cross-check reactions with old acquaintances, or with those whose only contact with the days of the “long march” were through reconstructions of that earlier time. Eight such meetings occurred—most accidental, but a few fortuitous-on-purpose by those eager that a rounded view emerge.

The expectations were not that these several sources would provide rigorous comparisons. Rather, the various cross-checkings would permit a relatively-informed, qualitative view of MARTA in the early 1990s—a kind of rough map of a complex territory. If tentative, the search might provide perspective on MARTA’s trajectory between then and the future.

Some Dominant Realities

Nine perspectives summarize the major features of the data-gathering. In general, they indicate that the 1982–83 expectations/probabilities proved inaccurate or always lacked specificity, on balance.

Funds Became Available

Perhaps paramountly, the expected closing of the financial spigot did not occur. That probability had been a major factor in the 1983 fear that MARTA would become a low-energy system suffering regret about the glory days. After an extended period of indeterminacy, substantial new funding made it possible for MARTA to undertake planned extensions of the core system, and that work was just coming to an end around the time of the 1991 interviews. In fact, over the eight prior years, MARTA received 21 percent of all federal funds available nationally for rail start-up.

One cannot say unequivocally that the original MARTA infrastructure both facilitated competing for, and also permitted taking optimal advantage of, these new funding opportunities. But that normative infrastructure clearly did not hurt.
At the very least, MARTA’s reputation for success was a large factor. The record clearly reflects the acceptance where it counts of MARTA as “not just another public agency.”

A New GM Takes Hold

Ken Gregor, who replaced MARTA’s original GM Alan Kiepper, played a major role in the inapplicability of the 1982–83 predictions. This conclusion rests on two sets of evidence.

At the level of expectations, the early 1980s betting was that few influentials wanted a heroic, high-profile replacement for Kiepper. He was widely seen as holding MARTA together in the early days, often by the force of his will, but wide agreement also existed that the continuing costs of such a GM were too high. Hence, the common views that MARTA leadership and operations would take on more conventional features. Consistently, Gregor—one of Kiepper’s long-time assistants—announced his unwillingness to take on the GM role in his predecessor’s hard-driving mode.

Once appointed, Gregor quickly grew into the GM’s job in ways that reinforce MARTA’s singularities and its OD infrastructure. Here, it is possible to illustrate only some senses of the present point. Illustratively, Gregor had

A very strong “sense of team,” which continues the early MARTA emphasis but without the cross-currents characteristic of earlier experience (Golembiewski and Kiepper, 1988, pp. 201–215);
A personal style congenial to OD values: confrontive, and less constrained by fear of failure than attracted to the pay-offs of reasonable risks prudently taken (Golembiewski and Kiepper, 1988, pp. 204–205, 211–212);
Preserved the sense of a diverse but unified workforce: at last count, MARTA employees were 60 percent black and minorities;
A strong credit-sharing orientation, different from that of predecessor Kiepper, with a substantial capacity to “take the heat” generated by MARTA decisions, whether his own or those of subordinates operating under the broad delegations Gregor favored;
Blended strong emphases on “efficiency with economy,” while earlier criticism often emphasized MARTA’s propensity toward expensive analyses and studies as well as high-salaried executives;
Reflected a strong sense of the need to keep employees at all levels “alive and committed,” as by encouraging delegation, discretion, and cross-training or job-switches to keep employees from becoming “zombies” who have “died” at their work, and of their work;
Kept alive the sense and reality of MARTA as a special kind of public agency with active, involved, and committed employees;
Maintained motivation while also holding rein on executive salaries, as by
internal promotions versus national searches, calculated risks in promo-
tions, and successful reliance on developmental or cross-training oppor-
tunities, but perhaps especially by giving his AGMs substantial room for
initiative and control; and
Dealt constructively with serious financial problems facing MARTA in the
mid-1980s

Consider some details related especially to the last emphasis, but relevant
to all of the others. Gregor was a “financial man,” and the mid-1980s MARTA
required precisely this focus. In contrast, Kiepper focused on “building the rail-
road,” with flush budgets and with a primary view to system effectiveness. Illus-
tratively, given a different day-and-time, MARTA wage increases averaged 11.5
percent per year during 1972–81 but only 2.9 percent per year since 1982. Gre-
gor’s approach is revealing of the man’s style. “The cupboard is empty,” he later
paraphrased his approach to unions and the general public. “Folks, you’ve got
to lower your expectations.” See also the later analysis of Likert Profile data.

As the Twig was Bent?

Several developments in the 1991 MARTA are consistent with the original OD
initiatives, and these developments deserve illustration even though their roots
must remain speculative. Hence the interrogative heading, despite the fact that
interviews typically reflected definite trails of associations between then and now.

Broadly, the pulse of life at MARTA seemed to have moderated, but the
essential high-energy character seemed to remain. As one Assistant General Man-
ger put it:

AGM: Well, one of the characteristics that I describe to people
when I interview them for management jobs here is the
dynamic pace of the organization. That was characteris-
tic of it in the beginning years, and that is very much
the style today. For a government bureaucracy, in many
ways it operates on a day-to-day basis like a private or-
ganization.

Interviewer: The early motto, you may remember, was “Not Just An-
other Public Agency.”

AGM: Many of us remember; and MARTA certainly isn’t, in
my opinion.

In large part, this pace reflects that MARTA is seen as a “mixed” organiza-
tion: it is clearly public, but it also has flexibility as an Authority, faces substantial
time-pressures, and sees a need to be client-centered in ways that many organiza-
tions in the public sector are not, or do not care to be. “If we have people wait-
ing in line to buy [monthly transit cards] at 7:45 AM when that office opens at
8:00,” says one AGM, “then we’ll open at 7:30. If the queue begins at 7:20, we’ll try hard to accommodate that as well.”

Relatedly, the MARTA system circa 1991 shows major signs of the organizational flexibility that was the prime goal of the original MARTA. Thus, the internal political force field is shifting in significant ways—roughly, development has a lessening role as major construction is moving toward completion, while operations and audit assume growing significance. These changes were expected in the mid-1980s (Golembiewski and Kiepper, 1983a 1983b), but their full impact was delayed by new funding.

Significantly, these critical adjustments seem to be accepted as new environmental determinants to be dealt with cooperatively, on clear balance, rather than as motivators of conflict, frustration, or withdrawal. Relatedly, various bus/rail activities have been shifted in ways that could have caused jurisdictional issues. Today’s executives speak of the “spirit of a matrix without its form,” in fact. This suggests a structural change-in-fact to complement regenerative interaction, although the change-in-form has not been established.

This flexibility may be severely challenged if MARTA seeks to expand its mandate. For example, one executive speaks of the “need for salespersons in MARTA over the next few years, just like the old days.” The reference is to the entrepreneurial spirit characteristic of the original MARTA officialdom, both before the passage of the referendum authorizing the MARTA system, as well as in the several years before the bus/rail system went operational. Unless MARTA expands its jurisdiction or product line, the system will become basically an operating unit, losing thereby much of the innovative character associated with design, planning, and early development.

Perhaps the clearest suggestion of past → present associations relates to the emphasis on the “team.” This reflects the style of GM Gregor, but also has antecedents going back to the original organization-building exercises. As one AGM notes: “Ken’s style is one that fosters teamwork. He insists on people being collaborative and understanding each other’s point of view. His ego isn’t large, and he doesn’t like people to be mad at each other and to be destructive in the organization.” Moreover, due to both socialization and perhaps especially selection, the 1991 Senior Staffers have a “close fit” that reflects and reinforces Gregor’s style. One AGM echoes the sentiments of virtually all colleagues in this exchange:

AGM: We don’t have the personalities that we once did. One by one, as you meet these folks that you haven’t met, you will find them trustworthy. They do have the best interests of the organization at heart. They in the long run want to work collaboratively for decisions that are good for MARTA. I don’t know of any one of us that I could say doesn’t fit that pattern.
Interviewer: Of course, the critic would say, “Well, that’s often true of organizations after the great leap forward.” In the next stage, people emphasize managing the system, lubricating the system a bit. Building a system requires different talents, more initiative, and risk-taking.

AGM: Yeah, we have a lot of people like that—practically the whole Senior Staff. The personality of the management group now is more in tune with, and more appropriate for, the maintenance of the organization. You know, rather than the first stages, which featured more in-fighting and stumbling toward the MARTA model that we now have pretty well-in-hand.

Also noteworthy is an ongoing in-house strategic planning initiative, which implies an effort to tighten-up procedures, policies, and structures as a prelude to a possible later shift in missions-and-roles. Consistent with the OD efforts in the early MARTA to include all levels, including the board (Golembiewski and Kiepper, 1976), this effort emphasizes broad involvement, including that of union members at operational levels. It is too early [at the time of observation] to estimate the effect of this exercise, which may well be a precursor for a broader strategic planning initiative involving powerful external stakeholders.

Several other MARTA activities also reflect the ongoing concern to avoid becoming “just another public agency.” They can be noted briefly, with sparse explanatory comments:

- The Board of Ethics continues, and robustly from all indications, reflecting MARTA’s thrust toward an open system—in this case, by aggressively evaluating the performance of its employees, exploring conflicts of interest, and so on
- Various efforts to continue to play a leading-edge role among regional agencies, which include initiatives related to customer-oriented behaviors and attitudes at all MARTA levels as well as to quality assurance, building on earlier efforts (e.g., Bates, 1981)
- A performance appraisal system linked to rewards for exceptional contributions of as much as 7 percent per year

The New Cadre

As expected, major personnel changes occurred at executive and managerial levels, as other organizations oriented their recruiting to MARTA’s success. Compared to their predecessors, the new executives are more likely to be younger, female, and the results of internal promotions. Specifically, only four executive holdovers from the mid-1980s were interviewed, and one of those had already shifted to a major role in Atlanta’s 1996 Olympic Games.
Checking Progress with Replication

However, the dire consequences considered probable have not occurred. These include concerns that efficiency and effectiveness would suffer, any hard-won chemistry would be diluted, tendencies toward bureaucratization would be heightened, and energy levels would plummet.

In fact, on clear balance, the dominant characterization of the 1990 MARTA executives seems a prevailing commitment to the "team" mode, which contrasts with the rambunctiousness of earlier executives, which so complicated early efforts toward a cohesive senior staff. Potentials for divisiveness in 1990s were revealed in several interviews, but senior staff still seemed willing to work with personal and organizational differences because the present balance of career costs/benefits in MARTA remains attractive, even desirable, under Gregor's evenhandedness. In some settings, the verbiage about "team" is window-dressing that merely submerges the expression of real differences or that puts a collaborative face on a Charge-of-the-Light-Brigade mentality. Neither case seemed to apply at MARTA, and certainly not in undiluted forms. As one executive noted of the 1991 senior staff, in special connection with the emphasis on internal promotion: "I don't think that it's incestuous, but I do think there is a lot of compatibility between some of the older heads and some of the newer heads."

This basic orientation no doubt has numerous explanations, of which two get emphasis here. Most obviously, the system has taken essential shape, which reduces the opportunities (or need) for sharp differences; the early MARTA executives always included several overt contenders for the general managership—which regenerative interaction encouraged be made overt—while neither the interviewees nor the author see an opponent of the GM in the present cadre.

More subtle issues probably also are at work. The national searches common in the earlier MARTA generated candidates who were more likely to be promineats with strong personal agendas who—due to scarcity—MARTA was glad to have even if fit was incomplete or in question. Indeed, a major motivation of the MARTA OD efforts was to manage, or even transform, precisely such lack of personal congruence, and to do so in tight time-frames. Willingness to be a team player gets more attention nowadays as the price of admission. Thus, one executive notes of himself that he finds MARTA congenial, but due to a prior fit rather than to socialization by the MARTA approach. His approach was "something I brought with me, always my style." The emphasis on internal promotions clearly also would inform judgments about prior fit.

Although there seem to be no major untoward consequences of this staffing pattern, a major test seems on the horizon. The ongoing strategic planning effort may constitute a major challenge for the present senior staff, and a possible cut-back due to a recent drop-off in sales tax revenues also may have divisive effects. Union contracts also were to be renegotiated at 1991's end.

One significant commonality of the old versus new MARTA seems worthy
of emphasis. In the old MARTA, regenerative interaction was necessary—to get people acquainted quickly, to encourage quick compatibility, to deal with some of the “rough edges” of executives and middle managers, especially those who remained after selling their bus property to MARTA. In the new MARTA, regenerative interaction was perhaps even more necessary—as in surfacing issues that developed despite the building effects of teamwork and continued employ by a high-profile Authority, and so on.

The “Older Heads”

A substantial stability has characterized MARTA at many levels, despite the expected flurry of losses around 1980–83. Gregor as GM in the 1990s was the national dean in point of continuous service with a single property, and he has been at MARTA for 19 years as a Senior Staffer. Industrywide, in contrast, GM tenure is 3–5 years. Until quite recently, two others of similar tenure were Gregor’s colleagues.

Many others at lower levels also “have been around,” but without “going stale.” In part, this reflects the rapid transitioning in major tasks that has characterized much of the MARTA history. As one executive notes: “Probably for 10 to 15 years I felt that we never did anything twice.” In addition, departures from the original MARTA senior cadre opened up developmental positions for those who were promoted early and stayed on. The result? One executive observed about the professional staff:

You know, when we first started building the system, we had a lot of new people, a lot of new learning processes, a lot of mistakes that were made that had to be corrected, and so on. And we now have a seasoned group of engineers that are doing a crackerjack job.

Note that MARTA turnover falls in the 6–7 percent range, which suggests the major attractions of employment in the Authority.

This particular coin has another side. Consider this exchange concerning an AGM’s comments about “things slowing down a bit, and bureaucracy . . . getting worse”:

**Interviewer:** I have no way to make a firm judgment, but my sense is that most of your colleagues still think that MARTA’s a pretty good public agency. A cut above—

**AGM:** Yeah, definitely, definitely. It has retained its reputation. Very professional and on the cutting edge in transit, but—

**Interviewer:** Reputation or reality? I am trying to smoke you out here. The “but” suggests you see some slippage from higher ground.
AGM: Oh, I do, I definitely do. I think we’ve been resting a little bit too much on our laurels.

Another senior staffer sees it similarly, but with a bit of a twist appropriate to that AGM’s turf. “Our ship is just a little more laid-back than when I first came on-board.” The AGM adds: “We are trying to energize it in the last year or so.” It may be a matter of where one sits, in fact. As another AGM notes:

Even though there are bureaucratic aspects of MARTA that have become necessary because we need good internal controls and to protect ourselves against audits in buying and contracting, I don’t believe that we have become a large murky bureaucracy. You can still get things done. . . . There is a lot of delegation downward. I believe that there is a good level of authority at the manager level [below the AGMs]. And you know, MARTA’s a very dynamic, fast-paced organization. That has never changed. . . .

Shared Aversion to “Bureaucracy”

A dominant agreement among senior staffers concerns a desire—in several cases, a passion—to preserve the flexibility and managerial lightfootedness they associate with the early MARTA. GM Gregor probably led the charge in this matter, arguing with passion against “no-brainers” and “organizational zombies” who lack discretion and initiative. Opinions differ about what constitutes appropriate discretion, but all place MARTA’s practices as well above common organizational practice.

The anti-bureaucracy theme was dominant in the early OD efforts and, with some qualifications, that survives. Thus, some MARTA executives do see a drift toward bureaucracy as inevitably growing, but as mainly a force to be resisted so as to delay and dilute its impact. Others take a variety of positions, as exemplified by this differentiated view of one AGM:

AGM: Yeah, it makes it more difficult to get things done. We are entrenched in reams and reams of administrative procedures. So we’ve become insecure—probably defensive to some extent—in what we do and how we go about doing things. Things are more regimented now.

Interviewer: There’s some good news and some bad news in that—right? So things have maybe slowed down a little bit?—

AGM: Exactly, exactly—

Interviewer: —so they’re easier to keep on top of now. On the other hand, they’re more of a pain in the ass.

AGM: Yeah.

Interviewer: Which would you rather have?
AGM: Oh, I guess the way it’s going now is probably the best way. Yeah. From my personal perspective, I guess that it is. But for a young person coming into the organization—a new employee who wants to achieve, a high achiever—it must be terribly frustrating for them.

A Helpful Board

All executives emphasize good relationships with the MARTA Board of Directors, who are appointed by the City of Atlanta and several counties. Indeed, many executives note that board relationships are the most attractive system feature, despite occasional public flare-ups.

It was not always thus, and early major OD efforts sought a more satisfying balance. The 1970–1980 tensions of moving toward an operating bus/rail system often were intense, and board vigilance—and on occasion, vengeance—had many early motivators. This understandable situation was exacerbated by the fact that some board members came from counties that did not earmark sale tax receipts for MARTA, but whose residents could free-ride on the system—as by driving their personal cars to MARTA park-and-ride facilities and then use the system. In another example, consultant Golembiewski worked directly with the board and especially its chairman for several years, on projects including reorganization and performance appraisal of the GM.

Explanations for the present situation include success in changing the early unfavorable board/staff relationships as well as MARTA’s reputation for managerial competence, but also extend beyond such factors. Thus MARTA has been fortunate in Board appointments, in general, who often have a policy versus operational orientation. Others emphasize that no one wants to “muss the feathers of a truly-golden goose, let alone harm that goose.” And most of the construction contracts, as well as the local enhancements associated with them, are past history.

A major test of the early 1990s board/staff relationships was seen as possible. MARTA may have to undergo a cutback due to falling sales tax revenues, and this might encourage board interest in many operating details. Generally, GM Gregor has been successful in maintaining the view that staff responsibility implies substantial operational discretion within policy guidelines. Consultant Golembiewski has long argued that expanding MARTA’s jurisdiction is the only long-run approach to saving the virtue of the innovative MARTA. Approaches might include becoming an interstate compact, on the general model of the Port of New York Authority. However, major stakeholders would resist such a development, at least short of a major consciousness-raising. That story is history in 2002, but will not be recounted here.

Gentle Perspective by the Numbers

The interviews also generated two quantitative estimates of the state of relationships in MARTA of 1991. These estimates help locate where MARTA is and,
Checking Progress with Replication

if with a real tentativeness, permit some gentle comparisons with where MARTA was.

**Likert’s Profile of Organizational Characteristics.** The earlier MARTA had made now-and-again use of Likert’s (1967) instrument, which, in its several forms, has several useful features. First, its several items are scored along a continuum of 20 equal-appearing intervals, which are differentiated into four major systems of management:

<table>
<thead>
<tr>
<th>Scores</th>
<th>System Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1–5)</td>
<td>System I Exploitative-Authoritative</td>
</tr>
<tr>
<td>(6–10)</td>
<td>System II Benevolent-Authoritative</td>
</tr>
<tr>
<td>(11–15)</td>
<td>System III Consultative</td>
</tr>
<tr>
<td>(16–20)</td>
<td>System IV Participative Group</td>
</tr>
</tbody>
</table>

Second, each item is anchored by four brief descriptive statements, one statement for each system. Consider the item: At what level are decisions made? The four Likert systems are anchored by these descriptions, respectively: I, mostly at top; II, policy at top, some delegation; III, broad policy at top, more delegation; IV, throughout [the organization] but well integrated.

Third, each item is scored twice. A Now score reflects an estimate of the existing condition, and an Ideal score reflects a respondent’s preference for what should exist.

Efforts were made to avoid contamination of Ideal and Now in the 1991 responses. Thus respondents provided Ideal scores early in their interviews, after the Profile approach was explained. Now scores were solicited much later in the interview, typically after a minimum of 30–45 minutes of other discussions. If anything, then, the intervening questioning may have deflated Now scores.

The twelve interviewees responded to five Likert items tapping four domains: Decisions, Goals, Control, and Motivation. Their Ideal scores average in the 13–15 range, which is to say mid- to upper-levels on System III. The Now scores have means of 9, plus or minus a tad, which is just within the upper boundary of System II. Although the point could be extended too far, some comparisons with earlier Likert scores can be suggestive. Briefly, distributions of Likert scores during the early development of MARTA can be characterized in these terms:

- Ideal scores averaged at mid-levels of System IV and above
- Now scores, before OD interventions, averaged in the 8–10 range
- Now scores, after OD interventions, averaged well into System III or low System IV

Two tendencies seem apparent. Aspirations of today’s MARTA executives reflected in Ideal scores are substantially more constrained than those of their
predecessors of 10–15 years earlier. Moreover, Now scores seem clearly lower. No statistical tests were performed.

Now, let us fast-forward to the early 1990s, although comparisons with the early MARTA self-reports must be hedged. The profile scores in MARTA circa 1991 also reflect a very wide dispersion of scores which, in general, coincide with the tenures of senior staff. The longer the tenure, the higher are both Ideal and Now scores. It is worthy of note that seven of the twelve interviewees in various ways drew attention to this difference between the “old guys” and the newer additions to the senior staff. About half of the executives are newer appointees.

Since the theory underlying both the Likert Profile and OD places special emphasis on System IV as a goal, broadly, the numbers suggest a definite drift in aspirations as well as in the perceptions of executives—a drift “downward,” as it were, toward the bureaucratic features that MARTA sought to avoid. The scores do encourage further probes over time to provide context for the numbers, as in detailed investigations of whether the Ideal/Now profiles will provide adequate support for MARTA at its present choice-point.

Regenerative Interaction. Interviewees also were asked to rate MARTA along the four dimensions thought to characterize individual and group relationships—openness, owning, risk, and trust, as estimated along a 5-point scale anchored by High, Medium, and Low. Mean scores cluster around 3.2–3.5, which are substantially beyond the mid-point. However, the scores certainly do not imply the “high” scores consistent with OD values and approaches. No direct comparisons with MARTA executives of 1975–85 can be made, unfortunately.

The individual scores again reflect a dispersion associated with differences in tenure. MARTA is more regenerative for the longer-tenured senior staff.

An Apparently Emerging Imperative

If with the qualifications appropriate to all dynamic situations, MARTA in 1991 was seen as standing in deep need of a booster shot reinforcing the kind of culture and relationships that were built a decade earlier. Directly, the Authority faces a choice-point that will shape its future. In this author’s terms, Choice I sees MARTA as struggling to stay even—as a metropolitan transit authority, serving (as it now does) Atlanta plus several of the metro counties. Choice II sees MARTA as enlarging its service area into the eight-county metropolitan area, or even beyond.

Directly, whatever the key choice, MARTA requires base-expansion as a transit authority, which can occur in one of two ways. Modestly, MARTA can strive to increase ridership through such devices as encouraging businesses and governments within its present service boundaries to subsidize employee travel much as they now subsidize parking for personal cars. This strategy is inherently
Checking Progress with Replication

limited, although it could help. MARTA’s downtown Atlanta base is already being affected as businesses (and hence potential passengers) move outside the fixed reach of the core rail system.

More aggressively, MARTA can seek extensions of its present system into the fast-growing suburbs. The market for mass transit burgeons, with the metro area having grown by about 30 percent during 1980–89. Indeed, the fastest-growing U.S. county, Gwinnett, held a referendum in 1990–1991 about entering into a contract with MARTA. The referendum did not carry, but it may be a harbinger of future extensions. MARTA officials, at Gwinnett’s request, did not play an active role in the referendum. More expansively still, MARTA could seek authority to build new rail linkages into a regional network.

These variants on base-expansion all imply political issues—in the less-expansive variants, local governments might resist; and in the most expansive variant, opposition from Georgia DOT seems probable, if not inevitable, and the support of the Authority’s local stakeholders is now unlikely.

Not being “just another public agency” can be a very useful reputation to maintain, whatever the base-expansion strategy. In an Armageddon view to its supporters, for example, MARTA would be folded into a state agency—most likely, Georgia DOT. The state in 1991 had no coherent transportation policy, with DOT remaining largely a department of highways. However, Georgia’s new governor may be seeking to redirect the DOT mission into transportation, as broadly conceived. A successful June 1991 initiative by several major Georgia executive and legislative figures induced the retirement of the DOT head, and arranged for his replacement by a more politically responsive appointee without department roots. Atlanta officials would resist even well-connected covetousness directed at MARTA. In all possibilities, MARTA’s reputation for efficiency as well as effectiveness remains a valuable adjunct to this built-in political support.

Not coincidentally at all, funding for mass urban transit seems to have had a suddenly high priority in congressional circles around 1990. Funds to fuel an expansion of base apparently will be available in recently unprecedented volume. If this occurs, it will add zestful urgency to how and when MARTA decides to resolve its present choice-point.

CONCLUSION

Four major points fairly characterize this update, given its limitations. First, most of the policies and practices in the 1991 MARTA seem to this observer to have deep roots, and suggest noteworthy persistence of early efforts in an OD mode. This characterization applies, despite major senior staff turnover and the passage of a decade since the last OD activity in the agency.

Second, various serendipitous features also contribute to the present sense of “MARTA.” The unexpected availability of federal funding clearly was of ma-
jor moment, for example, as was the growing-into-office of General Manager Ken Gregor, although he was not to survive in office very long after the time-frame of this revisit. Along with other factors, such buoyances helped blunt any major tendencies toward the dour 1982–83 probabilities. So one might argue that OD values and approaches were not subjected to a maximum test.

Granted. But the past decade or so has brought a goodly share of difficulties, and MARTA’s systemic development has continued.

Third, the quantitative data—given their clear limits—require major caveats about proposing that OD values have largely persisted in MARTA. To some, indeed, major tensions or even contradictions may seem to exist between the Profile data and much of the qualitative information reviewed above. For example, assume that MARTA’s board turns combative or even hostile in the face of state financial difficulties. The moderate scores on the Likert Profile and on the four components of regenerative interaction might then prove insufficient to inhibit a win-lose mentality from dominating, employee and union discontent might flourish, General Manager Gregor or successor(s) might substantially centralize and shorten the leashes on the other executives, and so on and on.

This author’s view has longer- and short-term components. These estimates attempt to take into account my 1991 personal inclinations or biases. I did not see myself as emotionally tethered to MARTA, having resigned in the mid 1980s. Moreover, along with General Manager Kiepper, I foresaw in 1982–83 great difficulties for MARTA beyond start-up. So history seems a wash to me, in this particular.

As for now, MARTA remains a qualitatively different experience—for most others, as well as for me. So as meaningful as the clear downward drift in “the numbers” is, the more attractive is the comparison between today’s MARTA and other public agencies, especially those in its niche nationwide and in its locality. As I see it, the advantage still remains with MARTA.

As MARTA moves more explicitly into rounds of strategic planning with multiple stakeholders about its future, and especially in any attempt at expansive base-building, appropriate concerns escalate. For example, in 1992 I reported to MARTA offici-aldom that reinforcement experiences for OD values seem strongly indicated, at a minimum. As the pressures mount in connection with the several choice-points now facing the Authority, as they very likely will, the significance of underlying interpersonal and group relationships will be heightened. The numbers here suggest that real deterioration has occurred in both the Now and Ideal aspects of these relationships, and this may negatively impact what gets done as well as how, especially when it comes to aggressive choices about MARTA’s near-future. An organization that “wants more” from its members, reasonably, has to “give more.” In general, such estimates of the health of underlying processes like those here are leading indicators of communication and decision-
making difficulties. However, qualitative data suggest no clear and major eruptions of such difficulties, as of October 1991.

Fourth, MARTA now seems poised for what in OD jargon is referred to as a “new orbit,” a “voyage of discovery” (in Neely Gardner’s words), or another “action research sequence.” The full process will encompass a strategic plan, built around what MARTA will aspire to become as the future unfolds.

Basically, the earlier MARTA was more focused on relatively “internal” development, with the quotation marks referring not only to intra-Authority issues but also to the local policy environment as well as to limited state and federal stakeholders.

From the perspective of this 1991 revisit, MARTA faces a basic choice. MARTA can remain in its historic frame, and face a future of stability, at best, or more probably a continuing decline in its passenger-base, with significant effects on Authority morale and aspirations. Or MARTA can choose various frame-breaking orientations to the future: to add to the passenger-base by expanding beyond its limited historic base; and/or to add to its product-line beyond rail and bus service, as by moving into other transportation modes or by various development activities along its present bus and rail corridors or in the airspace above station sites.

Of course, each strategic initiative has different mixes of political costs/benefits.

The point can be overdone, but the temptation is great to think of the choice-point in anthropomorphic maturational terms. Like people as they strive to “grow up,” MARTA faces the choice between standing pat or further development, of continuing to focus on a sense of the historic self or of refocusing in search of a broader sense of self (e.g., Hampden-Turner, 1966). The latter implies a risk and requires courage, of course, but it also represents an alternative to decline and even eventual senescence. This “looking outward more and inward less,” in Neely Gardner’s words, characterizes a new “voyage of discovery.”

It is premature to judge whether and how frame-breaking will occur, but two points seem clear enough. Thus, OD values and approaches imply an organic view of organizations, and hence are tailor-made for the strategic planning and other developments that MARTA faces. The conceptual rationale has been developed in great detail (e.g., Jellinek and Litterer, 1988); and operating models also exist (e.g., Golembiewski, Nethery, Shephard, and Hilles, 1992). OD’s relevance will certainly be high if MARTA adopts a frame-breaking orientation. If in different senses, the conclusion also will hold if the historic mission is maintained. For this choice would imply to many an acceptance of genteel decline, if not an abject failure to attempt to extend the MARTA model so as to meet major state and regional needs. Such shortfall would require subtle fine-tuning to avoid seri-
ous consequences, such as drops in performance and service as well as the human costs associated with them.

This analysis will resist the temptation to fill in the full history from 1992 until now. The limited goal was to test change effects through 1991 or so. That is ambitious enough a target. This chapter closes with a generalization about MARTA’s post-1992 history.

Clearly, any MARTA frame-breaking will imply the test of the degree of persistence of OD values developed in the earlier MARTA as well as of the degree to which they can be built upon. This highlights the significance of the apparent reduction in quantitative measures reviewed above. The existing levels may well be appropriate for the MARTA of today; these levels also may suffice if MARTA adopts a stand-pat posture for the future; but real concerns are appropriate about their appropriateness for frame-breaking in MARTA. This last condition will stress relationships between individuals and groups, probably severely, and no doubt would heighten the need for regenerative interaction. The apparently contradictory alternatives would be high levels of conflict, or low energy levels associated not only with low conflict but also with low aspirations.

As a postscript, the time-frame beyond this chapter now approaches a decade, and most indicators point to a steady-as-she-goes approach. But this chapter never sought to develop conclusions for all times, and hence this datum is merely accepted. The universals remain: OD values and approaches can contribute to large system development, in noteworthy and perhaps unique ways. And the human systems developed, as in MARTA, often will be worth preserving. Hence, the choice-points raised by the MARTA experience are likely to apply more broadly, and the dilemmas will be similar, whatever happens in Atlanta.

NOTE

1. Six of the cases are conveniently available in Mass Transit Management, U.S. Department of Transportation, Urban Mass Transportation Administration, (1981). Most others were accepted by the Intercollegiate Case Clearing House, Cases in Public Policy and Management.

REFERENCES


This chapter and the two following propose that it makes sense to view OD as “increasing responsible freedom,” and they consequently swim against a substantial tide, as it were, and quite self-consciously. Thus, powerful voices in recent years have called for OD to be less stiff-necked in the service of its traditional values, and thereby to be of greater service to managers and executives (e.g., Beer, 1989). But these three chapters seek to be more explicit, and in a global-tending way, about how and why traditional OD values are better served by thinking in depth about “responsible freedom.”

The present view is that being more normatively casual about OD may broaden its usefulness but, beyond some imprecisely known point, that would also risk a transmutation of OD into something other than OD.

From another perspective, these three chapters seek greater specificity about how and why respecting traditional values permits applications of OD far broader than common understanding has tended to acknowledge. To provide a bit of orienting detail, this first chapter develops some basic conceptual features that illustrate the normative groundedness of OD. A subsequent chapter places OD in the context of two normatively-loaded Western ethics applicable at work; and a concluding chapter begins exploring the fit of a Confucian ethic to OD practice and theory. Throughout, the focus is on workaday organization although, clearly, extensions of this approach to molar levels are not only possible but necessary—with “molar” being understood as relating to nation states, that is, to social, political, and economic aggregates.
Note one additional introductory point. This may seem a curious juncture at which to introduce such a central notion as increased responsible freedom as *the* normative tether for OD. But better later than never. Moreover, the six earlier chapters provide a useful context for this normative tethering of OD: in effect, these success rates imply the attractive proposition that normative sensitivity has benefits that outweigh costs, and substantially so on the face of it. From a variety of perspectives, those chapters relate to a significant technology for change that needs the protection of a normative orientation that serves several purposes. Thus, the tethering values should protect both labor and management as well as other possible stakeholders from misuse, manipulation, and misrepresentation of the OD technology.

Moreover, the normative orientation ought to be informed in two ways: by ethical traditions that have a broad acceptance and also by evidence of where and why specific OD designs seem to “work.” Neither the three chapters below, nor the rest of this volume, accomplishes these two types of tethering with completeness. But they certainly begin crucial demonstrations, and in ways that build upon *Ironies I* and other expressions of OD as a normatively-constrained technology for planned change.

**A CONCEPTUAL CONTEXT FOR IRONY III**

Chapter 7 begins a formidable task in relation to major thoughtways: in effect, to show how OD can be thought of as guides consistent with major approaches to life and work, rather than as some fad barely surviving in narrow circles of contemporary thought and action.

This demonstration comes in three parts. Immediately, discussion shows how OD fits with a major interpretation of Western thought whose goal is to increase responsible freedom in life. No one person has a monopoly on this approach as a useful perspective to meet modern problems, but here emphasis goes to Pope John Paul II, who has enriched us all by an integrative emphasis on responsible freedom as a conceptual key to the problems of contemporary life, and more specifically to the problems of contemporary religious experience.

The point of Chapter 7 is direct: to show how OD can at once derive momentum from Pope John Paul’s interpretation of religious experience and being, and at the same time show how OD values and approaches can extend the Pope’s view of responsible freedom to the exigencies and practicalities of the workplace.

**Organizations as Value-Loaded**

Put another way, OD sees organization as value-loaded—whether that happens in the normal course of events, by deliberation, or by mercurial mixtures of both. OD proposes to emphasize deliberative value-loading, and in that basic orienta-
tion differs only in characteristic values from much management thought convinced of the attractiveness of organizations that can be substantially “loose” in fitting the “tight” template of a normative system. In Saraide, for example, their comprehensive template is “Commitment-to-Excellence”—which encompasses the firm’s Vari Fare Philosophy, Excellence in Thought, and Excellence in Action.

The normative template proposal here—regenerative interaction and so on—is compatible with a range of such broader variants. The focus here is on essentials from one point of view, which can guide many broader templates tailored to specific worksites and conditions.

FREEDOM AS A CENTRAL THEME

It is tempting to preach on such a fundamental topic as the goal of organization change, but your author intends only to reflect on his own ruminations over the years—both published in sources here and there, as well as largely created for present purposes. Of course, the hope is that this review will aid readers in their own grappling with their practices but, whatever the case, your author nonetheless needs to fast-forward the tape of his thoughts over the years related to freedom in organizations.

Why be so responsive, and now, to such personal urges? The immediate stimulus is the clock and calendar—the ones that recently have told me that your observer has reached that stage in life when many others have retired. So there seems no time better than now for the present purpose, and there may be other times that are worse. Briefly, that clock and calendar both urge a greater certainty that one is emphasizing the right things in a correct manner, and also doing so with growing precision and increasing efficiency. This focus on freedom in organizations ranks very high in my priorities of what deserves careful and competent attention, as well as continuous review in both research and consulting practice.

For present purposes, which suggest an orientation rather than a settling of matters, three themes get attention. Thus, this first in a trio of chapters deals with some preliminary conceptual issues framing “freedom” for emphasis in organization change, and especially in OD. Two basic meanings are isolated: one is faux as well as simplified, and a second will challenge the best that is in us. A following chapter will deal more specifically with how a value-loaded “OD Ethic” can help normatively ground “freedom” in organizationally-relevant senses, if perhaps only from a Western perspective. The third chapter involves grounding “freedom” in terms of an Eastern normative framework—the Confucian ethic.

Adding motivation to raise these related topics is the ongoing review of John Paul II’s papacy, as framed in terms of his self-aware stewardship of the initiatives of Vatican II, which the Pope himself labeled the “council of freedom.”
He saw himself as “... witnessing an extraordinary global acceleration of the quest for freedom which is one of the great dynamics of human history” (Pope John Paul II, 1995, p. 1). Some observers may be overly generous (e.g., Weigel, 1999) in assessing papal progress in helping this initiative grow and mature, but we can do far worse than to test the degree to which the Pope’s basic insights can enrich organization change as well as the stewardship of humanity.

The initial prognosis seems very favorable. Narrowly, that is to say, this chapter thanks John Paul II for his thoughts about “freedom” and seeks to build upon them, and perhaps beyond them, in focusing on OD. Ideally, indeed, this trio of chapters might provide some useful detail for the Pope’s broader mission. In any case, the Pope reasonably enough did not give much attention to planned organizational change in his comprehensive pastoral view of “freedom.” This trinity of chapters provides such detail, while thankfully accepting the added motivation and content provided by the Pope’s thought as these chapters strive to demonstrate the usefulness of making “freedom” prominent in the theory of planned change.

FREEDOM FOR WHAT?

There was a time, and not so long ago, that was ebullient: employees in organizations had nothing to lose but their bureaucratic chains, with this fateful process already then being seen as well under way as well as inevitable (e.g., Slater and Bennis, 1964). Not long thereafter, however, once-optimistic prophets acknowledged that some strange and unexpected things had happened on the way to what had seemed a glorious unavoidable future (e.g., Bennis, 1970).

One easy way to characterize this interval is as the learning period for an old truism to surface. That is, freedom is not an end-goal, but a paradox to be carefully dealt with. That is to say, organizational freedom is at once related to fundamental human aspirations, but at the same time “freedom” also raises some of the most profound dilemmas facing all of us. “Freedom” is not just another term for “nothing left to lose,” to put it another way.

But we still need to be proactive about “freedom,” and one direction seems useful enough. There seems no one-best way to resolve the implied paradoxes, for once and all. What seems required is an intelligent and fanatical perseverance in working toward a resolution of a simple question that will permit some useful progress. In sum, freedom for what? It focuses primarily on what can be gained.

TWO CONCEPTUAL MODELS FOR A WORKING ANSWER

In the present view, the apparent and even disarming simplicity of this question should not be allowed to mask its significance. The answer is not “for everything.” So how can we move toward a useful working orientation? We can do much
Responsible Freedom as the Goal of OD

worse than to adopt the dual view of Pope John Paul II, expressed in a speech at the United Nations in 1995. To the Pope, “one of the great dynamics of human history” relates to the deep-seated yearning for freedom. As Pope John Paul II saw it, the primary battlefield encompassed extensions in space-and-time of universal human rights, everywhere. The yearning in one place and time may differ from others in degree or urgency, but is everywhere similar. Quite specifically, Pope John Paul II rejected any limitation of freedom to “any one part of the world [or to] the expression of a single culture.” He added:

Men and women throughout the world, even when threatened by violence, have taken the risk of freedom, asking to be given a place in social, political, and economic life which is commensurate with their dignity as free human beings. This universal longing for freedom is truly one of the distinguishing marks of our time. (Pope John Paul II, 1995, p. 1)

Such powerful yearnings notwithstanding, however, this burgeoning worldview also seems a fundamentally troubled one in the Pope’s view. Directly, this quest was in the gravest jeopardy of collapsing onto itself, of leading to its own negation because of the lack of attention to the primary query: Freedom for what? The Pope saw the modern crisis as another case of not finding the truth because the search was hasty and the prognosis was indiscriminate.

The conditions giving rise to the Pope’s dour conclusion are quite straightforward, and much the same view applies to organizational applications of “freedom.” To summarize by way of introduction, in recent times one concept of freedom has come to dominate; but only a fundamentally different concept offers real hope of dealing with the dilemmas inherent in the fruitful pursuit of freedom. The tragicomic situation brings to mind the myopic person, in the dead of night, searching for a wallet under a street lamp. The loss had occurred some distance away, but the searcher chose the site under the street lamp because the light was so much better there. We all need to go back to where the loss occurred, whether the issue is a lost wallet or freedom in organizations.

“Freedom” As Self-Assertion, Period

Modern times, and especially after the 1960s, featured an uncomplicated but unproductive view of freedom. In the words of Kris Kristofferson, I believe, many were tempted to see “freedom” as “just another word for nothing left to lose.” Put otherwise, freedom in this usage is operationally defined as self-assertion—“doing your own thing, no holds barred.”

Evidences of freedom as self-assertion may be found in much of the early OD literature, as well as in its numerous cultural artifacts. For example, many proposed that greater openness and owning in communication are sufficient to
bring us to an age of more flexible and productive organizations, with little or no concern about qualifications on such a simple and sovereign principle. In historic time, Rousseau comes to mind. People were born free, yet are everywhere in chains. Simply, to some, mankind had only to remove those chains.

More contemporary examples of short-circuit thought are everywhere. Without doubt, common interpretations of sensitivity training were most influential in recent days in developing this common view (e.g., Lakin, 1963) on turf associated with OD. Related social momentum for freedom as self-assertion also derived from the “drug culture,” as well as from the widespread reactions to Vietnam and urban warfare. All would be well enough, or at least far better, if institutions did not suppress the good that was in all of God’s children as their birthright. Bad institutions made good people crazy, in short. And attractive philosophies proposed the clear next step (e.g., Weick, 1979). Reality exists only as we enact it and, hence, we can rid ourselves of any or all historical detritus by unenactment. Man can lose those chains, in effect, by unenacting them, or enacting some more comely alternative. In its most developed form, this is the result held out by “affirmative inquiry,” one variant of “social constructionism” (e.g., Cooperrider and Associates, 1990).

These passages are not merely history retold at a distance. Indeed, your author himself almost adopted such a position (e.g., Golembiewski, 1979, 1993), but not quite; and he has elsewhere left an extensive mea culpa (Golembiewski, 1999). Being more careful here certainly will curb one’s style, but it has long been clear except to zealots that it matters profoundly who tells what to whom, why, and especially how (e.g., Culbert, 1973). Freedom as self-assertion can help lubricate such a process. However, that New Freedom falls fatally short of providing it with wise guidance and, far worse, can overload the total process.

The allures of freedom as self-assertion were widely aided and abetted, despite wise and early warnings to the contrary (e.g., Lakin, 1963). Beyond usefully emphasizing the significance of feelings and other aspects of what may be called “subjective realities,” some interpretations of this first meaning of freedom also were extended to the neglect of what might be called “objective realities”—extending to a denial of their relevance or even existence. Various forms of “social constructionism” provide the best contemporary examples of making too much of a useful notion. We create much of our own realities by enactment, as Weick (1979) reminds us, but to extend that useful view without limit leaves us all in the normative wilderness illustrated for us by The Lord of the Flies.

Establishing boundaries between usefulness and too muchness when it comes to “freedom” poses no insurmountable difficulties. Clearly, all of us define or “enact” many aspects of our own realities, and a vivid sense of that process can empower and embolden all of us. This puts power in the hands of everyone, but that potential has to be tethered short of a spurious freedom to change when we are seriously affected by previous multiple enactments, as when my percep-
tion, or interpretation, or evaluation of an event differ from yours. No need to just chafe under the weight of our past, in sum. We can just enact a more desirable future, that is the reachable ideal promised in freedom as self-assertion—as in social constructionism such as appreciative inquiry.

Despite the real attractions of freedom as self-assertion in various historic periods, then, this view was, and remains, seriously vulnerable. Even former ideological proponents now are coming to such a conclusion (e.g., Gergen and Thatchenkery, 1996). Moreover, elsewhere this writer goes into considerable detail to illustrate these vulnerabilities—empirical, theoretical, and practical alike (Golembiewski, 1999). Broadly, thought and experience combine to characterize this first concept as faux freedom, as some urged from the very start (e.g., Lakin, 1963). Hence, earlier ages insisted on the distinction between “freedom” and “license,” for example, with the latter providing unfettered but troubled reign for self-assertion. See also Chapter 25.

Overall, then, social constructionism clearly seems to simplify life, but in reality does so only by begging major questions of which self-assertions will serve common purposes, in what form, and when. At its worst, this first view risks making all of us victims of the self-assertions of others, and perhaps slaves of our own insistence on ignoble self-assertions. Practically, that is to say, insistence on self-assertion often leads to the awkward but stubborn position that might or manipulation make right. In technical senses, the critical limits of this first view were made clear by comparative research (e.g., Yalom and Lieberman, 1971), as well as by detailed prescriptions for effective personal action (e.g., Golembiewski, 1979, esp. chaps 1 and 2). Some kinds of self-assertion can hurt, in short, no matter how sincerely made.

Perhaps the most extreme insistence on self-assertion rests in the various forms of cultural relativity, which reflect a form of social constructionism (Golembiewski, 1999). Basically, that concept leaves us without standards and, hence, as fundamentally vulnerable. Oppositely, that is, recalcitrant realities also exist “out there” as well as “in here”—“out there” either in the sense of being what conventionally are called “objective realities” as well as in the sense of being subjectively defined realities that are “objective” in the sense of being variously beyond our efforts to unenact, easily or even at all.

All forms of cultural relativity sacrifice much by similarly urging that all truth-determining is entirely and resolutely “humanist,” that all truth-determining occurs only “in here,” in people. While useful for limited purposes, this basic view can easily undercut itself. Consider “value clarification,” once considered such a useful feature of change and development (e.g., Kirschenbaum, 1977), especially in schools. This potential was underexploited, in part because—as Leo notes (1999, p. 20)—“the qualifiers mistakenly assumed that most students . . . already had a moral grounding and just needed to sort out their own values.” Values clarification could not supply values, in short, and that was often the
probably growing need in the context of an often dominant “multiculturalism” or “cultural relativity,” which in effect threatened all freedom by the excesses of a purported equivalence of all values.

Our times constitute another point at which too much was made of a useful notion. Indeed, the tougher-minded among us might observe that the basic idea implies its own negation, and not only in the idea’s more extreme forms. That is, some might argue that the equivalence of values foretells the narrowed relevance of “clarification,” if not its irrelevance. This view proposes that man is both the measure and measurer of all nature, and this is a breath-taking simplification. As Pope John Paul II insists (1995, p. 3), this view neglects the fact that a vital difference-in-degree often moves toward a difference-in-kind, a difference he expresses in this way:

. . . it is a matter for serious concern that some people today deny the universality of human rights, just as they deny that there is a human nature shared by everyone. To be sure, there is no single model for organizing the politics and economics of human freedom; different cultures and different historical experiences give rise to different institutional forms of public life in a free and responsible society. But it is one thing to affirm a legitimate pluralism of “forms of freedom,” and another to deny any universality or intelligibility to the nature of man or to the human experience. The latter makes the international politics of persuasion extremely difficult, if not impossible.

In the several variants of self-assertiveness that extend to extremes, “freedom for what?” is either smoothly disregarded, or boldly finessed. Thus, some libertarians might propose self-assertion in all cases, except perhaps in those rarities in which others are hurt or (more narrowly) those cases in which my too-enthusiastic self-assertion acts to curb someone else’s license.

These rules-of-thumb may be practical, at times, and even necessary. But they nonetheless always and clearly beg the central and continuing concerns about defining some limits for freedom. Even the most strident libertarian would be unlikely to assert an unfettered right to yell “Fire!” in a crowded theater. Moreover, obviously limiting conventions clearly are required to indicate when people are “hurt,” when my self-assertion curbs somebody else’s, or when the threats of “hurting” or “curbing” are imminent enough in principle as to be prudently proscribed in practice.

To urge that there should be no limits on self-assertion—whenever, whatever, or however—clearly implies the real possibility of life as a war of all against all, an especially ironic result in the context of the first variety of freedom’s intent to increase personal reach-and-grasp. Where everyone can fully rely on self-assertion, no one can place any reliance on how their own self-assertions will be reacted to. In this case, the insistence on freedom-as-assertion, in direct pro-
portion to the degree of the insistence on it, negates experienced freedom. Not
to be churlish, but one must note that this negates what freedom-as-self-assertion
originally had in mind.

Hence, while seeming to settle all issues of who polices which limits, free-
dom as self-assertion only exacerbates what will always be difficult. Thus, it may
be in one’s interest to urge that a woman has “a right to control her own body,”
but that oversimplifies the issues of the impact on the rights of numerous real or
potential others. And these multiple impacts require attention, however tentative
or temporary. Thus, fetuses in the first trimester may be considered as non-right-
holders, as a matter of public policy, but other boundaries can be necessary and
terminally consequential.

“Freedom” as Normatively Grounded

Whether we acknowledge the reality willingly or kicking and screaming, then,
freedom as self-assertion has obvious conceptual shortfalls. This analysis thus
must extend its review of freedom in order to avoid being consumed by it.

My earliest impulse—perhaps “organizational humanism” best describes
it—usually led me to compromise in the cause of progress toward more humane
organizations. For example, in an early text on OD (Golembiewski, 1972), my
surrogate for moral grounding was a sufficient and simultaneous congruence of
major stakeholder interests—that is, “success” was defined as sufficiently meet-
ing individual and group needs as well as organization demands. Substantial ac-
tion research essentially was vetted in such terms, including:

Flexible workhours (e.g., Golembiewski, Hilles, and Kagno, 1974);
Confrontation designs as a basic model for team-building (e.g., Golem-
biewski, 1972, pp. 39–44, 291–298);
3-Dimensional image sharing for such purposes as conflict management
(Golembiewski, 1979, vol. 2, pp. 139–141, 191–194); and
Reorganizing a medical-surgical ward in a hospital (Golembiewski and

Later chapters give detailed attention to such approaches to freedom-as-norma-
tively-tethered.

In such work, more essentially, my working view moved toward OD as
generating a greater degree of responsible freedom for people—people in organi-
izations as well as people affected by them. Here, the operating word is respon-
sible, which directly raises that difficult question: Freedom for what? That is, some
freedoms can be irresponsible—that is, mere license—even as other freedoms
can be rooted in values that can lay real claims to universality and transcendence.

To illustrate, Pope John Paul II faced similar problems with the legacy of
the “freedom” he inherited from Vatican II, and which he saw as reflecting the
contemporary human challenge. How to retain “freedom,” and yet avoid mere self-assertion? In the words of one of his close observers (Weigel, 1999, p. 56), the Pope sought to normatively ground his version of “freedom” in Christianity—”... to vigorously underscore that freedom be ordered to truth and [also] finds its fulfillment in goodness.” Such grounding is never easy, the Pope realizes; but rejecting or finessing that point inevitably will be destructive.

In effect, at least in concept, Pope John Paul II at once perseveres in the pursuit of “freedom” and yet also provides a working set of value and ethical boundaries for its profound and progressive exploration. Put in colloquial terms, the Pope sought a balance between mere “copping-out” and “going to far.” Like John Milton, the Pope saw “freedom” as torn to shreds and tatters by modern life. So careful search was the order of the day, lest the form of the new freedom be poorly or coercively assembled. Perhaps unlike Milton, however, the Pope also offered a historic template for guiding the search for the tatters to be reassembled. This template provided guidance while also encouraging charity, tolerance, and forbearance for the no-doubt extended period in which consensus about the reassemblage would be forming and being tested. The double-barreled goal was to avoid a sense of “freedom” as merely conventional and relative while also guarding against the reality that history contains tragic examples of the good intentions of how many ethical traditions were inhumanely extended.

“RESPONSIBLE FREEDOM” AND OD

OD always has been grappling with its own version of “freedom for what?” at least in my view, and attention now turns to my related efforts to give form to this normative grounding while increasingly vivifying “responsible freedom.” In the most direct terms, these efforts in this and the two following chapters trend away from “freedom” seen as self-assertion and toward “freedom” as normatively-rooted by OD values in the context of other widely accepted systems of thought. Four features provide perspective on these central developments.

Macro-Level Normative Frame

From its inception (e.g., Bradford, Benne, and Gibb, 1957), OD intends to distinguish itself from conventional characterizations of traditional approaches to the person-at-work. For example, numerous contrasts are possible between Table 7.1, columns A and B, and the bureaucratic model: OD sees the person in a network of reciprocal social and psychological identifications, while bureaucracy is rightly criticized for its emphasis on the smoothly working machine and on “economic man” (e.g., Golembiewski, 1995).
Let my point be made plain. The macro-features in Table 7.1, column A, constitute one way of answering the question: Freedom for what? This at once distances “freedom” from excesses, and also details skills for real-life applications.

### Regenerative Interaction as Ideal

These macro-level features, in my view, are in most OD applications usefully reinforced by various values/skills at micro-levels (see Table 7.1B). Thus, the dynamic infrastructure supporting regenerative interaction is seen as a recursive model of four variables suitable for reinforcing macro-level goals via structures, policies, and procedures. That is, referring to Table 7.1A, the macro-level values of OD clearly are compatible with regenerative interaction—e.g., in giving life to an attitude of real inquiry, in expanding choices, and so on.

In contrast, a LoLoHiLo profile—or degenerative interaction—will tend to characterize bureaucratic forms of organization (e.g., Golembiewski, 1995). There, for example, jurisdictional squabbles will dominate, and they will reduce trust, openness, and owning, as well as increase risk, among other consequences (Ashkenas, Ulrich, Jick, and Kerr, 1995). The consequences of regenerative versus degenerative interaction are developed in detail at several points (e.g., Golembiewski, 1993, vol. 1, pp. 55–72 and 163–169).
biewski, 1996). Conveniently, it requires only a little effort to detail the probable effects of, for example, low trust in organizations both large and small which aspire to Table 7.1A values.

Again, let us be plain about column B in Table 7.1. In a second way, that normative context provides both tether and scope for OD seen as “responsible freedom.”

**Praxis and Normative Groundedness**

The crucial related point gets only brief mention here, but Table 7.1 supports a range of techniques or approaches—for example, giving or receiving feedback, which is at the heart of the development of basic OD theory (e.g., Golembiewski, 1993, pp. 37–72). Basically, I report, Table 7.1 helps answer the query “freedom for what?” in detailed terms, and thus provides guidance for what is (or is not) OD.

**The “New OD”**

Truth to be told, the normative tethers in Table 7.1 do not always get priority attention now, but once they were widely seen as critical “pre-work” underlaying organization change via OD (e.g., Golembiewski and Carrigan, 1970). As in “work out” or in “future conferences,” for example, many commentators nowadays prefer to assume that the required “culture of inquiry” can be easily and consistently attained without explicit design (e.g., Ashkenas and Jick, 1992). This seems too convenient, and may discourage precisely the normative guidance that this volume seeks.

Far more broadly, others not only question such normative boundaries but also go on to propose that those boundaries should be deliberately stretched (e.g., Beer, 1989), or even that they should be done away with so as to enhance OD’s usefulness to management. Such approaches may lead to the expected effects, but the present position here is that there is no certainty that the body of theory and experiences now reasonably labeled OD will survive. Hopefully, the present emphasis on “freedom for what?” will help restrain these recommendations for expanding OD’s boundaries, or for restraining efforts to pull OD out of shape.

In the present view, the “New OD” must avoid both the finesses illustrated above, as well as rise above alluring entreaties to loosen up a bit about normative boundaries. In combination, although they should be substantially inconsistent, the finesse and the loosen-up approach constitute a powerful combination to be avoided.

The pathway for the New OD will neither be an easy one nor a matter of avoiding suggestions to release its normative tethers. Proactively, to put the matter in essentially opposite terms, the two chapters that follow will expose OD to two types of tests: to introduce three worksite ethics that can help shape and
Responsible Freedom as the Goal of OD

Maintain the normative form of OD, and to begin the test of these alternative ethics against the success rates of OD applications, following early chapters of the second edition. The foci will be multiple: the what and the how and the why in which OD as a technology-cum-values can operate within several broad normative contexts. The demonstration is of necessity preliminary, but it is an unavoidable even if inadequate beginning. In sum, a reasonable case must be made for the substantial degree-of-fit of OD ways-and-means to alternative work ethics. That will constitute our working answer to this question: Freedom via OD for what?

So the OD ethic is not viewed as a closed system, basically. It will be tested against, as well as refined in, relationally other normative systems that have given shape to our lives, and can be relied on to guide the future development of the OD ethic. The next two chapters illustrate this effort to provide detailed treatment of: “Freedom for what?”

REFERENCES


With Chapter 7 having set the outlines of the case for “responsible freedom” as the foundation for OD, the question becomes whether or not that basic idea has been put to a real test. Here, that task involves four components: to sketch the value-loaded OD work ethic that can be said to constitute a reasonable normative framework for some half-century of theory and application; to suggest that a Judeo-Christian work ethic can be said to characterize North American management; to draw a convincing, if hardly complete, picture of reasonably good fit between the major features of those two ethics; and to present the rationale for substantial success rates characterizing that reasonably good fit.

This trio of chapters deal with the issue of “freedom for what?” The reasonableness of this analysis is first tested, in effect, by assessing the degree to which two widely accepted Western ethics serve to provide effective guidance for OD’s greater responsible freedom.

A CONCEPTUAL CONTEXT FOR IRONY III, CONTINUED

As the second in a trio of chapters highlighting the integration of OD with several “big ideas” of life and work, this chapter highlights the reinforcing support of two work ethics with a dominantly “western” history. Those work ethics include an OD ethic as well as a Judeo-Christian ethic, with the latter being a conceptual foundation for work of long standing, and the former being a more recent set of values-cum-techniques to provide direction for living the good life at work.
The interactive goal should be clear. The Judeo-Christian ethic ties us to our past and to its insights concerning collective life, and the OD ethic presents a contemporary guide for people in organizations.

To the degree that this interactive heightening of effects of two ethics can be demonstrated, this will root contemporary views in ideas that gained allegiance in the past and still retain influence. That demonstration also will show how current ideas can build upon long-standing ideals.

This task can only be begun here, and at best involves major multiple contingencies. If nothing else, to illustrate the energetic but hopefully-intelligent argument below, the estimation of North American OD success rates certainly is still a work-in-progress and may remain so for a long time. Nonetheless, this chapter details a substantial search for applications whose success rates can be evaluated. If incomplete, the present estimate is certainly no straw man. Fortunately, Chapter 7 already has begun some of the heavy work on which this chapter will depend.

In introduction, then, this chapter extends the conceptual development of “freedom” begun in the preceding chapter and slated for completion in a following chapter. Here, the focus is on two broadly Western work ethics to see if they can provide normative grounding of Organization Development, or OD, approached as “responsible freedom.” The effectiveness of this tethering by values is also suggested by the reviews of success rates in Chapters 1 through 3. These estimates, in effect, indicate the usefulness of the normative grounding of OD, which, in turn, implies the substantial degree to which those underlying values are desired and desirable at a large number of Western worksites. A following chapter will begin to test OD approaches against a third work ethic—a Confucian ethic whose most prominent home is East Asia.

TWO WESTERN WORK ETHICS AS THE NORMATIVE BASE

Responsible Freedom II has a direct but involved task. Given the inadequacies of freedom as self-assertion, is it possible to develop a satisfactory normative grounding for “responsible freedom”? The challenge seems clear enough, and has these components:

To describe a normative framework for OD, an ethic that can lay reasonable claim to practicality and broad acceptance, if not necessarily general acceptance

To describe a second ethic, the Judeo-Christian ethic, which is proposed as representative at least in outline of the broad value tethers common in North American and Western settings

To assess the consistency of the Judaeo-Christian and OD ethics
To illustrate the approaches and techniques associated with the normative framework of the OD ethic, which relates, among other targets, to macro- and micro-levels of organizational change.

To establish that major operational as well as philosophic purposes can be served by the OD ethic as well as by the approaches/techniques associated with that ethic—that is, that the OD ethic well serves the Judeo-Christian ethic.

In sum, these objectives require, in significant regards, going beyond Pope John Paul II’s treatment of freedom outlined in Chapter 8, while falling far short of his standards in other regards. That is to say, the Pope displays little concern with approaches/techniques for organizational change, let alone for their application in public agencies and businesses. Rather, the Pope’s focus is on rooting “freedom” in “human nature”—literally as a gift of God to all humanity. Here, the focus is at once more limited but also more practical than the Pope’s. Directly, the Pope’s work requires additional specification and reinforcement here. In addition, the test of truth implied here is selective and limited—“broad acceptance,” the text above notes, “if not necessarily general acceptance”; and the text also refers to various “operational” and “practical” consequences thought of in summary form as “success rates.” In sharp contrast, John Paul II deals with a universal normative standard.

To illustrate the contrast, this chapter concerns only work ethics, and in the bargain only two “Western” ones at that, clearly a deliberately limited perspective. In contrast, the Pope’s scope encompasses the essentials of humanity, always and everywhere, given the spark of divinity in all people. Rooting “freedom” in the “human nature” with which God in the Christian tradition has gifted humankind, the Pope explains that:

Freedom is not simply the absence of tyranny or oppression. Nor is freedom a license to do whatever we like. Freedom has an inner “logic” which distinguishes it and ennobles it: freedom is ordered to the truth, and is fulfilled in man’s quest for truth and in man’s living in the truth. Detached from the truth about the human person, freedom deteriorates into license in the lives of individuals and, in political life, it becomes the caprice of the most powerful and the arrogance of power. Far from being a limitation upon freedom or a threat to it, reference to the truth about the human person—a truth universally knowable through the moral law written on the hearts of all—is, in fact, the guarantor of freedom’s future (Weigel, 1999, p. 5).

Much the same could be written of other moral traditions, of course, but this second edition takes a convenient approach. Note also that Chapter 9 extends
major elements of this analysis to what is called a Confucian work ethic, which begins the comprehensive approach of this book to seeking normative grounding.

TWO WESTERN NORMATIVE FRAMEWORKS FOR RESPONSIBLE FREEDOM IN CHANGE

Let me recapitulate here in the service of an easy take-off, if at the possible cost of some repetition. This and the preceding chapter, yoked by a commitment to responsible freedom, highlight two Western normative frameworks for guiding change. They introduce two sets of values: the first relates to normative guides for change via OD, and the second provides micro-level values for guiding the interactions that provide structure and substance to large-scale management. More specifically, this chapter generates an OD work ethic consistent with such design details, and compares its fit to normative guides from a Judeo-Christian work ethic, which has a broad acceptance in Western settings (Golembiewski, 1965, 1993).

Together, then, these three chapters in their own ways seek to contribute to providing normative standards for the question: Freedom for what? Or put another way, the question becomes: What is OD’s view of “responsible freedom”? The OD Work Ethic and the Judeo-Christian work ethic each get attention below, in turn, for the working answers they provide to how OD fits into long-accepted tradition.

So, let’s get on it with complex cross-checking. Conveniently, Table 8.1 provides summary perspective on two central “work ethics.”

OD Work Ethic as Internal Normative Check

As OD evolved, largely in North America and in other Western representative systems, it became identified with a broad worldview—a weltanschauung, if you will. Column A in Table 8.1 sketches one version of this worldview common to OD and ODers, arrayed in terms of ten categories or themes. In total, column A provides a sketch of OD’s normative framework.

Several features of these OD normative themes deserve emphasis. Thus, this version is sui generis and was developed for the present purpose, but substantial consensus about it exists. In addition, the descriptions highlight the particular ways in which OD is value-loaded, or should be. This underscores the central fact that OD is at once a technology but also more than that. OD is more than a tool-kit of approaches and designs, if you will; rather, OD is a technology-with-values. Consequently, how a particular design is applied is equally as consequential as the character of the design itself. Table 8.1 permits, and even more so requires, the cross-checking of ten perspectives on this “how.” Not every OD
<table>
<thead>
<tr>
<th></th>
<th>A. OD Work Ethic</th>
<th>B. Judeo-Christian Work Ethic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. View of person</td>
<td>OD-1 Persons are perfectable, if often flawed by institutions.</td>
<td>JC-1 Persons are in essence flawed but salvageable by faith and good works, usually in combination; but some J-C ideation emphasizes faith.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Dominant concept of ideal state</td>
<td>OD-2 Humanist-centered, as in awareness of reactions by self and others to help guide and evaluate the development of more effective behaviors/attitudes.</td>
<td>JC-2 Basically rests on transcendent or divinity-centered ideals, but with the practical realization that individuals often will exist at several “levels of perfection” where humanist variants of the Golden Rule provide guidance: Do unto others as you would have them do unto you.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Dominant guide for choice</td>
<td>OD-3 Requires knowing the self in relation to others, and hence emphasizes empathic sensibilities as well as feedback and disclosure skills, so as to work toward the increasing realizations of needs similar to those postulated in OD’s early history by Argyris (1957) and others.</td>
<td>JC-3 The “spark of divinity” in each person provides the potential for coming to know transcendent ideals, but in practice this requires knowing the authentic self in relation to and with others. Progressively coming to act on that growing knowledge also requires relationships with others, via processes similar to OD’s emphases on feedback and disclosure involving individuals described by models of needs like those in humanist approaches.</td>
</tr>
<tr>
<td></td>
<td>A. OD Work Ethic&lt;sup&gt;a&lt;/sup&gt;</td>
<td>B. Judeo-Christian Work Ethic&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>4. Basic character of change</strong></td>
<td>Change is “linear-progressive” (Marshak, 1994, pp. 402–407): i.e., it intends movement from one state to another more desired by persons; is destination-oriented; often deliberately alters the equilibrium of the status quo; people plan and manage it to achieve their own goals; and change is difficult because much of life is in equilibrium, even if unsatisfactorily so to many.</td>
<td>Change is “linear-individual,” with the ideal being movement from a postulated “original sin” to levels of perfection as defined by transcendent values, with the levels being “final”—beyond human alteration, attainable only in degree, if immutable.</td>
</tr>
<tr>
<td><strong>5. Character of work</strong></td>
<td>Work must be psychologically acceptable to the individual, while collective requirements also are met, at least in the long run.</td>
<td>Work must be psychologically acceptable to the individual—that is, its performance should not generally threaten the individual, who is often in possession of a working definition of the desired/desirable that can be surfaced by the processes involved in J-C 1-4.</td>
</tr>
<tr>
<td><strong>6. Orientation to personal development</strong></td>
<td>Work often will permit developing individual faculties if group and system issues are taken into account.</td>
<td>Work must allow people to develop their faculties, consistent with an underlying sense of human needs generically like those in OD (e.g., Argyris, 1957).</td>
</tr>
<tr>
<td></td>
<td>OD-7</td>
<td>JC-7</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>7. Orientation to personal autonomy</td>
<td>Tasks at work must allow individuals considerable self-determination, and especially through groups.</td>
<td>Tasks at work must allow individuals considerable self-determination.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OD-8</th>
<th>JC-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Orientation to environmental influence</td>
<td>Group mediation often will influence or determine individual control over the work environment.</td>
<td>Individuals must have the possibility of controlling and influencing, in meaningful ways, the environment within which tasks are performed.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OD-9</th>
<th>JC-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Final arbiter in work</td>
<td>Many stakeholders will be involved in decision processes, where bargaining often will be dynamic, problematic, and involve multiple collectivities.</td>
<td>The organization should not be the sole and final arbiter of behavior; both individuals and the organization must be subject to an external moral order, but in practice this places great weight on individuals to resist illegitimate demands or commands.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OD-10</th>
<th>JC-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Roles for intervention in work</td>
<td>Intervention has a high priority—as in managerial roles that contain strong elements of change-agentry, and perhaps especially as part of everyone’s roles in regenerative and self-correcting systems.</td>
<td>The focus here is on the heroic individual, morally directed—e.g., as in entrepreneurial roles, or as in “ethical resistance” by “whistleblowers,” which may be aided by institutional features—e.g., via personnel departments or unions as well as legislation.</td>
</tr>
</tbody>
</table>

---

* From numerous sources, especially Golembiewski, 1993.

* Based on, but extending far beyond, Golembiewski, 1989.
design or effort will achieve an integration of all ten perspectives, but the implied goal is to maximize that integration in each application.

**Judeo-Christian Work Ethic as External Check**

The internal cross-checking features of Table 8.1 have their attractions, but does OD imply substantial congruence with the external normative system here labeled Judeo-Christian work ethic? Column B in Table 8.1 permits an approach to such a judgment, if a limited one. The normative framework there—the Judeo-Christian work ethic—is “external” in the sense that its characteristics and development are not only prior to and independent of the OD work ethic, but also that these characteristics and development have been broadly (but not exclusively) accepted as major features of Western normative traditions. For example, note that in column A humans are the measure of all things, as well as the measurers of all things. This basic humanism contrasts sharply in principle—if less so in practice—with deontically based column B. There, clearly, an external standard is implied. Broadly, the cross-checking permitted by Table 8.1 will prove of major usefulness in helping assess the degree of fit between that table’s two columns.

Does such an external check imply a substantial degree of tethering of OD to normative perspectives that have been widely accepted, even if not universally? The working answer, as details will soon show, is quite definite: Yes, overall. Certainly, however, this approach does not settle for all people and for every time the issues associated with our key question: Freedom for what? Nonetheless, demonstrating a substantial even if incomplete congruence between the OD and Judeo-Christian (JC) ethics will involve two important developmental traditions in several ways:

For humanists, the issue centers around the degree to which OD is normatively congruent with the features characteristic of people at work at their best—which is to say, at their most need-fulfilling.

For those emphasizing universal human rights, the issue is more the degree to which OD is normatively congruent with a “divine spark” view of humankind reflected in the Judeo-Christian work ethic.

Is the “fit” close-enough to play mutually supporting roles in work and life?

Much can be said that provides evidence about OD acceptable to humanists as well as deontists or universalists. One version of that demonstration dominates below with, as noted, the comparative reliance falling basically on the Judeo-Christian work ethic. Other commentary (e.g., Golembiewski, 1965, 1989) shows that the attributes listed in column B are consistent with deep religious and social traditions in North America, and have great influence elsewhere. The deontic approach of John Paul II is substantially the same as that of the JC ethic.
A Major Check on Congruence

Overall, inspection of Table 8.1 reveals that the JC and OD work ethics overlap substantially, although hardly completely. Perhaps the basic apparent difference, already noted, relates to the issue of the humanist bias of OD, as contrasted with the transcendent rooting of the JC Ethic.

Here, let us give explicit attention to this central feature of the two ethics. In principle, OD-1 and OD-2 vs. JC-1 and JC-2 suggest a major dysjoint in principle; in practical terms, however, substantial overlap exists. Thus, the J-C ethic resonates with the general sense of the universal human rights that Pope John Paul II emphasizes. In practical terms, moreover, both the JC and OD ethics often come down in a similar place, as OD-3 and JC-3 especially suggest. One speaks of the “divine spark” in people; and the other often emphasizes, and especially in OD’s earlier days, human needs that are seen as broadly shared, if not universally so (e.g., Argyris, 1957). The spirits in the machine seem quite similar. In addition, when push comes to shove, practically, both ethics depend on feedback and disclosure processes of individuals in dialog, with the shared goal of moving toward a clearer appreciation of self-in-community.

Differentiating the sources of the two ethics as deontic and humanist can have consequences for some purposes, but the two ethics also share major parallelisms that can comfortably guide practice even with the difference in source. Thus, the argument of Pope John Paul II is rooted in the divine creation of man, while humanists focus on developmental similarities somehow reflected by all or most individuals but—a la Argyris (1957)—without identifying the source or sources guiding this development.

While thus essentially differentiated, the JC and OD work ethics nevertheless share many detailed features, and this permits—even encourages—practical crosswalks between the two ethics such as those targeted here. Put directly, readers might consider this list of desired/desirable human features, among others:

- To realize a broadening set of skills and abilities
- To develop widening time horizons
- To decrease defensive and compulsive behaviors
- To increase cooperative and independent behaviors

This writer doubts whether readers can reliably distinguish whether the source of these features is deontic or humanist. Moreover, for many practical purposes and for most of principle, that difference in sources is neither obvious nor crucial. The similarities or parallelisms of humanist or deontic versions of humankind can take us a long way toward defining desired/desirable forms of the just and satisfying organization.

Relatedly, when it comes to the basic issue of change, similarities between the two work ethics in Table 8.1 also seem more prominent than differences. To
be sure, the OD ethic might give more attention to quasi-stationary equilibrium involving groups, and the JC ethic might place greater emphasis on individuals' striving toward a transcendent knowledge, always out of reach but nonetheless continually motivating effort. In practice, however, both ethics see the individual as flawed but salvageable, and as coming to know the self and others in substantially similar ways. Indeed, it is at once easy in principle and tempting in practice to see “universal human rights” and “humanist needs” as sharing much conceptual territory. Both ethics also come to be appreciated via dialog, given that the one can be seen as definitely, even defiantly, rooted in divinity and the other in humankind.

In other particulars, much the same picture emerges from other comparisons between the two columns of Table 8.1: apparent differences in principle can be accommodated in practice, at least substantially. As one example, consider the “basic character of change.” Here, OD-4 vs. JC-4 reflects a basic difference. Thus, one can imagine a situation in which both employee needs and organization objectives are being met simultaneously and in substantial degree, and in which “good enough” characterizes the meeting of both needs and demands, with a quasi-stationary equilibrium resulting. Here, ODers consequently might let well enough alone. In the same case, supporters of the JC work ethic may be characterized by an always-loftier view of the desired/desirable. In this second case, in contrast to the OD framework, the pressure for change presumably would continue, and any equilibrium should be more or less constantly disturbed in principle, so as to reach higher and higher.

Real tensions between the two work ethics are easy enough to imagine in principle, then, but practice implies many situations in which OD-4 and JC-4 (for example) could be accommodated. Thus, periods of consolidation can be expected in both JC and OD, and hence quasi-stationary tendencies might at times characterize both ethics. Similarly, there might be extended periods in which continuous change, or basic transformation, is broadly viewed as desired/desirable, whatever the dominant ethic. During such intervals, OD versus JC contrasts are moderated.

Not at all coincidentally, my personal ideal in OD is precisely the case highlighted above—when individual needs and organizational goals are not only being met, but when that condition is the dominant target. This choice suggests one approach to balancing humanist and deontic orientations.

**OD SUCCESS RATES**

Chapter 9 will return to a comparison of these two ethics and also will introduce yet a third, but here note only that the success rates of OD applications—in Western settings, as well as elsewhere—suggest that it is advantageous to rely on the OD work ethic as a major technology-cum-values for change as well as
for working toward the JC ethic. Let me put the essential point in plain terms. The OD ethic not only seems appropriate for acting on that set of values embodied in the Judeo-Christian ethic, but that OD ethic also generates attractive effects in a substantial proportion of cases. In short, the simultaneity suggests that the OD work ethic moves toward meeting broadly accepted principles while generating attractive practical consequences, as defined by both management and workers. To illustrate, “success” in an OD application typically means:

- Heightened employee satisfaction
- Greater job involvement
- More effective performance
- Lower rates of burnout and stress, among many other effects

This list suggests the attractive combinations to which OD leads, far more often than not.

Two estimates of the success of OD will be relied on here to demonstrate the efficacy of Western applications, largely in the United States (see Table 8.2). The first batch of applications goes back some years; and the second source amounts to a survey of surveys that permits estimates that are quite current. Generally, also, the “success rates” refer to a broad range of intended effects—increased productivity, decreased scrap, heightened satisfaction and morale, as well as many other outcomes conventionally seen as desired/desirable. In some estimates of success, over 300 outcomes were assessed to determine if intended effects occurred as a result of specific OD interventions (Porras and Berg, 1978).

Roughly, the two kinds of estimates of success cover applications during the interval 1970 through 1998. Details are not provided here, but they are conveniently available elsewhere (e.g., Golembiewski, 1998, 2000a). See also Chapters 1 and 2.

**Table 8.2** Summary of Success Rates in Western Settings, Largely United States (N = 574)

<table>
<thead>
<tr>
<th>Categories of Rated Success</th>
<th>Public Sector</th>
<th>Business Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Highly positive and intended effects</td>
<td>41%</td>
<td>40%</td>
</tr>
<tr>
<td>II. Definite balance of positive and intended effects</td>
<td>43%</td>
<td>49%</td>
</tr>
<tr>
<td>III. No appreciable effects, or balance of intended and contrary effects</td>
<td>7%</td>
<td>5%</td>
</tr>
<tr>
<td>IV. Contrary effects, and especially even those applications with a small proportion of estimates in a contrary direction that attain statistical significance</td>
<td>10%</td>
<td>6%</td>
</tr>
</tbody>
</table>

Early Success Rates

Two landmark studies dating from the 1980s (Golembiewski, Proehl, and Sink, 1981, 1982; Nicholas, 1982) reflect substantial success rates—indeed, formidable ones. Details of measurement will not be repeated here. The second study (Nicholas, 1982) surveys a much smaller panel of evaluations of OD applications and, although its criteria for selecting applications are more selective, the results are consistent with the Golembiewski research.

Replicating Success Rates

A recent study (Golembiewski, 1998) establishes that these two sets of early estimates are no flash in the pan. About sixteen separate surveys of applications were found and, in the aggregate, they reinforce attractive expectations about OD effects. Despite some double-counting, this survey of surveys encompasses thousands of cases. See also Chapters 1 and 2, which make the important general case that the OD success rate estimates relied on here are not narrow sectarian products.

REPRISÉ

All the seams in this analysis are not water-tight, then, but the pattern is attractively suggestive, if not definitive. Chapter 7 thus describes a value-loaded framework for change usually called Organization Development, or OD; and here OD approaches and techniques are tested against “internal” as well as “external” normative standards. That is, the OD ethic and the Judeo-Christian ethic seem quite consistent for operational and practical purposes, and both are well served by OD as a value-loaded technology for change, as judged by success rates estimated by multiple investigators, using different estimates of success, in applications covering the interval from about 1970 through recent days.

The following chapter will extend this analysis in two ways. It adds to the comparative analysis a third normative framework—the Confucian work ethic. And that third in this set of related analyses also reports on success rates, both from East Asian loci as well as from more numerous applications in nations reasonably described as “developing.” In general, this analysis implies that OD as a technology for change can be accommodated to work settings probably loaded by Confucian values in one collection of applications, and also encompassing 56 nation states in a second collection that more broadly represents “developing” nations.

REFERENCES


9

Responsible Freedom as the Goal in OD,
Part III
Confucian Work Ethic as the Base*

The goal here is transparent: It is to suggest the broad reach-and-grasp of OD to an ethic whose homesite is far displaced from that of the two work ethics considered in Chapters 7 and 8.

This demonstration cannot be definitive, but your author sees it as not only illustrative but as powerfully suggestive. Again, success rates of OD in a Confucian work setting—Korea—reinforces the suggestion of OD’s efficacy and flexibility in contributing to a growing responsible freedom.

This third chapter on a theme completes this line of commentary concerning a normative framework for OD, if only temporarily. A first chapter conceptually isolates our guiding normative goal of greater responsible freedom; the following chapter shows that OD approaches and theory can lay reasonable claim to fitting two normative templates with Western underpinnings, and with attractive practical effects. This final chapter presents a similar analysis for an East Asian normative framework—the Confucian work ethic.

Similar analyses could be applied much more broadly, of course. Clearly, the analysis could be extended to other alternative normative frameworks and to estimates of the success rates of OD applications on turfs within which alternative normative frameworks can be served. Much more expansively, attention could be directed at the kind of broader environment—social, political, and economic—that has a good fit with the several ethics considered here, plus others.

* Jong-In Yoon and Seok-Eun Kim contributed in meaningful ways to this chapter.
But enough for now. This trinity of chapters may encourage similar macro-extensions of the present effort. For example, this trinity restricts itself to the organization level, as distinguished from the levels of the nation-state or the inter-organization level.

CONFUCIAN WORK ETHIC AS THE BASE

This chapter develops some useful conceptual distinctions about OD as a value-loaded technology for approaching “responsible freedom.” Earlier, also, Chapter 8 illustrates the many senses in which OD is normatively tethered, not only internally by the OD work ethic but also externally by the complex ways in which this ethic compares with the Judeo-Christian (JC) work ethic. Not incidentally at all, this third chapter in a trinity also establishes the highly positive practical effects of this normative tethering—the high success rates of OD applications.

In short, OD not only proposes answers to the normative question, “Freedom for what?” but also implies major practical advantages of this tethering.

One major task concerning responsible freedom has not been addressed in these two chapters and, Colossus-like, that unfinished business dwarfs what this clusters of chapters can accomplish. To wit, to which other ethical sets can OD be said to apply?

This third chapter takes a further step toward establishing the double-barreled normative and practical features of OD, and when this chapter is done we will know more about the reach-and-grasp of OD applications. This time, the focus is on the Confucian work ethic—specifically, on its normative congruence with the Judeo-Christian work ethic as well as with the OD work ethic—far from exhaustively, but nonetheless substantially. Again, if more provisionally, success rates—in South Korea, specifically, and then in global OD applications—indicate the same happy practical outcomes of OD’s normative tethering even in national contexts in which the Judeo-Christian work ethic and the Confucian ethic presumably have smallish direct influence.

Let us get on with it, then. This chapter takes a large step forward. It asks, can the evidence of substantial OD success rates be extended trans-nationally, after assessing the overlap of three sets of leading normative guides relevant to managing work and planned change? Here, basically, an OD ethic is compared with a Judeo-Christian ethic as well as with a Confucian work ethic, with the cross-referencing being intended as a kind of sensitivity test for a sufficient congruence of the three ethics to encourage the hope that OD applications will have substantial success rates even in work contexts other than Western ones presumably influenced by the JC ethic.

Please underscore the tentativeness of this extension. Thus, no one of the “ethics” is comprehensively developed in this series, nor is any one logically consistent in all of its features. Moreover, the author associated with all three
Selected Differences Between Ethics

Tentativeness aside, Table 9.1 reflects preliminary but suggestive forms of three ethics relevant to work, and two general points can be made in summary. On the one hand, differences between the three ethics seem clear enough in several cases. Thus, most obviously, the three ethics differ in their fundamental views of humankind, with “original sin” making Judeo-Christian-1 the outlier distinguished from two humanist variants in Table 9.1. Even in this case, however, the three ethics share substantial conceptual territory. To illustrate, the Christian sense of “original sin” notwithstanding, the JC ethic preserves a substantially optimistic view of human nature, as do the two other ethics distinguished below. Broadly, this basic conceptual overlap in the three ethics is consistent in a general sense with the comparative and attractive OD success rates illustrated later.

The same compound point applies to Judeo Christian-2 in Table 9.1 as contrasted with OD-2 and C-2. Their conceptual overlap is undoubtedly incomplete, but the shared humanist-centered character of the Confucian and OD ethics further suggests the real potential for a good fit with OD designs. Chapter 8 already has developed the senses in which the JC and OD work ethics share normative meaning. As for the JC ethic, to complete this introductory triangulation comparison, its transcendent moral base shares basic conceptual territory with the humanist orientation of the Confucian and OD ethics. In the most obvious sense, and especially in JC-2 and JC-3, the principles of the Judeo-Christian work ethic move toward reliance on the humanist ideals consistent with the OD and Confucian ethics. Moreover, especially but not uniquely, Judeo Christian-2 and Judeo Christian-3 in practice imply that we come to know the specific features of “the human” in much the same ways in which the OD and Confucian ethics prescribe that we isolate humanist-related features. To be sure, of course, these features in one case are of divine origin, and in the other are somehow basically common to humans as they have evolved. In sum, the “needs” that Argyris (1957) and others attribute to (most or many) humans on the basis of studies of human development are close kin to the specific attributes usually associated with the “spark of divinity” view of humans, despite their different sources. Thus, persons in all three ethics share a similar developmental profile as they come to know self by transcending self with the help of others. See especially items 3, 4, 6 and 7 in Table 9.1.
<table>
<thead>
<tr>
<th>I. Judeo-Christian work ethic&lt;sup&gt;a&lt;/sup&gt;</th>
<th>II. OD work ethic&lt;sup&gt;b&lt;/sup&gt;</th>
<th>III. Confucian work ethic&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. View of person JC-1 Persons are in essence flawed but salvageable by faith and good works, usually in combination, but some J-C ideology emphasizes faith.</td>
<td>OD-1 Persons are perfectable, if often flawed by institutions.</td>
<td>C-1 Persons are capable of finding and maintaining a harmonious balance in their lives; in general, their “essential” self is heavenly but their “existential” nature may be good or bad.</td>
</tr>
<tr>
<td>2. Dominant concept of ideal state JC-2 Basically rests on transcendent or divinity-centered ideals, but with the practical realization that individuals often will exist at several “levels of perfection” where humanist variants of the Golden Rule provide guidance: Do unto others as you would have them do unto you.</td>
<td>OD-2 Humanist-centered, as in the awareness of reactions by self and others to help evaluate and help guide the development of more effective behaviors/attitudes.</td>
<td>C-2 Humanist-centered, as in the centrality of awareness of self and others to attain and maintain a harmonious balance for all in “families” of small/large size, where personal relationships beget ethical responsibilities, with the family as central (Ching, 1977).</td>
</tr>
<tr>
<td>3. Dominant guide for choice JC-3 The “spark of divinity” in each person provides the potential for coming to know transcendent ideals, but in practice this requires knowing the authentic self in relation with others. Progressively coming to act on that growing knowledge also requires relationships with others.</td>
<td>OD-3 Requires knowing the self in relation to others, and hence emphasizes empathic sensibilities as well as feedback and disclosure skills, so as to work toward the increasing realizations in dialog of needs similar to those postulated in OD’s early history by Argyris (1957) and others.</td>
<td>C-3 Basic loyalty is to transcending one’s self, to finding the authentic self capable of attaining and maintaining a harmonious balance with others, in a search for satisfaction of needs ideally motivated by altruistic goals rather than egotistical ones (e.g., Ching, 1977, pp. 71–79).</td>
</tr>
</tbody>
</table>
4. Basic character of change

<table>
<thead>
<tr>
<th>JC-4</th>
<th>OD-4</th>
<th>C-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change is “linear-individual,” with the ideal being movement from a postulated “original sin” to progressively higher levels of perfection as defined by transcendent values, with the end-states beyond human alteration although differentially attainable.</td>
<td>Change is “linear-progressive” (Marshak, 1994, pp. 402–407); i.e., it intends movement from one state to another more desired by persons; is destination-oriented; often deliberately alters the equilibrium of the status quo; people plan and manage it to achieve their own goals; and change is difficult because much of life is in equilibrium.</td>
<td>Change is “cyclical-processual” (Marshak, 1994, pp. 402–407); i.e., the constant ebb and flow of the universe moves in an orderly cycle; is journey-oriented in emphasizing how well one follows the Way; deliberately seeks to restore/maintain a harmonious equilibrium; it involves observing and following the Way to achieve it; and change is usual because everything is in dynamic flux except the Way.</td>
</tr>
</tbody>
</table>

5. Character of work

<table>
<thead>
<tr>
<th>JC-5</th>
<th>OD-5</th>
<th>C-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work must be psychologically acceptable to the individual—that is, its performance should not generally threaten the individual who is, after all, in possession of a working definition of the desired/desirable that can be surfaced by the processes implied in JC.</td>
<td>Work must be psychologically acceptable to the individual, while collective requirements also are met, at least in the long run.</td>
<td>Work must facilitate long-term and harmonious family-like (clan) relationships among individual members, with collective needs being primary and regenerative interaction being dominant.</td>
</tr>
<tr>
<td></td>
<td>I. Judeo-Christian work ethic&lt;sup&gt;a&lt;/sup&gt;</td>
<td>II. OD work ethic&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>6.</td>
<td>Orientation to personal development</td>
<td>Work must allow people to develop their faculties.</td>
</tr>
<tr>
<td></td>
<td>JC-6</td>
<td>OD-6</td>
</tr>
<tr>
<td>7.</td>
<td>Orientation to personal autonomy</td>
<td>Tasks at work must allow individuals considerable self-determination.</td>
</tr>
<tr>
<td></td>
<td>JC-7</td>
<td>OD-7</td>
</tr>
<tr>
<td>8.</td>
<td>Orientation to environmental influence</td>
<td>individuals must have the possibility of controlling and influencing, in meaningful ways, the environment within which tasks are performed.</td>
</tr>
<tr>
<td></td>
<td>JC-8</td>
<td>OD-8</td>
</tr>
</tbody>
</table>

<sup>a</sup> Judeo-Christian work ethic
<sup>b</sup> OD work ethic
<sup>c</sup> Confucian work ethic
<table>
<thead>
<tr>
<th>9. Final arbiter at work</th>
<th>OD-9</th>
<th>Many stakeholders will be involved in decision processes, where bargaining often will be dynamic, problematic, and involve collectivities.</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC-9</td>
<td>C-9</td>
<td>The top decision-maker is the final arbiter of behavior at work, but this &quot;rule of man&quot; presupposes intelligence, high ethical integrity, and correct use of power (Jacobs &amp; Guopei, 1995) in a relationship governed by eui — i.e., justice and morality.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. Roles for intervention in work</th>
<th>OD-10</th>
<th>Intervention has a high priority, as in managerial roles that containing strong elements of change-agentry, and perhaps especially as part of everyone's role in a regenerative and self-correcting system.</th>
</tr>
</thead>
<tbody>
<tr>
<td>JC-10</td>
<td>C-10</td>
<td>If conflict or imbalance exists, if failure occurs in attaining/maintaining a harmonious balance, the ideal is that a &quot;noble&quot; or &quot;sage&quot; will help restore and maintain a harmonious equilibrium (see C-9).</td>
</tr>
</tbody>
</table>

---

a From Golembiewski, 1995b, p. 64.

b From numerous sources, especially Golembiewski, 1993.

c Developed with the help of two Korean students in Georgia's Doctor of Public Administration program — Seok-Eun Kim and Jong-In Yoon. See also Ching (1977), Dollinger (1988), and Marshak (1994).
Noteworthy differences between ethics also characterize item 4 in Table 9.1—Basic Character of Change—with differences between OD-4 and C-4 being especially interesting. At the extremes, the descriptions imply major potential for poor fit of OD values/approaches in Confucian cultures. At the same time, these two ethics share substantial conceptual space. For example, C-4 encompasses movement from one developmental state to another in sequences; and OD-4 often will be associated with various forms of cyclicality—for example, as phases of group development succeed one another, or as generations of members succeed one another in various social collectivities.

This attention to differences between the ethics is usefully extended. “Social distance” is often emphasized among the differences between the ethics, with the Confucian ethic encouraging substantial emphasis such as that associated with “benevolent authoritarianism in leadership” (e.g., Oh, 1991), or with the primary Confucian respect for elders (e.g., Ching, 1977, pp. 96–97). These features suggest a poor fit with the OD ethic, perhaps nowhere more clearly than in OD-8 through OD-10. Here, in principle, some social inventions may be required to adapt OD approaches to Confucian contexts, as in the critical particulars of empathy as well as in the sharing of data via feedback and disclosure. Such OD exchanges in Western contexts are often central, if not critical, and their dynamics can be exuberant (e.g., Golembiewski, 1979, Vol. 1; and Boss, 1979).

How can we deal with such differences, broadly social in character? In general, there seems no need to violate OD to accommodate to Confucian values. To provide illustrative detail, it does not seem advisable in such comparisons of ethics to rely exclusively on extreme interpretations concerning “social distance,” and for at least two classes of reasons. Paramountly, the results of personal disclosure and feedback may be important quite broadly, and perhaps nowhere more so than in settings that discourage both. This is suggested by the extraordinary levels of burnout in cultures that sanction depersonalization, or the distancing of self from others (e.g., Golembiewski, Boudreau, and Munzenrider, 1996, esp. p. 168-175).

Relatedly, the significance of social distance differences is implied by the convenient alternative ways that exist in which disclosure and feedback can be delivered so as to enhance cultural fit, or to threaten it less, no matter what the ethic. For example, Ha (1986) emphasizes that in Korean settings feedback and disclosure can be as salient as in the most “with it” Western setting, but delivery in the former might well emphasize modifications in how empathic materials are fed-back or disclosed to others. Rather than direct, person-to-person feedback in a program for training local Korean administrators, Ha observes that each participant composed long letters to each other participant so as to convey materials whose public sharing might have been generally uncomfortable to fellow Koreans, but which materials were nonetheless critical to a person intent on developing
the authentic sense of self required by (for example) Confucian-1 through Confucian-3 as well as by the OD ethic, overall.

At the very least, then, enhanced empathy and feedback are of relevance in all three ethics—see especially items 2 and 3 in Table 9.1—but applications of the OD ethic often can avoid poorer fits by adaptive resourcefulness. Presumably, for example, reticence to openness in Confucian contexts would be greatest when organizational superiors or elders are involved. Hence, where the C work ethic dominates, intact work groups may be less featured, or survey/feedback interventions may be of special value because of the relative anonymity they can permit while giving feedback. Arguably, also, the high individualistic quotient in the JC ethic also seems a less-than-perfect fit with the OD work ethic, on average. As with the Korean example above, however, the success rates detailed below nonetheless suggest an important compound conclusion: that feedback and disclosure are broadly salient, and that the OD ethic usually can be accommodated to a broad range of cultural conditions.

How to facilitate such adaptations? Perhaps paramountly, and especially related to the issue of social distance, note that OD not only encompasses interaction-centered designs but also includes designs related to organization structure as well as policy and procedures. Reliance on the two latter classes of designs—in which superiors and elders often can be the champions for change, and not necessarily the primary targets in that change—can moderate awkward social distance effects. In addition, no doubt of greatest relevance is OD’s emphasis on diagnosis of specific worksites, along with the involvement of the client in defining of desired/desirable conditions as well as in developing local ways to approach them. Such features facilitate “good enough fit” of the OD and C ethics and, on definite balance, probably often will reduce resistance to change in Confucian contexts, for example. Flexibility in intervention has a clear priority.

Overall, then, the differences between the three ethics can be real enough in principle, but this need not be overwhelming in practice, or even consequential. Each ethic has definite central tendencies, to put the same point another way, but each also tends to shade into the spaces of the other ethics in numerous particulars. This common ground encourages convenient ways to increase the goodness-of-fit of OD design to specific sites.

In addition, also note that each ethic seems to admit some slack in its acceptance: Thus, your author observes that the JC ethic has an individualistic predisposition in principle, but in practice approaches to it might admit (or even require) group-centered features. For example, the J-C work ethic may be associated with more-or-less individualistic approaches as well as with more communitarian predispositions. Both cases could require adaptation and innovation on the basic OD model, typically in an action-research mode that feature client involvement in diagnosis, prescription application, and appraisal with possible fine-tuning.
Selected Similarities Between Ethics

To consider the second generalization suggested by Table 9.1: the sketches there also reveal substantial conceptual overlaps as well as differences between all three of the work ethics. Specifically, six aspects of this variable overlap get illustrative attention here to suggest these commonalities/similarities. First, all three ethics variously reflect an underlying moral order with similar features. This is most obvious in the JC ethic, on the face of it, but the same characterization can be applied to the other two ethics in various ways and at different levels of analysis. For example, the “divine spark” attributed to people by JC-3 can be seen as consistent with the humanist needs models underlying conventional OD (e.g., Argyris, 1957). Moreover, in Confucian-9, the “rule of man” exists there in principle, but the fuller meaning seems rooted in deeper and value-loaded concepts such as justice and morality. Even more clearly, the Confucian ethic differentiates humans from other animals because the former have a “sensitive heart” with four components:

- A sense of compassion
- A heart of shame
- Courtesy and modesty
- A sense of right and wrong

The first three components are clearly and largely humanistic, but the last has obviously transpersonal implications consistent with Pope John Paul II’s sense of universal qualities of humans that extend far beyond the merely conventional. In common, then, all three ethics are clearly value-loaded, and often in compatible if not similar ways.

In addition to such semi-transcendent footings, overall, the practical exigencies of approaching all three ethics often involve a similar “humanistic approach” (e.g., Zikiye and Zikiye, 1993). Thus, both the OD and Confucian ethics have obvious and strong human-centered features, as in the empathy each requires to gain access to the ideas and feelings of self and others that both ethics propose that individuals should take into account in moving toward the good life—for example, in C-6 and C-7, as well as in OD-5, OD-6, and OD-7. Similarly, openness in relationships and personal owning are critical in moving empathic knowledge into action, as “responsible freedom I” requires. Here, the OD and Confucian ethics feature what the JC ethic may not emphasize, but that is nonetheless useful or even necessary in much of the life to which the JC ethic is applied. The broad sense of the point is implied by the notion of “levels of perfection” in the JC tradition, which preserves the sense of a transcendental end-value while realizing that humans often will attain or even aspire only to some asymptotic approximations of that end-value, perhaps even far removed from the ideal. The reader may remember the young man who expressed an interest to Christ of being
like Him. Clear enough, the young man was told: “Give up all you own, and come follow me.” But that was more than the young man had bargained for, and the biblical Christ was willing to prescribe more suitable ways-and-means for the lower (but still acceptable) level of perfection that the young man really targeted, at least for the then-present.

Relatedly, of course, the JC ethic rests on a clear transcendent foundation, but even in its essentials one can see prominent humanistic emphases. See JC-2 and JC-3, especially, which highlight the roles of coming to knowledge-of-self in community, and acting on that knowledge in collective contexts. Put another way, the divine spirit is free in the JC ethic to intervene dramatically in nature but, for the far larger part of their choice-making, people will see truth only through the often-dark lens of their own self-awareness, as tested against that of others. Much the same seems to hold in the OD and Confucian ethics, neither of which accepts a transcendent view of desires/desirables.

Second, several ethics seem to share basics even as their sketches in Table 9.1 suggest some differences in degree. Thus, JC-6 through JC-10 clearly reflect a definite theme of individualism, but group-oriented OD designs also have a clear relevance in highlighting what is both often necessary in practice but also can go against the JC grain. As for OD and Confucian ethics, they both give priority to collectivist themes, which encourage such group effects as “teamwork” but also may awkwardly obscure individual claims versus those of collectivities like clans or families (e.g., C-1, C-5, C-6, and C-7). Here, also, OD approaches and values may be relevant, as adapted to different cultural contexts by providing a counterweight to help balance contrasting central tendencies in the two other work ethics. Here, that is to say, OD approaches provide a map to help attain what the Confucian ethic prescribes but always needs to practice better; and OD also can help attain what the JC ethic does not emphasize but which nonetheless will be useful, or even necessary at times, but which does not give the highest priority to individual growth and even survival.

Third, all three ethics have a definite integrative orientation to work, as contrasted with the fragmentation of work associated with the bureaucratic model. This bias is at OD’s heart, to begin, and also is especially clear in JC-5 through JC-8, as other sources develop in detail (Golembiewski, 1965, 1985, 1989a, 1995b). C-3 through C-5 also trend in a similar direction. Consistently, Oh (1991) traces the reputation for teamwork attributed to “East Asians” directly to the Confucian work ethic; Dollinger (1988) attributes a heightened sense of collective responsibility to the same source. The full case for the Confucian ethic as it gets reflected in detailed work designs has not yet been made, but Ouchi (1980, 1984) seems instructive on the key point of associating the Confucian ethic with integrative versus fragmenting concepts of work as well as of organization structures. And no doubt, “the family” in the Confucian ethic provides powerful integrative forces as well as a model for other and larger collective forms.
Fourth, and perhaps of paramount relevance, whatever the dominant ethic or combination of them at a particular worksite, it seems easily arguable that common forces like those associated with globalization are acting with some uniformity on all systems. If so, this implies a significant twist on the “good fit” argument. To illustrate, the OD ethic may be said to provide a useful normative template related to where many or most organizations are now heading, given their different starting points. In contrast, the usual “good fit” argument focuses on the conditions of origin. In this fourth commonality, in contrast, the focus is on the condition of destination and its increasingly congruent fit with (for example) the OD ethic, which was constructed with fast-paced learning and integration uppermost in mind. In fact, the OD ethic (for example) may be more desirable when paired with an opposed ethic that has a poor fit with emerging environmental demands.

Fifth, such a twist may apply with special force to what seem noteworthy differences between the several ethics. Hence, JC versus Confucian ethics suggest different comfort zones concerning the direct confrontation of superiors, with the Confucian implying the greatest social distance; and OD admits, even idealizes, a range of confrontations far beyond the other ethics. See items 3, 4, and 5 in Table 9.1 for JC and Confucian ethics, and then compare to OD-3 through OD-5. Relatedly, the balance between individualistic versus collective pressures differs most between JC and Confucian ethics and seems more similar between OD and Confucian ethics. Differences also exist about the ideal degree of quasi-stationary equilibrium (QSE). Emphasis on QSE seems greatest in Confucian ethics, but the notion definitely is not absent in the two other ethics, and is certainly not abhorrent to them.

Some observers also may see the three different concepts of change—items 4 and 10 in Table 9.1—as posing the greatest challenges to a good fit between OD versus Confucian ethics. On closer examination, however, the dominant emphasis on equilibrium in the Confucian and perhaps OD ethics is complemented by a backup strategy for change. Similarly, JC may emphasize “linear-progressive” change as an ideal, but this clearly admits both the theoretical and practical possibility of “punctuated equilibria”—i.e., substantial periods of relative stability between “big bang” before attempting further progress, which may not differ much in the long run from practice in (for example) the Confucian ethic.

Perhaps most directly, the reliance on “action research” designs in OD provides useful flexibility in all cases of fine- or gross-tuning, whatever ethic dominates. That is, both the start-up condition as well as the terminal desired/desirable position often derive from dialog among the targets of change, or at least are influenced by that dialog. In the OD idiom, the “here” for start-up can have very different empirical and normative features. As for the “there” highlighted by OD ideals, the only requirement is that movement occurs from “here” to more closely approach the OD “there.” The “here” will seldom if ever become the “there”—
not in OD ideation, let alone in practice. Hence, pressure to disturb an equilibrium often will be felt—as people change their views about what is desired/desirable, or as environments change, among many other stimuli.

Sixth, although this is clearly a matter of opinion, if not conjecture, the differences between the three ethics do not seem so great as to preclude building on their similarities. For example, the Confucian ethic in recent years has been associated with economic and personal developments or has even inspired them. Earlier, most observers saw the C Ethic as a force restraining change. One essential part of these bridgeable differences derives from the value placed on equilibrium somewhere near the present condition, which some observers associate with the Confucian ethic and see as an anti-change bias. However, all three ethics also can encourage the pursuit of distant ideal equilibria, if the present conditions are seen as unattractive enough, or if the “there” takes on a great attraction.

SUCCESS RATES, EXTENDED: GLOBAL AND SOUTH KOREAN WORKSITES

If the ideation above is reasonably correct, one condition should generally exist. OD approaches should “work” in many situations. That crucial point gets tested here. In sum, success rate estimates will be significant in assessing the mutual ties in practice of the three ethics, as suggested above. Earlier, the first chapter in this trinity as well as Chapter 1 demonstrate the substantial success rates of OD and QWL applications in Western settings where, one presumes, the JC ethic should be prominent. We still need to test for OD success rates in “developing” contexts and especially in Korea, which is arguably the “most Confucian” nation in all of East Asia. The conventional view (of Max Weber, for example) has been that the Confucian ethic inhibits economic development specifically—and all change, in general. But recent developments question this facile view and may undercut it. We shall see.

Sparse Methodological Notes

The several summaries below come from multiple panels of evaluations of OD applications that use similar measurement conventions and present details related to three generalizations. Thus, searches for applications were mostly directed toward published studies, but with some success at accumulating unpublished agency and consultant reports; and the searches initially focused on English-language evaluations, but now encompass a growing range of languages. Typically, “success” was rated by several observers—often three individuals. Finally, for present purposes, conservative conventions dominate. Thus, rater reliability is 100 percent in all cases. That is, in evaluating the OD applications summarized below, all raters either agreed to the same success rating or, when disagreement
<table>
<thead>
<tr>
<th>Categories of rated success</th>
<th>A. Western countries, largely U.S. (%) (N=574)</th>
<th>B. Developing countries (%) (N = 240)</th>
<th>C. South Korean(^{\dagger}) applications (%) (N = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public sector</td>
<td>Business Sector</td>
<td></td>
</tr>
<tr>
<td>I. Highly positive and intended effect</td>
<td>41</td>
<td>40</td>
<td>26</td>
</tr>
<tr>
<td>II. Definite balance of positive and intended effects</td>
<td>43</td>
<td>49</td>
<td>48</td>
</tr>
<tr>
<td>III. No appreciable effects, or balance of intended and contrary effects</td>
<td>7</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>IV. Contrary effects, and especially even those applications with a small proportion of estimates in a contrary direction that attain statistical significance</td>
<td>10</td>
<td>6</td>
<td>5.4</td>
</tr>
<tr>
<td>Results not available</td>
<td>0</td>
<td>0</td>
<td>4.2</td>
</tr>
</tbody>
</table>

\(^{\dagger}\text{Yoon, 2000.}\)
TABLE 9.3 Verbal Descriptions of Success Rate Categories

<table>
<thead>
<tr>
<th>Success rate Categories</th>
<th>Verbal descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Most pre vs. post estimates of intended effects are statistically significant or “large” (&gt;10 percent), as well as in the expected direction, with the minority of cases in a contrary direction being largely or entirely random</td>
</tr>
<tr>
<td>II</td>
<td>Many changes are large or statistically significant; differences can fall in a contrary direction; although as many as a quarter or so fall in a contrary direction, few or none attain statistical significance or are “large”</td>
</tr>
<tr>
<td>III</td>
<td>No appreciable effects means few or no “large” differences.</td>
</tr>
<tr>
<td>IV</td>
<td>Many effects fall in a contrary direction; also, many attain significance or are “large.”</td>
</tr>
</tbody>
</table>

persisted after discussion, the lowest rating in contention was assigned to the case.

Success Rates in Four Loci

Table 9.2 summarizes the analysis of three populations of OD applications in a number of panels of evaluative studies. Although, some double-counting is involved, reinforcing considerations suggest the estimates are realistic. Broadly, these four success categories (Table 9.3) usually provide summary estimates of effects.

OD SUCCESS RATES IN MIXED SETTINGS

This book at several earlier points reports on the success rates in Western settings, and overall findings indicate “success” by anyone’s standards, in both business and public sectors. Two non-Western settings add to this material on success rates. Table 9.1 provides a convenient setting.

OD Success Rates in “Developing Countries”

Although the common wisdom at one time was myopic—expecting few OD applications in non-Western settings, and presuming poor success in these rare applications—that caricature has become increasingly untenable. Specifically, an initial batch of 100 applications in “developing” countries generated substantial success rates (Golembiewski, 1991) and, relying on similar measures of success,
two subsequent addenda to that base now have accumulated 240 cases in 56 countries (Golembiewski and Luo, 1994; Golembiewski, 2000a). See Table 9.1B.

“Developing countries” include those large political jurisdictions—countries or nation-states—which had Gross National Product per capita in 1980 of $5000US or less. Kurian (1984, pp. 98–99) lists 38 countries with 1980 GNP per capita of $5000 or more, and 133 with less than $5000 per capita. The countries Kurian numbered 39–43, to ground the cutting-point in Table 9.1 for the reader, include, in order, Ireland, Martinique, the old Soviet Union, Greece, and Israel.

Table 9.1B summarizes the success rates assigned to cases in the 56 “developing” countries in which searches through 1999 yielded 240 OD applications. Clearly, the estimates are less attractive than for Western settings but nonetheless remain substantial. In addition, the latest applications seem to generate ascending curves—in numbers of applications, countries of application, and attractiveness of success rates.

All these features imply that—whatever the unspecified differences in organizational or social cultures—reasonably effective adaptations of OD values, approaches, and techniques can be made to them. Presumably, this implies that OD applications can be characterized as “broad spectrum” in that the 56 countries aggregated in column B of Table 9.2 encompass JC and Confucian work ethics, and more besides. Finally, even acknowledging measurement issues, the estimates in column B seem far beyond those reasonably assignable to chance, or to the mere passage of time.

Note, again, virtually all the cases summarized in columns A and B of Table 9.2 appear in the English language, whatever their locus of application. Hence, the cases observed do not necessarily measure the total level of OD activity within a jurisdiction. Note also that the two most frequent loci for OD applications in column B are otherwise very different—India (70 cases) and Israel (14).

**OD Success Rates in Korea**

Finally, investigators whose first language is Korean recently have began to add to the batch of OD applications in what are here labeled “developing countries.” See column C in Table 9.2 for a summary of results of this East Asian effort (Yoon, 1999). Early search of Korean sources has generated twenty cases, and the worksites included no doubt reflect a dominant influence of the Confucian work ethic.

Two generalizations characterize the Korean batch summarized in column C and detailed elsewhere (Yoon, 1999). Thus, despite the small population, the provisional success rate seems at least comparable to other assays. In addition, the closer one looks, the more cases are isolated. The search for OD cases reported in the Korean language has only fairly begun—investigators have not yet system-
atatively contacted Korean-speaking consultants working in South Korea. Nonetheless “old Korea hands” are surprised by the recently available volume of evidence of OD activity.

DISCUSSION

So, the sketch of a major argument has been completed for present purposes, although much remains undone. To facilitate this follow-on work, let us remind the reader of the flow of this argument. In sum:

Organization Development and its associated work ethic provide major evidence of becoming a useful technology-cum-values—certainly in Western settings and, if with less but nonetheless considerable confidence, also in “developing” countries; and

A review of three ethics relevant to work implies that the OD ethic has a good fit with work settings in which the Judeo-Christian ethic and the Confucian ethic seem prominent. As a major value-added, this review also suggests a tolerable fit between the OD ethic and other work contexts, with special attention to Korea, where the Confucian ethic is highly relevant, as well as to certain “developing nations.”

Overall, then, the several assays above suggest a “good enough fit” of OD, on balance, to a broad range of work ethics as well as cultures, whether in organizations or within nation states. However, several qualifications should tether interpretations. The basic generalization is not that East/West or other differences can be disregarded in OD applications. Rather, the following five nuanced generalizations seem apt.

First, one or more OD designs, but hardly all, seem applicable under all/most worksite conditions, wherever. In part, this reflects the coverage of OD designs—in applications that emphasize interaction, policies or procedures, or structures, singly or in various combinations. Thus, interaction-centered designs might be more acceptable in “Western” settings, whereas structural change might be more favored by legitimate authorities in “traditional” settings.

Second, on definite balance, OD interveners have the requisite skills and knowledge to diagnose the conditions operative in a broad range of organizations, in Western as well as other countries or cultures. To put it otherwise, the high proportions of “successful” impacts reported above could hardly have occurred by chance, where “poor fit” existed, or where fine-tuning capabilities were weak.

Third, the goals or values underlying OD seem widely desired, if not necessarily to the same degrees, in a broad range of settings. This generalization encompasses several meanings. Typically, OD applications can begin from many initial normative positions—that is, an OD intervention may begin from a “here” that may be distant from the “there” of OD ideals, as long as it constitutes a real
step toward them. Perhaps more significantly, OD prominently features “action research” or “participatory research” (e.g., Brown and Tandon, 1983). In the many variants of these related approaches, participants at the site of application prescribe the desired or desirable states, movement toward which is facilitated by OD approaches and techniques. In sum, OD methods and approaches essentially encourage “good fit” to sites of application by emphasizing the inputs of clients at the worksite, not only in diagnosis but also in prescription and evaluation (e.g., Golembiewski, 1993).

Fourth, for many reasons, organizations or social systems often will differ in many regards, but OD accommodations to them seem possible in a substantial proportion of cases. This seems a reasonable interpretation of the fact that (for example) nearly 60 nation-states—and we are still counting!—have hosted the OD applications summarized in Table 9.1.

Fifth, the JC and Confucian work ethics are variously relevant to OD applications. Thus, isolating work contexts influenced by these or similar work ethics can help choose among possible OD designs as especially relevant or as contraindicated. And these work ethics can help in the gross and subtle adjustments always necessary to fine-tune applications based on the OD work ethic. The latter is clearly differentiated from (for example) a bureaucratic work ethic but nonetheless will profit from fine-tuning that adjusts to the specific properties characteristic of each particular worksite (e.g., Golembiewski, 1995b).

REFERENCES
The Confucian Work Ethic

ach (Ed.), *Research in the sociology of organizations* (pp. 201–263). Greenwich, CT: JAI Press.


Not Every Learning Design Works Every Time
Toward an “Optimum Discrepancy” to Better Target Interventions

The present point is an elemental one, but it still has not attracted enough attention. It is this: No one should expect that any learning design will work everywhere. This expectation may take many forms, all of which are related to ways of becoming more efficient in designing change. For example, knowing what design will not work will either inhibit applications, or it will motivate the search for the appropriate applications.

How, then, to determine which design works best under specific conditions? One basic approach relates to this old chestnut in social psychological research, as in the theory of cognitive dissonance (e.g., Festinger, 1957). If a discrepancy between a stimulus anchor has motivating potential, what degree of discrepancy is most conducive to the largest degree of desired change?

OD has achieved a substantial degree of success without agreement, or even concern, about such a model of ideal discrepancy. Said otherwise, to increase productive and applied usefulness, work in OD must become increasingly specific about what effects can be expected under which circumstances. To be sure, general agreement exists that learning typically occurs in response to some hitch or incongruence between what we expect or believe, as contrasted with what we come to know or experience. This can be viewed as a gap between “ideal” and “actual.” And reasonable OD advice proposes intervening where trauma exist,
or where hurt is present. The trauma or hurt, in effect, generate energies that may be used to power choice or change—that is to say, learning.

But what is the optimum hitch or incongruence—small, medium, or large?

Despite the substantial OD success rates sketched in early chapters, we have a long way to go in OD being usefully specific about the kind of ideal/actual discrepancy that best induces learning, choice, or change. Assume we wish to effectively bring to someone’s attention some piece of feedback. What is the recommended strength of the dosage? Should the feedback, consistent with essential honesty, be mild, tough, or somewhere in-between? Having a reasonable orientation to this question should add much to OD efficiency and effectiveness.

Several varieties of the common wisdom propose rules of thumb concerning such specificity, but little relevant OD research exists. Hence this chapter contributes to one aspect of the required specificity, and this by dealing with one issue—the different degrees of motivational force that can be attributed to varying degrees of discrepancy between an individual’s preferences about desirable organizational climate and perceptions of the actual situation. In sum, how much discrepancy is too much? It is bad news that what follows has been published before (Golembiewski, 1990) and still remains rare or even unique.

A CONCEPTUAL CONTEXT FOR IRONY III

The design in OD of appropriate discrepancies patently must deal with the properties of individual learners as well as with the characteristics of their organizations. Major early signs of concern about the point (Bennis, 1969, Harrison, 1965) were expressed, but largely unexplored issues still loom large, even in the face of recent confirming evidence that the early concerns are right-on (e.g., Bowers, 1973; Bowers and Hausser, 1977; Pasmor and King, 1978). For example, central issues may be phrased in the form of a general question of relevance here:

What degree of discrepancy between design properties and existing conditions is too much, or too little?

We are not very far down the trail toward having detailed answers to specific versions of this question. Opinions about such key questions abound and tend to be strongly held. Not only does little research address these questions, however, but the research that does address them seems to be methodologically vulnerable. Witness the criticisms of Bower’s major effort to differentiate interventions appropriate for different hosts (e.g., Torbert 1973; Pasmor 1976). And my own narrower work on a related theme (Golembiewski, 1970) can also be faulted on methodological grounds.

Some progress is possible despite the unknowns about optimum discrepancies, and that possibility motivates this chapter to address aspects of the issue. The data come from a mass team-building experience at one site for personnel
from thirty-three sales regions of a national marketing organization, including all first-line managers and all salesmen reporting to them.

The data here are self-reports about behavior and attitudes, which are interpreted as reflecting changes in organizational climate and, if by inference only, as changes in behaviors due to an OD intervention. Respondents initially described the organizational climate of their immediate work units as typically falling short of their ideal, which was basically consistent with the values underlying OD. The intervention thus had two tasks: to induce attitudes supporting the possibility of reducing the gap between ideal and actual organizational climate, and, especially, to increase the performance of behaviors appropriate to reducing that gap. Impressionistically, the experience was considered successful—by the consultants and by several levels of management, as well as by most other observers. More rigorous efforts to document effects generally support this global and impressionistic evaluation (Golembiewski, 1979a, pp. 132–136), albeit with significant reservations.

**THREE ALTERNATIVE MODELS OF AN “OPTIMUM DISCREPANCY”**

Conceptually, this chapter focuses on the issue of an optimum discrepancy for learning. That is, a discrepancy can exist between how a person behaves and some challenging stimulus. Motivating energy is often attributed to such discrepancies, with change in attitudes or behaviors being a convenient way to reduce the discrepancy and thus to satiate a motivational urge. If I believe I am a bit pudgy, well, that discrepancy might motivate some direct action. Some call this the “dilemma/invention model” of change.

Hypothetically, the simple models in Figure 10.1 sketch three alternative but not exclusive discrepancy/change linkages. The models are analogues of ones that have received much attention, but in other ways. They are used here as a tentative starting point, and the qualifier “tentative” is appropriate for several reasons. For example, the models obviously homogenize the complex mix of behavioral and attitudinal changes in individuals generated by an OD intervention. In addition, the present approach is also too simplistic to extend to Heider’s (1958) balance model, which has a variety of attractive features. These are inelegant features but unavoidable at this early stage of research.

Model I in Figure 10.1 is “accelerating” because it proposes that the greater the discrepancy between the stimulus of an OD intervention and an individual’s normal anchors, the greater will be the tendency to move those anchors toward that stimulus. That is to say, the individual reduces dissonance between self and stimulus by making adjustments in self, in direct proportion to the magnitude of the dissonance. Model I shares a basic conceptual thrust, then, with approaches such as Festinger’s (1957) theory of cognitive dissonance. He proposes that “the
discrepant reality which impinges on a person will exert pressures in the direction of bringing the appropriate cognitive elements into correspondence with that reality... The strength of the pressures to reduce the dissonance is a function of the magnitude of dissonance” (1957, 18).

Model II is “decelerating,” in direct contrast. Essentially, it implies that change in OD occurs more by assimilation than by contrast. High degrees of discrepancy, consequently, will tend to restrict the degree of change, or negate it entirely. Model II is consistent with a variety of experimental research by the Sherifs (1967), who argue that Model I describes attitudinal change only where “ego involvement” is low. There, learning occurs by contrast. Model II implies that learning or change in OD programs — where high ego involvement is intended and often will exist — will tend to occur by assimilation. Hence, low discrepancies will be associated with larger changes, if Model II describes reality.

Model III in Figure 10.1 perhaps dominates in OD theory (e.g., Bennis 1969; Harrison 1965). Model III, in effect, provides an “accelerating” effect for low-to-intermediate discrepancies, as well as a “decelerating” effect for greater discrepancies. To illustrate, in the case of choosing a suitable strength for a piece of significant feedback to an individual two effects are to be avoided: too little may fail to get the message across, and too much may be so threatening to the target that the message is misheard or rejected. For example: the target is an individual who sees self as mentally healthy but who persistently resorts to dra-
matic outpourings of language that describe a fearful fantasy. One could express a concern about this behavior with any of these three statements:

   Gee, you tell a good story.
When you talk so long, and use such rich metaphors, it is hard for me to follow you. And I often don’t know where you are going, even though I want to understand you.
Pretty clearly, you are concerned about schizoid tendencies in yourself.
   You sound sick to me. You have real reason to be fearful.

In the context of Model III, the first alternative may be “too little” to encourage attention to the other’s behavior, and the third alternative formulation is almost always “too much.”

**OD LEARNING AND DISCREPANCIES: INDIVIDUAL PREFERENCES AND ORGANIZATIONAL CLIMATE**

One particular discrepancy gets attention here—an “optimum prior discrepancy”—as does its association with the degree of learning or change attributed to an OD program. This optimum prior discrepancy refers to the most effective degree of differences for stimulating change, with the difference referring to the gap between what an individual says he would like his organization’s climate or style to be, and how he perceives that climate before an OD experience. “Discrepancy” here has both attitudinal and behavioral components, in sum. In effect, the OD intervention in this case provides most participants with an experience that legitimates moving toward their preference as to organization climate, which is distinctly different from the climate they attribute to their real organization, in general. Directly, the OD intervention highlights this prior discrepancy while it also encourages skill practice with behaviors appropriate for more closely approaching the preferred climate.

**Operationally Measuring Discrepancy**

The present approach to measuring prior discrepancy uses respondent self-reports on the Likert (1967) Profile of Organizational Characteristics, form E. For an overview of the profile and its features and uses, see also Likert’s later work (1977). Form E consists of the twenty-four items that measure aspects of an organization’s comprehensive climate or style, rated both as Now, or the way the respondent sees his organization on each item, and Ideal, or the way the respondent would like his organization to be.

Differences between Ideal and Now ratings before the OD intervention provide the operational measure of prior discrepancy. Ideal scores measure per-
sonal preferences about organization climate, in sum, and Now scores measure actual organization conditions as perceived by the respondents.

The twenty-four Likert items are not reproduced here, to conserve space, but they are intended to tap a broad phenomenal range. In capsule, seven basic organizational process are said to be tapped by the profile items:

- Items 1 and 2 deal with leadership.
- Items 3–6 deal with motivation.
- Items 7–13 deal with the character of communication.
- Items 14–17 deal with interaction and influence.
- Items 18–20 deal with decision making.
- Item 21 deals with goal setting.
- Items 22–24 deal with the character of control.

The profile has another useful property. Each item is measured along a 20-point scale of equal-appearing intervals with four major segments, each of which is anchored by a brief descriptive statement. Likert conceives of these four segments as qualitatively different “systems of management,” that is, as distinct managerial climates or styles. The systems are:

- System 1: Exploitative-authoritative (1–5 on the scale)
- System 2: Benevolent-authoritative (6–10 on the scale)
- System 3: Consultative (11–15 on the scale)
- System 4: Participative group (16–20 on the scale)

Systems 1 and 4 are taken to be the anti-goal and goal, respectively, of OD applications. Almost all respondents saw their organization as too far removed from high-System 3 or System 4, toward which they desired to move.

**Massive Team Building as the Goal**

Overall, the OD design in this case intends to enhance movement of the climate or style of thirty-three regional sales units in the direction of Likert’s System 4, the Participative Group System. In effect, the design has three purposes. First, it seeks to legitimate looking at possible discrepancies between individual preferences and organization style. Second, the design intends to make participants aware of actual discrepancies, where consciousness is lacking. Third, the design provides a model for reducing any discrepancies via a learning experience with appropriate values, attitudes, and behavioral skills.

The design lasted two days, and follow-up experiences were possible for individual regions. This undermines a view of the team building as a one-shot effort. The data reported here were gathered before any region had a follow-up experience.

The details of the design have been elaborated elsewhere (Golembiewski...
1979a, pp.132–133), consequently, brief notice here will suffice. By intent, the
design is of modest intensity and basically seeks to develop skills as well as
attitudes consistent with OD values and especially with regenerative interaction.
Practice was also emphasized, with contagion effects being important in a large
ballroom where learners were exposed to high-tech graphics and to one another’s
excitement. A team of six facilitators was available. Their help could be requested
by any regional assemblage that raised a flag at their table. The design’s motif
was: Tell it like it is in the effort to move toward regenerative and away from
the degenerative. This won no style points, but contributed to the sense of an
elevated and up-tempo mood.

Each of the design’s five major components takes a common approach to
increasing the quantity and improving the quality of job-relevant feedback avail-
able to members of each region. Thus, brief cognitive inputs in large general
sessions attempt to help participants organize their experiences as team members.
The dual purposes are to aid participants in seeing themselves and others more
clearly in job-relevant activities, and to begin building more effective and satis-
fying relations at work.

After each conceptual input, commonly, individual regions spent an hour
or two applying the substance to their relationships at work. For example, task
and maintenance roles were introduced, and participants then were given lists of
behaviors consistent with such roles and asked to identify specific team members
whom they saw as having performed one or more of the behaviors attributed to
each role type. These perceptions were shared, providing not only some experi-
ence with openness and cross-checking of perceptions, but also often indicating
a deficit. Action planning appropriate for each region followed, as a scheduled
event.

Analysis of the Results of Team Building

In effect this research design can be summarized as: Observation 1 → OD inter-
vention → Observation 2. To explain, Observation 1 is a pre-test with the Likert
Profile, which preceded the OD intervention by a month, and Observation 2 is
a post-test that was administered approximately three months after the interven-
tion. The four-month interval between the two administrations was a guesstimate
about the length of time required for any attitudinal or behavioral changes to be
tested in the regional units, whose members came together only episodically.
This sought to focus on behavioral changes, as opposed to the good intentions
that might well directly follow the OD experience but also might quickly evapo-
rate back in the real world. Both administrations were by mail, sent from the
firm but returned to the author’s university address. For convenience, data from
the first administration are identified as “Now 1” and “Ideal 1,” and data from
the second administration carry the labels “Now 2” and “Ideal 2.”
Response Rates

The response rate for salespersons falls somewhere between 75 and 85 percent, depending on the denominator one chooses. At Day 1, that is, respondents numbered 430. Of these, 406 responded to the first Likert administration, and 341 to the second one some four months later. The sample here is the 329 salespersons who responded both times, or some 76 percent of the original 430. By Day 120, due to turnover, approximately 345 respondents remained. So the response rate could also be calculated at somewhat over 83 percent. In either case, no major sampling errors seem probable. Normal attrition accounts for about one-third of the differences between the numbers of respondents to the administrations.

Strategy for Organizing Data

Two major points suggest the range of alternatives for organizing the data, and they also sketch a rationale for the eventual comparisons. These comparisons involve individual respondents, the four Likert systems, and individual Likert items.

First, the level of analysis here is the individual respondent. In effect, this opts against analysis at two other levels: by the thirty-three regions, or by the individuals within regions.

Several factors motivate this compound choice. Paramountly, the issue of optimum discrepancies for change by individuals has been a major and continuing focus for research and speculation, as in the cognitive dissonance literature, and this choice avoids well-known measurement and conceptual problems at the regional level of analysis. For example, convenient operations for estimating regional climate—much as averaging the self-reports of all members of a region— imply severe interpretive difficulties. If self-reports are equally weighted, as is usual, that may undervalue the reports of opinion leaders, and so on. Finally, analysis at the individual level avoids a variety of troublesome statistical issues with unequal sample sizes and differential response rates in the several regions.

Second, the analysis will focus on individual Likert items for each of the 329 respondents. In effect, this opts against two other possible ways of treating responses to the profile: using a total score based on each individual’s responses in terms of the seven dimensions into which Likert subdivides the profile items.

The decision to focus on individual items reflects some difficult judgments. Thus, a principal components analysis strongly implies that it is reasonable to view the Likert Profile as having seven dimensions. Specifically, the first principal components of Now 1 and Now 2 scores account for 41.1 and 58.0 percent of the variance, respectively. See Table 10.1 for details. If anything, given that the lowest item loading on the two first principal components for Now 1 and Now 2 scores is a substantial 0.53, Table 10.1 provides more support for calculating one total Likert score than for distinguishing scores on seven dimensions.
### Table 10.1  Principal Component Analysis of Likert Scores, 0₁ and 0₂

<table>
<thead>
<tr>
<th>Likert Item</th>
<th>Likert Now 1 scores</th>
<th>Likert Now 2 scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.73</td>
<td>-.34</td>
</tr>
<tr>
<td>2</td>
<td>.66</td>
<td>-.45</td>
</tr>
<tr>
<td>3</td>
<td>.63</td>
<td>-.14</td>
</tr>
<tr>
<td>4</td>
<td>.64</td>
<td>.01</td>
</tr>
<tr>
<td>5</td>
<td>.57</td>
<td>.27</td>
</tr>
<tr>
<td>6</td>
<td>.66</td>
<td>-.13</td>
</tr>
<tr>
<td>7</td>
<td>.73</td>
<td>-.09</td>
</tr>
<tr>
<td>8</td>
<td>.69</td>
<td>-.22</td>
</tr>
<tr>
<td>9</td>
<td>.65</td>
<td>-.08</td>
</tr>
<tr>
<td>10</td>
<td>.65</td>
<td>-.01</td>
</tr>
<tr>
<td>11</td>
<td>.53</td>
<td>-.14</td>
</tr>
<tr>
<td>12</td>
<td>.72</td>
<td>-.40</td>
</tr>
<tr>
<td>13</td>
<td>.72</td>
<td>-.22</td>
</tr>
<tr>
<td>14</td>
<td>.72</td>
<td>.11</td>
</tr>
<tr>
<td>15</td>
<td>.68</td>
<td>.18</td>
</tr>
<tr>
<td>16</td>
<td>.66</td>
<td>.04</td>
</tr>
<tr>
<td>17</td>
<td>.53</td>
<td>.21</td>
</tr>
<tr>
<td>18</td>
<td>.66</td>
<td>.07</td>
</tr>
<tr>
<td>19</td>
<td>.73</td>
<td>.13</td>
</tr>
<tr>
<td>20</td>
<td>.60</td>
<td>.09</td>
</tr>
<tr>
<td>21</td>
<td>.58</td>
<td>.33</td>
</tr>
<tr>
<td>22</td>
<td>.58</td>
<td>.51</td>
</tr>
<tr>
<td>23</td>
<td>.53</td>
<td>.53</td>
</tr>
<tr>
<td>24</td>
<td>.64</td>
<td>.10</td>
</tr>
</tbody>
</table>

Loadings on principal components

Eigen root 10.10
Eigen root variance 41.10%
Percent variance 48.33%
Cumulative variance 53.15%
Support for such an approach also derives from the fact that aggregating items thought to tap a common phenomenal realm can increase reliability via the randomization of unique variance and the pooling of common variance.

In sum, the comparisons below focuses on three classes of targets: 329 respondents, four Likert systems, and twenty-four Likert items. If those comparisons yield a consistent and statistically significant pattern, that outcome is consistent with a dominant covariation of individual discrepancies and changes in self-reports about organization climate. Readers should be cautious in interpreting Tables 10.2 and 10.3, however, because they clearly do not refer to tests of twenty-four separate phenomenal domains.

Trends in Data

Table 10.2 provides some preliminary data about individual respondents and their self-reports, with this summary supporting three generalizations. First, 239 of the 329 respondents rate their region’s Now 1 climate as System 1, 2, or 3, which

<table>
<thead>
<tr>
<th>Table 10.2</th>
<th>Some Summary Data from Member Self—Reports About Climate of Regional Units as Measured by Likert Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial Likert System of management by individual respondent</td>
</tr>
<tr>
<td></td>
<td>System 1</td>
</tr>
<tr>
<td>(1) Average number of respondents rating their region on Now 1 as</td>
<td>12</td>
</tr>
<tr>
<td>(2) Mean Now 1 scores</td>
<td>3.1</td>
</tr>
<tr>
<td>(3) Mean Ideal I scores</td>
<td>16.5</td>
</tr>
<tr>
<td>(4) Mean discrepancy scores, or Ideal 1–Now 1</td>
<td>13.4</td>
</tr>
<tr>
<td>(5) Mean Now 2 scores</td>
<td>9.8</td>
</tr>
<tr>
<td>(6) Mean Now 2–Now 1 scores</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Notes: Overall F-values for rows 2–6 were statistically significant in all cases.

As is conventional when overall F-values are significant, each paired comparison in rows 2–6 was tested for statistical significance using Duncan’s Multiple-Range test. The test permits comparisons of each of the six possible pairs of the four Likert Systems on each of the twenty-four items, or 144 comparisons in all for each row.

* All 144 possible pairs of differences surpass the .01 level.
+ Seventy of the 144 possible pairs of differences do not attain the .05 level. Fifty-one of the 72 comparisons involving only Systems 2, 3, and 4 do not reach the .05 level.
@ All 144 possible pairs of differences surpass the .05 level.
† 126 of the 144 possible pairs of differences surpass the .05 level.
‡ 132 of the 144 possible pairs of differences surpass the .05 level.
implies that most individuals see their organization unit as falling short of OD ideals. The differences between the four clusters of Now 1 scores attain extremely high levels of statistical significance.

Second, Ideal 1 scores in Table 10.2 indicate a very strong preference for mid-System 4, which is to say that respondents attitudinally prefer an organization climate consistent with the normative thrust of OD. Essentially, then, the challenge to the OD program was dual: convincing participants that they could safely behave in ways that approach their preferences as to organization climate, and providing them with sufficient skill-practice dealing with appropriate attitudes and behaviors.

Third, respondents who initially rate their regions in System 1 have the highest prior-discrepancy scores by far, whereas System 4 respondents approximate a mean discrepancy of half a point. Row 4 in Table 10.2 suggests this regularity (which column 5 in Table 10.2 establishes more directly). Tests of the differences between the means of the discrepancy scores, classified in terms of the four Likert systems, achieve huge F-levels.

The Ideal 1–Now 1 discrepancies in Table 10.2 (i.e., row 4) provide the take-off point for the central concern of this chapter. Specifically, which model of change in Figure 10.1 best accounts for the actual pattern of movement in Now 2 scores?

The accelerating model of discrepancy/change gets strong support from the data—in fact, it gets almost unqualified support. (See rows 4 and 6 in Table 10.2, for general support.) As the mean discrepancy score decreases, in sum, so do respondents report less change in pre- versus post-intervention Likert self-reports. For the lowest prior-discrepancy scores, in fact, the mean change has a negative value.

Table 10.3 provides far more detail on the applicability of the accelerating model to the present data, via a breakdown by each of the twenty-four Likert items, and this approach establishes a pattern of massive consistency. Table 10.3 also reflects more sophisticated data treatment, an analysis of variance which—among other virtues—permits the use of covariance techniques to control for differences in the initial levels of prior discrepancies reported by individuals. That is, the two measures of post-intervention self-reports—Now 2 and Now 2–Now 1 scores—are adjusted for differences in the levels of Ideal 1–Now 1 discrepancies. The control is a vital procedure, assuring as it does that any post-intervention differences are not simply an artifact reflecting differences between the standards of individuals in designating their various levels of discrepancies. This approach permits substantial confidence in the results in Table 10.3.

That the accelerating model also best accounts for discrepancy/change interaction in Table 10.3 can be established by a two-stage analysis, the first stage of which suggests that prior discrepancy does not seem to be associated with learning or change. Consider columns 4 (Now 2) and 7 (F-values) in Table 10.3.
### Table 10.3: Size of Initial Discrepancy and Two Measures of Learning or Change

<table>
<thead>
<tr>
<th>Likert System</th>
<th>Number of respondents</th>
<th>Now 1</th>
<th>Now 2</th>
<th>Prior discrepancy, or ideal 1–Now 1</th>
<th>Now 2– ideal 1 discrepancies</th>
<th>F-ratios for differences in levels of Now 2 scores, adjusted by differences in ideal 1–Now 1 discrepancies</th>
<th>F-ratios for differences in levels of Now 2–ideal 1 discrepancies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>1</td>
<td>7</td>
<td>3.1</td>
<td>6.0</td>
<td>12.9</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>38</td>
<td>8.0</td>
<td>12.4</td>
<td>8.2</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>154</td>
<td>12.8</td>
<td>14.3</td>
<td>4.3</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>130</td>
<td>17.2</td>
<td>16.2</td>
<td>1.0</td>
<td>-1.0</td>
<td>23.88(^a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.31(^b)</td>
</tr>
<tr>
<td>Item 2</td>
<td>1</td>
<td>7</td>
<td>2.9</td>
<td>6.4</td>
<td>15.0</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>55</td>
<td>8.0</td>
<td>11.0</td>
<td>11.1</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>168</td>
<td>12.7</td>
<td>14.5</td>
<td>5.1</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>99</td>
<td>17.9</td>
<td>16.2</td>
<td>0.7</td>
<td>-1.7</td>
<td>12.89(^b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.39(^b)</td>
</tr>
<tr>
<td>Item 3</td>
<td>1</td>
<td>37</td>
<td>3.5</td>
<td>9.3</td>
<td>13.3</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>120</td>
<td>8.2</td>
<td>10.9</td>
<td>8.6</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>137</td>
<td>12.6</td>
<td>12.6</td>
<td>4.4</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>59</td>
<td>17.4</td>
<td>14.5</td>
<td>1.3</td>
<td>-2.7</td>
<td>5.04(^b)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.04(^b)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>3.5</td>
<td>10.2</td>
<td>13.5</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>51</td>
<td>8.2</td>
<td>12.3</td>
<td>7.2</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Item 4</td>
<td>215</td>
<td>12.8</td>
<td>13.5</td>
<td>4.4</td>
<td>0.7</td>
<td>5.01&lt;sup&gt;b&lt;/sup&gt;</td>
<td>21.04&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Item 5</td>
<td>215</td>
<td>12.8</td>
<td>13.5</td>
<td>4.4</td>
<td>0.7</td>
<td>2.50</td>
<td>20.32&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Item 6</td>
<td>217</td>
<td>12.9</td>
<td>13.6</td>
<td>5.1</td>
<td>0.1</td>
<td>3.74&lt;sup&gt;a&lt;/sup&gt;</td>
<td>19.34&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Item 7</td>
<td>135</td>
<td>12.8</td>
<td>13.7</td>
<td>5.0</td>
<td>0.9</td>
<td>3.77&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.70&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Item 8</td>
<td>147</td>
<td>12.8</td>
<td>14.2</td>
<td>5.0</td>
<td>1.5</td>
<td>10.89&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.93&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Item 9</td>
<td>148</td>
<td>12.9</td>
<td>13.8</td>
<td>4.9</td>
<td>0.9</td>
<td>11.15&lt;sup&gt;b&lt;/sup&gt;</td>
<td>8.31&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Optimum Discrepancy to Better Target Interventions
<table>
<thead>
<tr>
<th>Item</th>
<th>Likert System</th>
<th>Number of respondents</th>
<th>Now 1</th>
<th>Now 2</th>
<th>Prior discrepancy, or ideal 1–Now 1</th>
<th>Now 2–Now 1</th>
<th>F-ratios for differences in levels of Now 2 scores, adjusted by differences in Ideal 1–Now 1 discrepancies</th>
<th>F-ratios for differences in levels of Now 2–Now 1 scores, adjusted by differences in Ideal 1–Now 1 discrepancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3</td>
<td>172</td>
<td>13.2</td>
<td>13.7</td>
<td>5.1</td>
<td>0.5</td>
<td>10.32&lt;sup&gt;b&lt;/sup&gt;</td>
<td>15.50&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>136</td>
<td>17.3</td>
<td>15.4</td>
<td>1.3</td>
<td>−1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>2.8</td>
<td>12.8</td>
<td>12.2</td>
<td>10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>60</td>
<td>8.2</td>
<td>12.3</td>
<td>6.0</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>145</td>
<td>13.0</td>
<td>13.7</td>
<td>3.6</td>
<td>0.6</td>
<td>8.57&lt;sup&gt;b&lt;/sup&gt;</td>
<td>25.83&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>119</td>
<td>17.3</td>
<td>14.9</td>
<td>1.0</td>
<td>−2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>17</td>
<td>3.4</td>
<td>7.1</td>
<td>11.2</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>72</td>
<td>8.2</td>
<td>11.2</td>
<td>6.1</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>141</td>
<td>12.8</td>
<td>13.3</td>
<td>3.4</td>
<td>0.5</td>
<td>22.55&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.26&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>99</td>
<td>17.6</td>
<td>16.2</td>
<td>1.1</td>
<td>−1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6</td>
<td>3.0</td>
<td>11.5</td>
<td>13.7</td>
<td>8.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>43</td>
<td>8.3</td>
<td>10.7</td>
<td>8.0</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>177</td>
<td>12.8</td>
<td>13.2</td>
<td>4.4</td>
<td>0.4</td>
<td>12.74&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9.21&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>103</td>
<td>16.8</td>
<td>15.1</td>
<td>1.8</td>
<td>−1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>2.4</td>
<td>11.6</td>
<td>16.0</td>
<td>9.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>37</td>
<td>8.2</td>
<td>11.4</td>
<td>8.9</td>
<td>8.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Optimum Discrepancy to Better Target Interventions

<table>
<thead>
<tr>
<th>Item</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1.4</td>
<td>1.2</td>
<td>1.6</td>
<td>1.8</td>
<td>2.0</td>
<td>2.2</td>
<td>2.4</td>
<td>2.6</td>
<td>2.8</td>
<td>3.0</td>
<td>3.2</td>
<td>3.4</td>
<td>3.6</td>
<td>3.8</td>
<td>4.0</td>
<td>4.2</td>
<td>4.4</td>
<td>4.6</td>
</tr>
<tr>
<td>15</td>
<td>1.5</td>
<td>1.7</td>
<td>1.9</td>
<td>2.1</td>
<td>2.3</td>
<td>2.5</td>
<td>2.7</td>
<td>2.9</td>
<td>3.1</td>
<td>3.3</td>
<td>3.5</td>
<td>3.7</td>
<td>3.9</td>
<td>4.1</td>
<td>4.3</td>
<td>4.5</td>
<td>4.7</td>
<td>4.9</td>
</tr>
<tr>
<td>16</td>
<td>1.6</td>
<td>1.8</td>
<td>2.0</td>
<td>2.2</td>
<td>2.4</td>
<td>2.6</td>
<td>2.8</td>
<td>3.0</td>
<td>3.2</td>
<td>3.4</td>
<td>3.6</td>
<td>3.8</td>
<td>4.0</td>
<td>4.2</td>
<td>4.4</td>
<td>4.6</td>
<td>4.8</td>
<td>5.0</td>
</tr>
<tr>
<td>17</td>
<td>1.7</td>
<td>1.9</td>
<td>2.1</td>
<td>2.3</td>
<td>2.5</td>
<td>2.7</td>
<td>2.9</td>
<td>3.1</td>
<td>3.3</td>
<td>3.5</td>
<td>3.7</td>
<td>3.9</td>
<td>4.1</td>
<td>4.3</td>
<td>4.5</td>
<td>4.7</td>
<td>4.9</td>
<td>5.1</td>
</tr>
<tr>
<td>18</td>
<td>1.8</td>
<td>2.0</td>
<td>2.2</td>
<td>2.4</td>
<td>2.6</td>
<td>2.8</td>
<td>3.0</td>
<td>3.2</td>
<td>3.4</td>
<td>3.6</td>
<td>3.8</td>
<td>4.0</td>
<td>4.2</td>
<td>4.4</td>
<td>4.6</td>
<td>4.8</td>
<td>5.0</td>
<td>5.2</td>
</tr>
<tr>
<td>19</td>
<td>1.9</td>
<td>2.1</td>
<td>2.3</td>
<td>2.5</td>
<td>2.7</td>
<td>2.9</td>
<td>3.1</td>
<td>3.3</td>
<td>3.5</td>
<td>3.7</td>
<td>3.9</td>
<td>4.1</td>
<td>4.3</td>
<td>4.5</td>
<td>4.7</td>
<td>4.9</td>
<td>5.1</td>
<td>5.3</td>
</tr>
<tr>
<td>20</td>
<td>2.0</td>
<td>2.2</td>
<td>2.4</td>
<td>2.6</td>
<td>2.8</td>
<td>3.0</td>
<td>3.2</td>
<td>3.4</td>
<td>3.6</td>
<td>3.8</td>
<td>4.0</td>
<td>4.2</td>
<td>4.4</td>
<td>4.6</td>
<td>4.8</td>
<td>5.0</td>
<td>5.2</td>
<td>5.4</td>
</tr>
</tbody>
</table>
### Table 10.3 (continued)

Mean scores, classified by system

<table>
<thead>
<tr>
<th>Likert System</th>
<th>Number of respondents</th>
<th>Now 1</th>
<th>Now 2</th>
<th>Prior discrepancy, or ideal 1–Now 1</th>
<th>Now 2–Now 1</th>
<th>F-ratios for differences in levels of Now 2 scores, adjusted by differences in Ideal 1–Now 1 discrepancies</th>
<th>F-ratios for differences in levels of Now 2–Now 1 scores, adjusted by differences in Ideal 1–Now 1 discrepancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 21</td>
<td>3</td>
<td>213</td>
<td>12.9</td>
<td>13.8</td>
<td>3.7</td>
<td>0.9</td>
<td>25.33&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>67</td>
<td>17.1</td>
<td>15.0</td>
<td>0.4</td>
<td>-2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7</td>
<td>1.9</td>
<td>8.4</td>
<td>15.1</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>52</td>
<td>8.1</td>
<td>12.1</td>
<td>8.6</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Item 22</td>
<td>3</td>
<td>156</td>
<td>12.7</td>
<td>13.4</td>
<td>4.6</td>
<td>0.7</td>
<td>3.06&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>114</td>
<td>17.1</td>
<td>14.7</td>
<td>1.3</td>
<td>-2.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>23</td>
<td>2.9</td>
<td>9.2</td>
<td>10.9</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>108</td>
<td>8.1</td>
<td>10.6</td>
<td>7.1</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Item 23</td>
<td>3</td>
<td>148</td>
<td>12.7</td>
<td>13.0</td>
<td>3.5</td>
<td>0.3</td>
<td>6.90&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>50</td>
<td>16.9</td>
<td>13.6</td>
<td>0.3</td>
<td>-3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>17</td>
<td>5.3</td>
<td>9.8</td>
<td>14.0</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>59</td>
<td>7.8</td>
<td>11.9</td>
<td>9.0</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>Item 24</td>
<td>3</td>
<td>165</td>
<td>12.7</td>
<td>13.4</td>
<td>4.6</td>
<td>0.7</td>
<td>5.20&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>88</td>
<td>17.1</td>
<td>15.2</td>
<td>1.4</td>
<td>-1.9</td>
<td></td>
</tr>
</tbody>
</table>

Notes: <sup>a</sup> designates that P < .05. F = 2.62 for 3, 324 d.f.

<sup>b</sup> designates that P < .01. F = 3.85 for 3, 324 d.f.
Note that twenty-two of the twenty-four differences between Now 2 means for the four Likert systems still achieve usually accepted levels of statistical significance, using analysis of covariance techniques whose F-ratios are reported in column 7. The OD intervention, in sum, did not eliminate the major differences in Now 1 scores.

Second, however, this suggestion of quite uniform change between T 1 and T 2 does not stand up to the acid test. Specifically, columns 6 and 8 in Table 10.3 imply that the greater the initial discrepancy, the greater the relative change toward System 4. Indeed, the individuals who initially perceive their organization climates as being System 4 not only change the least in the expected direction, but on the average, they learn negatively, as it were. In sum, the apparent similarities in relative Now 1 and Now 2 scores at times can hide more than they reveal.

The accelerating model of discrepancy/change in Figure 10.1 applies to over 90 percent of the cases in columns 6 and 8 of Table 10.3, in sum. Two cases—Items 1 and 21—are deviant. They follow the reversible model in Figure 10.1. All F-ratios for Now 2–Now 1 scores, as adjusted, far surpass usually accepted levels of statistical significance for all twenty-four items. That is to say, all twenty-two cases fitting the accelerating model reflect statistically significant between- versus within-variance for the four Likert systems. The two cases of the reversible model also attain statistical significance. Consistently, although autocorrelation effects are possible, average $r = .46$ for Now 2–Now 1 scores and Ideal 1–Now 1 scores on the twenty-four Likert items.

GREATER SUCCESS BY SPECIFYING A LEARNING MODEL

The data above generate a number of implications which, if necessarily tentative, touch profound concerns in OD programs as well as in such areas as budgeting and advertising. Clearly, the determination of an optimum discrepancy will affect the answers to two questions:

What implications for enhanced OD success rates do the present data suggest?

Can the literature help us test the present perspective on a learning model featuring discrepancies?

Under this first heading, four points get attention. They all relate to the probability of greater success in OD efforts that specify in detail the properties of their learning theory. Most dramatically, first, the data urge major attention to the tailoring of OD designs to the characteristics of change targets. This does not outrage common sense, and it also gets support from research resting on different designs and methodologies (e.g., most prominently, Bowers and Hauser, 1977) as well as from emerging prescriptive formulations (Bowers, Franklin,
The data above imply that the prior discrepancies between perceived and ideal organization climate are massive covariants of change.

What do the present results imply for designing OD interventions? Only illustrations are possible here, but they can help. For example, high discrepancies might well respond to different kinds of designs than low discrepancies. In the latter case, for example, the primary goal might focus on raising attention to ideals that are not so lofty as to make them seem like pie-in-the-sky, by-and-by. Various “visioning” designs seem applicable in such cases (e.g., Lindaman and Lippitt, 1979), as do activities focusing on developing a broad “culture statement” or a “mission statement” at an operational level (e.g., Walton and Hackman, 1986).

For organizations whose members have high prior discrepancies, on the other hand, the primary focus would be on raising the perceived or actual level of functioning. A broad range of conventional designs seems applicable here. For example, three-dimensional images (Golembiewski 1979a, pp. 318–323) seem to have high success rates in alerting persons to unsatisfactory conditions and in motivating change. Role negotiation designs have a similar impact on actual or perceived functioning, with real but subsidiary effects on the ideal (e.g., Harrison, 1972).

The data above imply the usefulness—if not the necessity—of making constructive response in OD designs to differences in ideal/actual discrepancies. Consider the post-design regression in Likert scores reported by the average low-discrepancy respondent. What causes that major trend? One might argue that the OD design simply worsened the objective situation for the low-discrepancy respondents. But other plausible explanations also demand attention. Perhaps low-discrepancy respondents had unrealistic original perceptions. In this case, the public team-building design would have impressed on them the unrealism of their initial assessments of their units’ climate. This is a benign effect, as effects go, and probably beneficial. But it does suggest that separate subdesigns might have to be developed for prior discrepancies.

Other possibilities similarly encourage reliance on two-stage learning designs. For example, “social desirability” (Crowne and Marlow, 1964) might be related to the reporting of low prior discrepancies, with some persons having a strong tendency to see and report what is socially acceptable or desirable, as contrasted with what exists. Available evidence shows such an effect (e.g., Golembiewski and Munzenrider, 1975), and this encourages more intensive diagnosis in the pre-intervention phases. See also Irony III below, which deals with the impact on OD success rates of differentiating people and their properties. A two-stage design in the case of high social desirability would first focus on sensitizing people to their sunny bias, and then perhaps on moderating that bias. Conventional confrontation designs seem to have both effects (e.g., Golembiewski, 1983), and these effects also have been associated with T-groups or sensitivity.
training (e.g., Golembiewski, 1972, pp. 252–71). The second stage would then rely on these preparations to enhance problem-solving.

Whatever the case, the next step is patent. Research using different OD designs for individuals and groups at different take-off points is indicated. For example, low-discrepancy respondents might profit from a more impactful design than the present one. In other cases, as when respondents report that relationships in their organization are more or less where they prefer them to be, it is better to let well-enough alone.

The second implication for enhanced success derives from the suggestion in the data that prior discrepancy is directly associated with the change or learning. Qualifications, however, are appropriate. Thus, a more impactful design—such as a week-long family T-group—might generate a different pattern of change or learning. It is credible, for example, that a high-impact design might threaten high-discrepancy respondents so as to inhibit or preclude attitudinal or behavioral change. On the other hand, even the largest discrepancies in the present data batch may not have been big enough to trigger effects more consistent with Models II and, especially, III. Fortunately, this hypothesis is researchable. The sparse evidence available (e.g., Golembiewski, 1970, esp. 23–34) does not permit unqualified interpretation, but it implies that the accelerating model also best explains the effects of even more impactful interventions than are focal here.

Third, as is the case with all research reporting changes in extreme scores, it is possible that the data above reflect correlation effects, or regression toward the mean (Campbell, 1963). Table 10.3 in part controls for such effects by adjusting both Now 2 and Now 2–Now 1 scores for differences in prior discrepancies, but such statistical manipulations clearly do not eliminate the alternative hypothesis that correlation effects are more operative than experimental effects.

Fourth, most OD interventions probably tend to favor Model III, or the reversible model; and, if so, surprise is appropriate. Indeed, results may hold only for measures like the Likert Profile. Some measures of prior discrepancy are open-ended, and beyond some point escalating differences between anchor and stimulus might lose credibility or even become grotesque. In such cases, Model II might not apply. Hitler’s “big lie” technique might not work, in short. His advice was that if one is going to tell a lie with the intent of changing someone’s behavior, one should use the biggest whopper conceivable as a stimulus.

**Theoretical Leads Toward Enhanced Success**

The present research design has significant limitations, in sum, but the literature also contains some direction for improving the reach and grasp of the search for an optimum discrepancy. In short, the results above may motivate research on a broader range of possible models, and in the long run that will lead to greater success in OD applications. Thus enlarged, the conceptual range of discrepancies
will relate to crucial aspects of choice and change: for example, to fitting OD
designs to existing conditions, to determining the kind and intensity of feedback
appropriate for eliciting hearing and learning, and so on.

The passage of some ten years since Ironies I does not require much change
in the following text. The challenge still essentially remains, that is, and so does
the potential leverage in responding to Irony III.

**Model I: Greater Discrepancy, Greater Change**

What does the literature tell us about optimum discrepancies for inducing learning
or change, beyond the results and the trio of models depicted in Figure 10.1? To
review, Model I is no doubt the most popular view. Basically, it predicts that
the greater the discrepancy between a stimulus and an individual’s internal an-
chors, the greater the probable change in attitudes and opinions so as to reduce
that discrepancy. The tension induced by discrepancy acts as a motivator, propose
such syntheses as Festinger’s theory of cognitive dissonance. As he notes, to
respect: “. . . the discrepant reality which impinges on a person will exert pres-
sures in the direction of bringing the appropriate cognitive elements into corre-
spondence with that reality. . . . The strength of the pressures to reduce the disso-
nance is a function of the magnitude of dissonance” (1957, p. 18).

Native cunning raises strong questions about Model I, however, and this
courages putting the results above in a special perspective. In OD terms, for
example, Model I may not provide an optimum discrepancy under degenerative
interaction—where risk is high, as well as where openness, owning, and trust
are low. The OD learning design in the present case may have faced only a mildly
degenerative system, to put it directly. Hence the results above favoring Model
I may not apply universally, but only where OD values already exist or can be
vivified easily.

**Model II: Greater Discrepancy, Lesser Change**

To continue, Model II proposes a stark alternative. This view gets no prominent
theoretical support, but practice is another matter. Many of us act this way, much
of the time, as in tethering short our feedback to others. The underlying expecta-
tion is that people can bring massive forces to bear to fend off even incisive
thunderbolts of feedback—as by denial, misinterpretation, or whatever. So the
common guide often is: Easy as she goes, and especially to pressure the relation-
ship if resistance develops.

In sum, one can argue that Model II is more appropriate for advanced de-
generative conditions—that is, for what often exists in relationships between indi-
viduals and groups. Under regenerative interaction, in contrast, Model II might
provide an overcautious guide for discrepancies.
Model III: Reversible When Ego Involvement Varies

To complete the review of Figure 10.1, Model III encourages introducing an intervening variable for discrepancy-to-change linkages—“ego involvement,” or the condition when a person feels strongly about an issue that is seen as a central one.

Two conditions seem involved. When ego involvement is low, Model I obtains. That is, the reasonable course of least resistance is for an individual to move attitudinally toward the discrepant stimulus or to perceive it as less discrepant than it is. In sum:

<table>
<thead>
<tr>
<th>Low Ego Involvement</th>
<th>Discrepancy</th>
<th>Probable Effect</th>
<th>Change Toward Discrepant Stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Small</td>
<td>Assimilation</td>
<td>Small</td>
<td></td>
</tr>
<tr>
<td>(2) Great</td>
<td>Assimilation</td>
<td>Great</td>
<td></td>
</tr>
</tbody>
</table>

When ego involvement is high, in contrast, different dynamics may be expected. For example, beyond some point, growing discrepancies may induce high ego involvement—that is, “That’s outrageous; I can’t accept that.” Or ego involvement may be high for some other reason. In either case:

<table>
<thead>
<tr>
<th>High Ego Involvement</th>
<th>Discrepancy</th>
<th>Probable Effect</th>
<th>Change Toward Discrepant Stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) Small</td>
<td>Assimilation</td>
<td>Great</td>
<td></td>
</tr>
<tr>
<td>(4) Great</td>
<td>Contrast</td>
<td>Small</td>
<td></td>
</tr>
</tbody>
</table>

That is, increases in discrepancy will generate increasing change only up to a point, some students propose. As Whittaker explains: “Small discrepancies yielded small change, moderate discrepancies yielded the maximum change, and the degree tended to diminish as the discrepancies became even larger” (1967, pp. 168–169). In sum, the first two cases above (that is, when ego involvement is low) refer to the rising part of the curve in Model III. The third and fourth cases (that is, when ego involvement is high) refer to segments of the curve beyond the point of reversal.

So, how do the present results apply to Model III? Directly, the model does
not get as much support as the descriptor of data trends in Table 10.3. Moreover, one assumes that ego involvement was high in the learning design, but the results do not tend to fall along tracks 3 or 4 above; one cannot easily argue that discrepancies in the present case are not large enough to "go over the hill," as it were. Table 10.2 shows a mean discrepancy of 13.4 for Likert’s System 1, however, and that on a scale of twenty equal-appearing intervals.

**Model IV: Balance Hypothesis**

The present research design neglects an important alternative model: Heider’s emphasis on the balanced state, which “designates a situation in which the perceived units and the experienced sentiments co-exist without stress; there is thus no pressure toward change either in the cognitive organization or in the sentiment” (1958, p. 177). Heider adds: “That sentiment, unit formation, and balanced state have something to do with each other can be stated as a general hypothesis, namely: the relationship between sentiments and unit formation tends toward a balanced state” (1958, p. 177).

In general, Model IV gets much support from the OD literature, on those rare occasions when the present issues get self-conscious attention. Thus, OD start-up often begins by trying to disturb a balanced state—as by indicating its perceived unattractiveness to many group members, or by modeling an alternative to the existing condition. Very often, individual observers may know they dislike a situation, but remain unaware that colleagues share their view. Or many individuals may suffer an unattractive situation because they see no alternative to it, and least of all an alternative with general peer support.

More specifically, Model IV also implies significant guides for choice and change, especially in that model’s acceptance of Model II as a guide to effectively moving unbalanced systems. As Heider notes: “It has been assumed, and with good support from experimental findings, that ‘Assimilation appears when the differences between the substructures [of a unit] are small; contrast appears when the differences are large’ ” (1958, p. 211).

Model IV also relates to optimum discrepancy and choice, please note, in multiple ways. Consider one unbalanced situation—a case in which person A does not like person B, but A discovers that B is the author of a poem of which A said: “Only a beautiful spirit could have written it.” In sum (Golembiewski, 1972, p. 104), a new balance will develop between A’s sentiments toward B and his original definition of B as a unit. The unbalanced combination of positive sentiments about the poem and negative evaluation of the author can be resolved in several ways:

- A might see the person-author B as a total positive unit; consequently, A comes to like B more.
- A might soon decide he does not really like the poem, thereby creating a
balance with B seen as a totally negative unit; consequently, B is disliked at least as much as originally.
A might doubt or deny B’s authorship of the poem; in which case B is disliked as much or more than originally, and so on.

The choice between these alternative resolutions presumably will depend on the magnitude of the differences between B’s substructures, that is, B as person and B as author. The greater the differences, the more likely are alternative balances like the two above.

IN-PROCESS CONCLUSIONS, RELEVANT IN 1990 AND 2002 AS WELL

This chapter does and does not, in sum. It does illustrate the usefulness of a learning model with discrepancy features, for which it also does detail one research design and some supporting results. Discrepancy models have long been central in OD—as in the dilemma/invention model that goes back to the earliest days, or as in the sharing of three-dimensional images, which similarly use discrepancies to motivate exploration and possible change.

This consciousness-raising with some data has a value, but there are also many things the chapter does not do. Thus, the chapter does more modeling of what can be done, than it does accomplishing the job. Moreover, this chapter does not provide definite guidance for OD intervenors, even as it does highlight major choice-points in seeking to induce a specific discrepancy—whether by a simple piece of feedback or by a global learning design.

Finally, this chapter does not take into explicit account that more than one kind of change occurred. If that did happen—and especially if what Chapter 24 calls “gamma change” did occur—the interpretations of empirical findings badly need revision. But that possibility only complicates this analysis, it does not invalidate it. See later chapters of this book for emerging details leading to the conclusion that a useful concept needs to accommodate a plural notion of change.

REFERENCES

Not Every Design “Works” Everywhere
Greater Sensitivity to Interaction of Situations and Designs

The planned-change literature contains many general guides for praxis, but cannot confidently specify the situational features that are associated with high success rates of specific designs; hence, this enriched replication of earlier research concerning the power of 12 often-prescribed guides to predict success rates in Quality of Working Life (QWL) applications.

A pilot study (N = 231) generated interesting but limited results (Golembiewski and Sun, 1990). Thus, no consistent associations were found to exist for the total batch, but clear patterns existed for each of four major types of QWL applications—Human Processual, Sociotechnical Systems, Technostructural, and Combined. The size of this panel permits the breakout of QWL subtypes, which was beyond the capacity of all previous surveys of QWL applications.

The conclusions about situational covariants are essentially the same for two measures of success rates—hard criteria and global assessments. Inclusion in the panel requires objective measures of success, but all cases also provide “soft” measures—self-reports of attitudes, and so on. Global assessments of QWL effects combine the two measures of outcomes.

This study (N = 314) replicates the results of the earlier pilot study. The replication permits confidence beyond the pilot, and its results require no basic change in interpretation. In sum, QWL success rates are substantial in both business and government, and the data reveal a definite pattern of how situational features are associated with success of QWL applications.
A CONCEPTUAL CONTEXT FOR IRONY IV

Irony III relates to substantial OD success without specifying a learning theory that details the situational features associated with success or failure. Such associations seem highly profitable: indeed, the OD literature has proposed self-consciously working guidelines for just such effects. For example, OD shoptalk often has proposed the greater difficulties of change in the public versus business sectors, typically due to the lack of a “profit motive” in government.

However, the status of such working guidelines remains provisional. Those guidelines are not beyond confirmation or disconfirmation, but the exercise has seldom been attempted, and never satisfactorily so. In large part, this has historically been due to the small number of comparable OD applications in almost all research panels. But as several earlier chapters indicate, this condition need no longer be accepted as a given.

Hence, the centrality of the effort here. It will seek to access the readiness of OD evaluative studies to support the required test, and this effort also may provide guidance as to how that evaluative literature needs to be developed in order to provide the most telling evidence about which situational features are associated with heightened success rates and which features imply lower success rates.

In either case, knowledge about such situational features will heighten success rates, beyond the present substantial levels established in earlier chapters. The existing association—the irony of high success rates along with the inability to take explicit advantage of favorable situational features—suggests aspects of two generalizations. First, some overall OD designs may be suited to broad ranges of problem situations. Or, second, many intervenors may have the capacity to tailor generic designs to different specific features, although largely in implicit or intuitive ways.

GETTING BEYOND PRESENT STATES IN OD AND QWL

All effective praxis involves a serious leap-frogging toward greater comprehensiveness in attaining intended effects, based upon accumulating evidence from experience as well as from deliberate tests of the conventional wisdom. Basically, we can learn how praxis can be successful, or how and possibly why existing theory-guided practice needs refining or replacement.

In these terms, the two most popular approaches to planned change—Organization Development (OD) and Quality of Working Life (QWL)—remain stuck, if comfortably. Applications of both genres seem to “work” in a high percentage of cases, as thousands of applications in more than a score of independent surveys establish for both QWL and OD (Sun 1988, pp. 76–85; Golembiewski, 1998). See also Chapters 1 through 3.
Despite such attractive success estimates, we still lack a validated view of the specific situational features that enhance success rates, or depress them. An earlier study with a design similar to the present one is available (Golembiewski and Sun, 1990, 1990d), but even its substantial size \((N = 231)\) left a number of cells in the data matrix with few or no cases. Of course, this hedged the interpretation of results, and urged an extension of the period during which cases were sought—from 1965–86 in the pilot to 1996 in the present panel. Now, \(N = 314\).

In part, our lack of certainty about situational features does not surprise. Thus, there seems only a single post-1990 study that targets some situational features (Robertson, Roberts, and Porras, 1993), but it relies on only 47 studies and hence has strict limitations; hence, the centrality of this study and its pilot. As Beer (1988, p. 6) reminds us, OD research has been infatuated with the consultant and intervention methods. He concludes (1988, p. 6) that the associated neglect of situational features is significant, even momentous: “. . . the external environment and the internal organizational features that lead to the felt need for change . . . are far more important than the intervention method.” That seems reasonable, but no one can yet act on that position with certainty and specificity.

Many reasons help explain this shortfall about the situational features predisposing change applications toward favorable or unfavorable outcomes. First, a satisfactory assessment of situational features must rest upon an incredible amount of prior effort to generate the required critical mass of theory and experience in OD and QWL. That critical mass is only now becoming available, and few attempts to assemble sufficiently large panels have been made.

Second, most evaluations of QWL or OD applications deal with small batches of cases, and this limits the design of analyses. For example, Cummings, Molloy, and Glen (1975) use 57 cases; Cummings (1978) makes do with 16; and Schuster (1984) relies on 38 applications. Only Pasmore et al. (1982) employ a population even approaching the size of the one utilized here, and their panel includes 134 cases, or less than 43 percent the size of the present population. Of course, the granddaddy of all OD panels—\(N = 574\)—is reviewed in early chapters.

Third, more specifically, OD or QWL applications rely on diverse designs and use multiple measures of numerous targets of change. Hence, small populations typically will disguise or camouflage real differences, as by precluding breakouts of classes of interventions that can be cross-checked for associations between situational features and success rates.

Fourth, with only a few exceptions (e.g., Golembiewski, 1980; Golembiewski and Sun, 1990a; Nicholas, 1982; Park, 1990), surveys of both OD and QWL applications rely on “soft” variables or self-reported outcomes. For some observers, of course, only “hard” or “objective” measures of outcomes will deter charges of artifactuality. The view here is not so limiting, but multiple measures of outcomes clearly have much to recommend them.
Fifth, in the main, surveys of evaluative studies do not allow for the differences in the methodological rigor that exist between individual pieces of research. Of course, rigor has been central in some analyses (e.g., Morrison, 1978; Porras, 1979), but not in combination with either success rates or the situational features associated with them. See especially the detailed analysis in Chapter 3.

These five contributors to stuckness set the agenda for this analysis. Briefly, this paper relies on 314 applications to test twelve hypotheses about the conditions associated with favorable outcomes of QWL efforts. This large batch also permits the breakout of four classes of interventions. Moreover, inclusion in the batch requires that each eligible case report “hard” or objective outcomes—productivity rates, costs or savings, personnel turnover, and so on.

Our research design also permits estimating the effects on success rates of differences in methodological rigor, but that analysis is reserved for other times and places. Already published research with the N = 231 panel implies no major concerns about the interpretation of past or present results (e.g., Golembiewski and Sun, 1989, 1990a,b,c,d). In sum, different degrees of methodological rigor are not associated with significant variations in success rates; and no strong “positive findings bias” exists, as would be the case where the degree of rigor and success are inversely related.

Dealing with this agenda requires four major steps. They include a description of twelve hypotheses linking situational features with the success of QWL outcomes, a brief introduction to methods, the detailing of results, and a discussion of the findings as well as their implications.

**TWELVE HYPOTHESES AND THEIR RATIONALES**

Basically, this study tests major aspects of the common wisdom about which situational features predispose successful QWL application. For convenience, Table 11.1 refers to twelve “hypotheses,” but these might be better viewed as rules-of-thumb about success that get general respect in shop-talk. The original and reasonable supporting rationales appear elsewhere (Golembiewski and Sun, 1990b, pp. 102–106), and are summarized here in Appendix 11.A. The original rationales there rely on fuller citations, which are both pruned and updated in the appendix below. Early and later views of the rationales differ only in details.

The measure of association used in this analysis is Cramer’s V, especially to facilitate comparisons with earlier studies (Dunn and Swierczek, 1977; Golembiewski and Sun, 1990d). Cramer’s V provides not only a measure of the strength of any association in an asymmetric table, but also avoids chi-square’s sensitivity to small cells. Cramer’s V ranges from 0 to +1: a large value of V signifies that a high degree of association exists, but *without expressing direction*. Following an earlier study (Dunn and Swierczek, 1977), a value of V between 0 and .20 will be considered as negligible to weak; .21 to .30 is treated as weak; .31 to .50 is moderate; and .51 to 1.0 is rated as strong. Since the square of Cramer’s V...
## TABLE 11.1 Twelve Hypotheses Underlying QWL Praxis

<table>
<thead>
<tr>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. QWL tends to be less successful in an existing organization than in a start-up situation.</td>
</tr>
<tr>
<td>2. QWL applications face fewer obstacles in small organizations than in medium-sized or large organizations.</td>
</tr>
<tr>
<td>3. QWL efforts in business organizations are generally more effective than those in public-sector organizations.</td>
</tr>
<tr>
<td>4. QWL efforts in economic or profit-based organizations will be more successful than efforts carried out in other types of organizations.</td>
</tr>
<tr>
<td>5. QWL programs have a greater likelihood of success in nonunion settings.</td>
</tr>
<tr>
<td>6. Multiple QWL interventions are more likely to increase productivity and employee well-being than a single intervention.</td>
</tr>
<tr>
<td>7. QWL efforts that focus on more than one level in an organization will be more successful than those that focus on one level only.</td>
</tr>
<tr>
<td>8. QWL programs that are introduced throughout the whole organization are more successful than programs limited to part of the organization.</td>
</tr>
<tr>
<td>9. QWL programs that involve high levels of participation induce intended effects more frequently than those involving low levels of participation.</td>
</tr>
<tr>
<td>10. QWL programs introduced by both internal and external consultants are more likely to be successful than QWL programs introduced by either one or the other.</td>
</tr>
<tr>
<td>11. QWL programs introduced with the help of internal consultants are more likely to be successful than QWL programs introduced by external consultants.</td>
</tr>
<tr>
<td>12. The longer the system exposure to QWL activities, the greater the intended changes.</td>
</tr>
</tbody>
</table>

provides an estimate of the explained variance. a “weak” association in this study accounts for 4.4 to 9 percent of the variance in the association of a specific situational feature and an outcome measure, and a “strong” association explains over 25 percent.

Two other conventions complete the introductory task. All hypotheses whose V falls below .20 or below will be considered “rejected,” as a convenience. In cases where V > .20, the direction of paired associations for each situational features X success rates will be individually assessed. That is, V only estimates magnitude, but direction also requires specification for many purposes.

### FINDINGS

This analysis of situational features and QWL outcomes generates three basic conclusions. Overall, the 314 QWL cases in the present panel have substantial success rates.
Moreover, when the QWL panel is considered as a whole, no notable associations are isolated. This second conclusion implies that QWL’s common wisdom is not broadly helpful. This finding seems to support both Dunn and Swierczek’s (1977) seminal efforts, as well as our pilot study (Golembiewski and Sun, 1990d).

Finally, these associations became quite regular and even robust for three of the four classes, when analysis considers breakouts of the four classes of QWL interventions (see Exhibit 11.1). This conclusion implies that guides for QWL praxis must be intervention-specific—with Sociotechnical System interventions having the distinctly weakest pattern of associations between situational features and QWL success rates. The other three classes of interventions—Human-Processual, Technostructural, and Combined—have virtually identical patterns. This goes well beyond Dunn and Swierczek’s path-breaking analysis, and elaborates on the finding of our pilot study.

Four emphases develop these three central conclusions. An initial section reports on the general associations between situational features and QWL outcomes in the total batch of 314 applications. A second section tests for the specific associations of each of the twelve hypotheses with each of the two measures of effects, considering the aggregate of 314 QWL cases. Then, a third section summarizes similar tests, but this time for each of the four major classes of QWL applications (see Exhibit 11.1). A fourth and concluding section speculates about why specific situational features show signs of statistical association with three classes of interventions but not with the fourth.

All QWL Cases, Success Estimates by Situational Features

Table 11.2 summarizes the data in a broad sense—for all QWL cases, for 12 situational features, both only for hard-criteria estimates of success. Of course, most observers consider hard criteria as providing a more demanding test of QWL effects. However, a similar table details QWL outcomes based on a “global assessment,” which incorporates both hard and soft criteria, as explained above. This counterpart table is not reproduced here because it generates a pattern of results that differ from Table 11.2 only in minor ways.

Note also that Table 11.2 is so designed that for each hypothesis, it first lists the coding category expected to be associated with the most positive QWL outcomes. Subsequent coding categories are then listed in order of the hypothesized declining attractiveness of their expected outcomes. Hence for Hypothesis 6, three or more interventions are expected to generate the most positive QWL effects; one intervention will do poorest; and any pair of interventions is expected to generate an intermediate success rate.

Two firm conclusions characterize this treatment of all 314 QWL applications as an aggregate. With only minor qualifications, first, Table 11.2 establishes that QWL interventions of all kinds, for all situational features, have substantial success rates, even formidable ones.
EXHIBIT 11.1 Four Main Classes of QWL Interventions, with Component Types

A. Human-Processual
   1. Process-oriented training that involves T-groups, encounter groups, behavior-modeling techniques, or any other group training experiences that focus on interpersonal behavior and group process issues and are structured to permit specific learning that can be transferred to work setting.
   2. Team-building encompasses a variety of activities in which intact work groups experientially learn to increase their skills for effective teamwork by utilizing a structured agenda, usually with the help of a behavioral-science consultant.
   3. Grid organizational development (Blake and Mouton, 1969), which in its full form involves six separate but interrelated phases.
   4. Process consultation to improve group and personal functioning.
   5. Survey feedback.
   6. Goal-setting is commonly associated with MBO. Yet there are important differences between the two. Although the setting of specific performance goals is an essential feature of MBO applications, they typically attempt to accomplish much more than simply setting goals and clarifying roles. See also D1 below.
   7. Joint union-management settings involve a voluntary, cooperative relationship, outside of the collective bargaining process. It is usually co-chaired by one representative of labor and another from management and serves primarily as a formal action-planning body to which proposals for increased productivity and quality of working life can be submitted by employees (e.g., Frucher, 1980; Hendrix and Lloyd, 1982).

B. Sociotechnical Systems
   This approach to organizational design integrates technological and social subsystems to create a flexible structure that can deal with environmental turbulence while affording organizational effectiveness. This intervention involves the restructuring or redesigning of work methods, technology, social structures, and relationships among roles or tasks. Self-maintaining, semi-autonomous work groups are typical products of this approach.

C. Technostructural
   1. Work redesign approaches specify job content, methods, and relationships in order to satisfy the technological and organizational requirements of work as well as the social and personal requirements of the job-holder. This often involves job enrichment or job enlargement.
   2. Alternative work schedules include two major forms: flexible working hours; and redistribution of hours in a workweek of (e.g.) four 10-hour days.
3. Typical compensation plans tie monetary rewards to individual, group, or organization-wide gains. Some programs are purely financial (e.g., encouraging increased productivity via simple financial incentives), others mix monetary rewards with employee participation and feedback (e.g., Scanlon, Impreshare, and Gain Sharing).

4. Organizational redesign deals with patterns of responsibility, authority, activity and communication at macro levels. Three subcategories under the rubric are workflow reorganization, parallel organization, and matrix organization.

D. Combined

1. Management by objective (MBO) involves processes by which organizational purposes are met by joining superiors and subordinates in the pursuit of mutually agreed-upon goals and objectives, which ideally are specific, measurable, time-bounded. And joined in an action plan.

2. Quality circles (QC) involve small groups (usually 5 to 12 members) performing similar work who meet, usually voluntarily, on a regular basis, to learn and apply techniques for identifying, analyzing, and solving work-related problems (e.g., Park, 1990). Employees can use their individual strengths in problem-solving to improve product quality while satisfying personal needs.

3. Multifaceted approaches employ more than one type of QWL intervention, simultaneously or in succession.

Second, the aggregated cases provide only modest support for the expected effects of the 12 situational features. Thus, using some reasonable if tedious conventions, overall, 65.8 percent of all the paired-comparisons in Table 11.2 fall in the expected directions. To illustrate, estimates of success rates are greater for new versus established sites, as well as for nonunion sites, as hypothesized.

All QWL Cases, Success Estimates by Twelve Hypotheses

However, no suggested effects stand up under close examination. Indeed, Table 11.3 shows not even a single case of nonrandom variation of success estimates with the 12 situational features, for the total batch, considering both global as well as hard-criteria assessments of success. Moreover, $V^2$ based on Table 11.3 implies that only minor portions of the variance are explained by the 24 tests. Specifically, the mean $V^2$ approximates .022, for global and hard-criteria assessments of performance. This falls in the “negligible” portion of the range of $V$ as a standard for association.
### Table 11.2: Summary Data re Twelve Hypotheses Relating Situational Features and Outcomes of Aggregated QWL Applications, Hard-Criteria Assessment, N = 314 Applications

<table>
<thead>
<tr>
<th>Hypothesis and coding categories</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Sub-N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>1. New plant organization</td>
<td>7</td>
<td>(87.5)</td>
<td>1</td>
<td>(12.5)</td>
<td>0</td>
<td>(0.0)</td>
</tr>
<tr>
<td>Established plant/organization</td>
<td>214</td>
<td>(69.9)</td>
<td>67</td>
<td>(21.9)</td>
<td>14</td>
<td>(4.6)</td>
</tr>
<tr>
<td>2. Small size (&lt;50)</td>
<td>50</td>
<td>(69.4)</td>
<td>15</td>
<td>(20.8)</td>
<td>4</td>
<td>(5.6)</td>
</tr>
<tr>
<td>Medium size (≥200)</td>
<td>96</td>
<td>(72.2)</td>
<td>31</td>
<td>(23.3)</td>
<td>5</td>
<td>(3.8)</td>
</tr>
<tr>
<td>Large size (&gt;200)</td>
<td>57</td>
<td>(69.5)</td>
<td>14</td>
<td>(17.1)</td>
<td>4</td>
<td>(4.9)</td>
</tr>
<tr>
<td>3. Private</td>
<td>170</td>
<td>(71.1)</td>
<td>51</td>
<td>(21.3)</td>
<td>10</td>
<td>(4.2)</td>
</tr>
<tr>
<td>Public</td>
<td>47</td>
<td>(67.1)</td>
<td>16</td>
<td>(22.9)</td>
<td>4</td>
<td>(5.7)</td>
</tr>
<tr>
<td>4. Economic</td>
<td>151</td>
<td>(68.9)</td>
<td>51</td>
<td>(23.3)</td>
<td>8</td>
<td>(3.7)</td>
</tr>
<tr>
<td>Service</td>
<td>25</td>
<td>(73.5)</td>
<td>6</td>
<td>(17.6)</td>
<td>2</td>
<td>(5.9)</td>
</tr>
<tr>
<td>Commonweal</td>
<td>41</td>
<td>(71.9)</td>
<td>11</td>
<td>(19.3)</td>
<td>4</td>
<td>(7.0)</td>
</tr>
<tr>
<td>Mutual benefit</td>
<td>0</td>
<td>(0.0)</td>
<td>0</td>
<td>(0.0)</td>
<td>0</td>
<td>(0.0)</td>
</tr>
<tr>
<td>5. Nonunion</td>
<td>37</td>
<td>(78.7)</td>
<td>6</td>
<td>(12.8)</td>
<td>2</td>
<td>(4.3)</td>
</tr>
<tr>
<td>Union</td>
<td>59</td>
<td>(61.5)</td>
<td>31</td>
<td>(32.3)</td>
<td>3</td>
<td>(3.1)</td>
</tr>
<tr>
<td>6. Three or more interventions</td>
<td>14</td>
<td>(87.5)</td>
<td>2</td>
<td>(12.5)</td>
<td>0</td>
<td>(0.0)</td>
</tr>
<tr>
<td>Two</td>
<td>16</td>
<td>(61.5)</td>
<td>10</td>
<td>(38.5)</td>
<td>0</td>
<td>(0.0)</td>
</tr>
<tr>
<td>One</td>
<td>191</td>
<td>(70.2)</td>
<td>56</td>
<td>(20.6)</td>
<td>14</td>
<td>(5.1)</td>
</tr>
<tr>
<td>Hypothesis and coding categories</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Multiple levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managers</td>
<td>104</td>
<td>(81.3)</td>
<td>14</td>
<td>(10.9)</td>
<td>6</td>
<td>(4.7)</td>
</tr>
<tr>
<td>Staff</td>
<td>8</td>
<td>(88.9)</td>
<td>0</td>
<td>(0.0)</td>
<td>1</td>
<td>(11.1)</td>
</tr>
<tr>
<td>First-line supervisors</td>
<td>6</td>
<td>(75.0)</td>
<td>1</td>
<td>(12.5)</td>
<td>1</td>
<td>(12.5)</td>
</tr>
<tr>
<td>Line employees</td>
<td>97</td>
<td>(64.7)</td>
<td>40</td>
<td>(26.7)</td>
<td>6</td>
<td>(4.0)</td>
</tr>
<tr>
<td>Total organization</td>
<td>105</td>
<td>(77.8)</td>
<td>22</td>
<td>(16.3)</td>
<td>3</td>
<td>(2.2)</td>
</tr>
<tr>
<td>Several parts of organization</td>
<td>44</td>
<td>(65.7)</td>
<td>18</td>
<td>(26.9)</td>
<td>5</td>
<td>(7.5)</td>
</tr>
<tr>
<td>Single part of organization</td>
<td>72</td>
<td>(64.3)</td>
<td>28</td>
<td>(25.0)</td>
<td>6</td>
<td>(5.4)</td>
</tr>
<tr>
<td>Total high level participation</td>
<td>57</td>
<td>(70.4)</td>
<td>16</td>
<td>(19.8)</td>
<td>5</td>
<td>(6.2)</td>
</tr>
<tr>
<td>Low level participation</td>
<td>33</td>
<td>(61.1)</td>
<td>17</td>
<td>(31.5)</td>
<td>3</td>
<td>(5.6)</td>
</tr>
<tr>
<td>No participation</td>
<td>34</td>
<td>(65.4)</td>
<td>14</td>
<td>(26.9)</td>
<td>1</td>
<td>(1.9)</td>
</tr>
<tr>
<td>Total high and external change-agents</td>
<td>24</td>
<td>(82.4)</td>
<td>3</td>
<td>(10.3)</td>
<td>1</td>
<td>(3.4)</td>
</tr>
<tr>
<td>Internal or external</td>
<td>62</td>
<td>(74.7)</td>
<td>15</td>
<td>(18.1)</td>
<td>4</td>
<td>(4.8)</td>
</tr>
<tr>
<td>External change-agent</td>
<td>18</td>
<td>(85.7)</td>
<td>3</td>
<td>(14.3)</td>
<td>0</td>
<td>(0.0)</td>
</tr>
<tr>
<td>Longer than 24 months</td>
<td>43</td>
<td>(75.8)</td>
<td>9</td>
<td>(16.1)</td>
<td>2</td>
<td>(3.6)</td>
</tr>
<tr>
<td>19 to 24 months</td>
<td>15</td>
<td>(78.9)</td>
<td>2</td>
<td>(10.5)</td>
<td>2</td>
<td>(10.5)</td>
</tr>
<tr>
<td>13 to 18 months</td>
<td>11</td>
<td>(64.7)</td>
<td>5</td>
<td>(29.4)</td>
<td>1</td>
<td>(5.9)</td>
</tr>
<tr>
<td>7 to 12 months</td>
<td>70</td>
<td>(75.3)</td>
<td>17</td>
<td>(18.3)</td>
<td>4</td>
<td>(4.3)</td>
</tr>
<tr>
<td>1 to 6 months</td>
<td>76</td>
<td>(67.9)</td>
<td>25</td>
<td>(22.3)</td>
<td>5</td>
<td>(5.4)</td>
</tr>
</tbody>
</table>

**Table 11.2 Continued**

<table>
<thead>
<tr>
<th>Degree of success in outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Highly positive</td>
</tr>
<tr>
<td>and intended effects</td>
</tr>
<tr>
<td>2 Definite balance of</td>
</tr>
<tr>
<td>positive and intended</td>
</tr>
<tr>
<td>effects</td>
</tr>
<tr>
<td>3 No appreciable effects</td>
</tr>
<tr>
<td>4 Negative effects</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Hypothesis and coding categories</td>
</tr>
<tr>
<td>Multiple levels</td>
</tr>
<tr>
<td>Managers</td>
</tr>
<tr>
<td>Staff</td>
</tr>
<tr>
<td>First-line supervisors</td>
</tr>
<tr>
<td>Line employees</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Several parts of organization</td>
</tr>
<tr>
<td>Single part of organization</td>
</tr>
<tr>
<td>Total high level participation</td>
</tr>
<tr>
<td>Low level participation</td>
</tr>
<tr>
<td>No participation</td>
</tr>
<tr>
<td>Total high and external change-agents</td>
</tr>
<tr>
<td>Internal or external</td>
</tr>
<tr>
<td>External change-agent</td>
</tr>
<tr>
<td>Longer than 24 months</td>
</tr>
<tr>
<td>19 to 24 months</td>
</tr>
<tr>
<td>13 to 18 months</td>
</tr>
<tr>
<td>7 to 12 months</td>
</tr>
<tr>
<td>1 to 6 months</td>
</tr>
</tbody>
</table>
**Interaction of Situations and Designs**

TABLE 11.3 Summary, Statistical Tests of Twelve Hypotheses, Global and Hard-Criteria Assessments, All QWL Cases

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Global assessment</th>
<th>Hard-criteria assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rejected (V = .08)</td>
<td>Rejected (V = .07)</td>
</tr>
<tr>
<td>2</td>
<td>Rejected (V = .18)</td>
<td>Rejected (V = .19)</td>
</tr>
<tr>
<td>3</td>
<td>Rejected (V = .16)</td>
<td>Rejected (V = .14)</td>
</tr>
<tr>
<td>4</td>
<td>Rejected (V = .15)</td>
<td>Rejected (V = .10)</td>
</tr>
<tr>
<td>5</td>
<td>Rejected (V = .17)</td>
<td>Rejected (V = .16)</td>
</tr>
<tr>
<td>6</td>
<td>Rejected (V = .11)</td>
<td>Rejected (V = .13)</td>
</tr>
<tr>
<td>7</td>
<td>Rejected (V = .14)</td>
<td>Rejected (V = .12)</td>
</tr>
<tr>
<td>8</td>
<td>Rejected (V = .18)</td>
<td>Rejected (V = .16)</td>
</tr>
<tr>
<td>9</td>
<td>Rejected (V = .12)</td>
<td>Rejected (V = .16)</td>
</tr>
<tr>
<td>10</td>
<td>Rejected (V = .14)</td>
<td>Rejected (V = .15)</td>
</tr>
<tr>
<td>11</td>
<td>Rejected (V = .13)</td>
<td>Rejected (V = .17)</td>
</tr>
<tr>
<td>12</td>
<td>Rejected (V = .17)</td>
<td>Rejected (V = .16)</td>
</tr>
</tbody>
</table>

Classes of QWL Interventions, Success Estimates by Twelve Hypotheses

The enlarged size of the present QWL panel permits distinguishing the success rates of each of the four classes of interventions. To summarize Exhibit 11.1, these classes of interventions are:

- **Human Processual**, which class includes process consultation, most team building, the Grid, and so on.
- **Sociotechnical systems**, which seek to integrate technological and social subsystems into a system flexible enough to deal effectively with environmental turbulence.
- **Technostructural**, which includes role analysis, the redesign efforts usually called job enlargement and especially enrichment, alternative work schedules, compensation plans like Gainsharing, and so on.
- **Combined**, which relies on several types of interventions, either simultaneously or successively. MBO and Quality Circles are two typical combined applications.

Distinguishing the four classes of interventions proves useful, but in two very different ways. First, as Table 11.4 illustrates for hard criteria, success estimates vary in only a narrow range by the four major classes of interventions. Even the most deviant class fits this generalization. Thus, Sociotechnical Systems designs have the greatest proportion of Category 4 assignments, more than doubling the average of the other three classes of interventions. At the same time,
Sociotechnical Systems designs achieve the largest proportion of Category 1 assignments.

A separate table shows the same pattern as Table 11.4—and virtually the same numerical estimates—for the global measure of success. This table is not reprinted here. Of interest, moreover, the hard-criteria assessment generates the higher success rate estimates—69.1 versus 62.4 percent, respectively, for highly positive and intended effects. Of course, most commentators attribute a superior status to hard-criteria or objective measures.

Second, the break-out of the four classes of QWL in Tables 11.5 and 11.6 interventions also generates two basic findings: a substantial majority of the tests of association between QWL outcomes and situational features show statistically significant variance; and the random cases decidedly cluster in one of the classes of QWL designs. This contrasts sharply with Table 11.3, which finds not even a single statistically significant association of the 24 tests of hypotheses dealing with situational differences, for the total panel of QWL applications, with success estimates in two ways.

Table 11.5 illustrates the pattern of results common to all of the counterpart tables for the four classes of QWL interventions. Specifically, eight of the eleven cases in which V signals a rejection of a hypothesis linking a situational feature to success of an application involve Sociotechnical Systems.

Much detail reinforces this dual pattern. Table 11.5 illustrates the dominant pattern for three of the classes of QWL interventions, relying here on only Human Processual interventions. In sum, V achieves acceptable levels in all 22 tests of the magnitude of the association of situational features and success rates, as Table 11.5 clearly shows; and all of the 22 cases trend in the expected direction, as is revealed by an examination of paired-comparisons related to Table 11.5. Somewhat more than 13 percent of the variance is accounted for by each of the two kinds of assessment of success.

With only three exceptions in 44 additional tests of association, the same pattern characterizes Technostructural and Combined interventions. V^2 falls in the .12–.16 range in the four relevant arrays of data. To conserve space, the relevant tables are not reprinted here.

The situation for Sociotechnical Systems differs substantially, although even there over 63 percent of the tests of association achieve acceptable levels. That is, 14 of 22 cases generate V > .20. V^2 estimates that 8.7 and 7.1 percent of the variance are explained for global and hard-criteria assessments, respectively. See Table 11.6 for details.

In sum, then, the magnitudes of the associations of success rates by classes of QWL interventions are substantial. When all four classes are aggregated, 87.5 percent of the cases achieve V > .20; and when Sociotechnical Systems are excluded, the appropriate percentage is 95.5 percent. Note that these estimates are appreciably higher than in the pilot study, perhaps as a result of the larger
### Table 11.4 Summary, Four Classes of QWL Interventions, Hard-Criteria Assessment

<table>
<thead>
<tr>
<th>Major classes of interventions</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human processual</td>
<td>76</td>
<td>69.1</td>
<td>25</td>
<td>22.7</td>
<td>6</td>
<td>5.5</td>
<td>3</td>
<td>2.7</td>
<td>110</td>
</tr>
<tr>
<td>Sociotechnical systems</td>
<td>24</td>
<td>77.4</td>
<td>4</td>
<td>12.9</td>
<td>0</td>
<td>0.0</td>
<td>3</td>
<td>9.7</td>
<td>31</td>
</tr>
<tr>
<td>Technostructural</td>
<td>81</td>
<td>66.4</td>
<td>29</td>
<td>23.8</td>
<td>7</td>
<td>5.7</td>
<td>5</td>
<td>4.1</td>
<td>122</td>
</tr>
<tr>
<td>Combined</td>
<td>36</td>
<td>70.6</td>
<td>14</td>
<td>27.5</td>
<td>1</td>
<td>1.9</td>
<td>0</td>
<td>0.0</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>217</td>
<td>69.1</td>
<td>72</td>
<td>22.9</td>
<td>14</td>
<td>4.5</td>
<td>11</td>
<td>3.5</td>
<td>314</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes of QWL interventions</th>
<th>1 Highly positive and intended effects</th>
<th>2 Balance of positive and intended effects</th>
<th>3 No appreciable effects</th>
<th>4 Negative effects</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>76</td>
<td>69.1</td>
<td>25</td>
<td>22.7</td>
<td>6</td>
<td>5.5</td>
</tr>
<tr>
<td>24</td>
<td>77.4</td>
<td>4</td>
<td>12.9</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>81</td>
<td>66.4</td>
<td>29</td>
<td>23.8</td>
<td>7</td>
<td>5.7</td>
</tr>
<tr>
<td>36</td>
<td>70.6</td>
<td>14</td>
<td>27.5</td>
<td>1</td>
<td>1.9</td>
</tr>
<tr>
<td>217</td>
<td>69.1</td>
<td>72</td>
<td>22.9</td>
<td>14</td>
<td>4.5</td>
</tr>
</tbody>
</table>
### Table 11.5  Summary, Tests of Twelve Hypotheses, Two Kinds of Assessment of Outcomes, Human Processual Interventions

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Global assessment</th>
<th>Hard-criteria assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—*</td>
<td>—*</td>
</tr>
<tr>
<td>2</td>
<td>Strong support ($V = .52$)</td>
<td>Moderate support ($V = .48$)</td>
</tr>
<tr>
<td>3</td>
<td>Weak support ($V = .21$)</td>
<td>Weak support ($V = .22$)</td>
</tr>
<tr>
<td>4</td>
<td>Weak support ($V = .23$)</td>
<td>Weak support ($V = .24$)</td>
</tr>
<tr>
<td>5</td>
<td>Weak support ($V = .28$)</td>
<td>Weak support ($V = .27$)</td>
</tr>
<tr>
<td>6</td>
<td>Moderate support ($V = .34$)</td>
<td>Moderate support ($V = .32$)</td>
</tr>
<tr>
<td>7</td>
<td>Moderate support ($V = .42$)</td>
<td>Moderate support ($V = .39$)</td>
</tr>
<tr>
<td>8</td>
<td>Weak support ($V = .23$)</td>
<td>Weak support ($V = .22$)</td>
</tr>
<tr>
<td>9</td>
<td>Strong support ($V = .65$)</td>
<td>Strong support ($V = .62$)</td>
</tr>
<tr>
<td>10</td>
<td>Moderate support ($V = .39$)</td>
<td>Moderate support ($V = .37$)</td>
</tr>
<tr>
<td>11</td>
<td>Weak support ($V = .24$)</td>
<td>Moderate support ($V = .32$)</td>
</tr>
<tr>
<td>12</td>
<td>Weak support ($V = .31$)</td>
<td>Weak support ($V = .28$)</td>
</tr>
</tbody>
</table>

* Analysis is not possible in this case because one or more of the cells contains no cases.

### Table 11.6  Summary, Tests of Twelve Hypotheses, Two Kinds of Assessment of Outcomes, Sociotechnical Systems Interventions Only

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Global assessment</th>
<th>Hard-criteria assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—*</td>
<td>—*</td>
</tr>
<tr>
<td>2</td>
<td>Moderate support ($V = .33$)</td>
<td>Moderate support ($V = .31$)</td>
</tr>
<tr>
<td>3</td>
<td>Rejected ($V = .18$)</td>
<td>Rejected ($V = .14$)</td>
</tr>
<tr>
<td>4</td>
<td>Rejected ($V = .16$)</td>
<td>Rejected ($V = .14$)</td>
</tr>
<tr>
<td>5</td>
<td>Rejected ($V = .18$)</td>
<td>Rejected ($V = .17$)</td>
</tr>
<tr>
<td>6</td>
<td>Moderate support ($V = .32$)</td>
<td>Moderate support ($V = .33$)</td>
</tr>
<tr>
<td>7</td>
<td>Moderate support ($V = .35$)</td>
<td>Rejected ($V = .18$)</td>
</tr>
<tr>
<td>8</td>
<td>Weak support ($V = .22$)</td>
<td>Weak support ($V = .24$)</td>
</tr>
<tr>
<td>9</td>
<td>Moderate support ($V = .43$)</td>
<td>Moderate support ($V = .42$)</td>
</tr>
<tr>
<td>10</td>
<td>Moderate support ($V = .36$)</td>
<td>Moderate support ($V = .34$)</td>
</tr>
<tr>
<td>11</td>
<td>Weak support ($V = .21$)</td>
<td>Rejected ($V = .19$)</td>
</tr>
<tr>
<td>12</td>
<td>Moderate support ($V = .36$)</td>
<td>Moderate support ($V = .34$)</td>
</tr>
</tbody>
</table>

* Analysis is not possible in this case because one or more of the cells contains no cases.
population here of QWL applications. However, note that despite increased size of the QWL sample, Hypothesis 1 still could not be tested.

**Why Situational Features Differ by Classes of Intervention?**

The analysis of situational features by the four classes of QWL applications distinguishes this analysis from all other attempts to test the common wisdom about situational determinants of planned change (e.g., Dunn and Swierczek, 1977), and this uniqueness encourages speculation about why the classes of QWL intervention have different profiles of situational guides for action. Let us first review the major findings by classes of interventions, and then speculate about the differences. The deviant class (sociotechnical) will be considered last.

**Human Processual Interventions**

QWL applications classified as Human Processual attain $V > .20$ in all 22 cases, and the appropriate paired-comparisons are consistently in the expected directions.

**Technostructural Interventions**

This class basically supports the rationales summarized in Appendix 11.A. Thus, the 22 tests reject only one hypothesis—to be specific, 4 for the hard-criteria, which proposes the supposedly greater success rates of QWL interventions in economic or profit-based organizations.

Why this deviant case? Most simply, one can note that $V$ in this case equals .19, which is close to our convention for assigning a “support” vs. “rejected” label. More enterprisingly, Technostructural interventions in the public sector may occur in those arenas more susceptible to economic analysis. There, reasonably, the supposed advantages of the business sector in making such estimates may be attenuated, or even largely absent.

**Combined Interventions**

Again, this class of QWL interventions basically supports the rationales detailed in Appendix A. Specifically, 20 of the 22 tests achieve $V > .20$.

The deviant cases here are hypotheses 3 and 4, for hard criteria. These hypotheses relate to two versions of the basic public versus business comparison, with the operational definition of 4 being more detailed than 3. Several explanations could apply here, with the simplest one pointing to the mixed record of efforts to establish the often-argued greater receptivity of business to planned change. The presumed advantage does not exist in some surveys of applications, and differences tend to have small magnitudes even when business sites have higher success rates (e.g., Golembiewski, 1998).
Sociotechnical Systems

Relevant tests of the twelve hypotheses show the sharpest record of rejection of the ideation on which the present tests of association are based; but even here a general pattern supporting the hypotheses exists. Thus, eight of the twenty-two tests of situational hypotheses are rejected: 3, 4, and 5 for both global and hard-criteria assessments, and 7 and 11 for hard criteria only.

The cases for the deviant status of 3, 4, and 5 seem clear enough. As for the first two hypotheses, which emphasize public versus business sectors, one can guess that public agencies employing STS designs face worksite problems very much like business counterparts at operating levels, where performance measures are similarly applicable. Hence, the reasonableness of a “Rejected” status for 3 and 4 in Table 11.7. As for 5, which argues that change is easier where labor unions do not exist, STS adoptions again may favor similar situations in business and government—as in cooperative agreements between labor and management that are common facilitators of STS applications. Or, alternatively or conjointly, the issue may be increasingly the different features of labor unions rather than their mere existence/absence.

As for the rejection of hypotheses 7 and 11, both for hard-criteria assessments, they focus on the proposed greater efficacy of change initiatives that operate at multiple levels and are introduced by internal versus external consultants.

Why might this be so? For one thing, its \( V = .19 \) is on the cusp of acceptance by present conventions. Or for 7, STS applications may emphasize low and simple levels of organizations; and/or they may have such high success estimates as to attenuate variance. Table 11.3 provides evidence supporting both hunches.

The speculation for hypothesis 11 may encompass the same features, in some part. More revealingly perhaps, STS interventions may require a blend of skills and experiences that are unlikely to be possessed by many internal OD consultants. Moreover, STS interventions involve fundamental transformations in structures, cultures, and behaviors, and internal consultants may be too much associated with local conditions to be credible architects of their replacements.

**DISCUSSION**

Six themes summarize the results of this replication of the covariation of 12 situational features with the success of QWL applications. First, the guides or rules-of-thumb underlying QWL praxis receive little overall support that meets our standards, when undifferentiated QWL applications are used to test the twelve hypotheses. Such results imply a pessimism about the leading ideas of OD and QWL about situational features as a whole.
Second, however, the guides get substantial support when the four classes of QWL applications are distinguished. In a pilot study motivating this effort (Golembiewski and Sun, 1990d), each class of applications had a somewhat distinct profile—most guides were associated with successful QWL outcomes, but cases-in-opposition tended to differ for each class. The present ability to break out four classes of applications rests on the larger size of the QWL population, and distinguishes this study from its comparable predecessors (e.g., Dunn and Swierczek, 1977; Golembiewski and Sun, 1990d). This capacity permits acknowledging the value of the major guides for praxis underlying QWL and OD, while it also highlights the need for fine-tuning those guides relative to some differences between the classes of interventions. In the present study, one class of designs differs substantially from the other three, even while sharing some situational features with them.

Third, the individual guides account for noteworthy proportions of the variance in outcomes, when the panel of QWL applications is separated into four classes of interventions. A separate analysis is necessary to assess the effects of the guides in explaining the variation in QWL applications, when the situational features are considered in combination. This analysis constitutes a major item in our future research agenda.

Fourth, the present panel of QWL applications cannot claim ideal status. Directly, the high rates of successful outcomes here may involve an artifact—for example, this panel contains only a small proportion of outcomes rated as “no appreciable effects” or “negative effects.” Obviously, this pattern contains much good news but, at the same time, the attenuation in the range of outcomes requires tentativeness in interpreting results and may in turn reflect the limitations of the present search for QWL applications. Illustratively, only 13 percent of the pilot panel of applications came from unpublished sources, which clearly underrepresents them; and that proportion has diminished in the enhanced panel.

However, some progress has been made beyond the panel of applications in the pilot study. The present panel includes more applications from developing countries, which contributes to the representativeness of the panel as well as to concerns about the generalizability of findings. Moreover, this panel includes a greater incidence of Human Processual interventions than the pilot study. The same is true of QWL applications in unionized settings.

Fifth, other situational variables could have been added to this analysis. These enrichers include profit (loss) estimates, and other bottom-line measures. In the present search, of course, we had to take what the literature gave us. Future studies stand better advised, to be sure.

Sixth, separate analyses establish that results of the pilot study probably are not artifactual in several major senses, and a replication of these findings
with the population $N = 314$ seems useful. Thus, high success rates in the pilot do not seem to reflect low levels of methodological and design rigor; in short, little evidence supports a positive-findings bias in QWL evaluations (Golembiewski and Sun, 1989, 1990a). See also Chapter 3.

Moreover, although rigor of evaluative studies tended to increase over time in the pilot study, the effect is not a dominant one; and, in any case, success rates seem to remain essentially constant over time and by classes of interventions, whatever the degree of rigor (Golembiewski and Sun, 1990c). Finally, for present purposes, the success rates of published vs. unpublished applications did not differ significantly in the pilot study (Golembiewski and Sun, 1990a), even though it seems prudent not to simply accept the present panel of evaluations as is.

Briefly, then, two summary views seem justifiable on the basis of the present study. Thus, QWL approaches—a technology-cum-values—seems to generate very attractive consequences, on balance. Moreover, these success rates can be influenced positively, if at times marginally, by a knowledge of supportive situational features, as well as by distinguishing classes of QWL applications.

In sum, Irony III suggests direct ways of heightening OD success rates by greater clarity about situational associations with success.

NOTES

1. Table 11.2 involves 38 possible paired-comparisons for the situational variables, given the absence comparisons, in each of the success categories, of mutual benefit organizations in our panel. Support for each of the 12 situational hypotheses requires that outcomes should be rated most favorable for the coding category listed first in Table 11.2, less favorable for that listed second, and so on if more than two coding categories are involved.

Hence, four success rate categories are used to assess the direction of success rates for each situational feature: Category 1; Category 1 + 2; Category 3; and Category 4. Support of the expected impact of the situational variables requires highest success estimates for coding categories 1 and 2 above. For Category 1 in Table 11.2, 71.1 percent of the paired comparisons fall in a direction consistent with the twelve situational hypotheses. For Category 2, 65.8 percent of the paired-comparisons fall in the expected direction.

2. Paired-comparisons are not attempted for the breakout tables. Given the dominance of positive effects, however, the general direction of associations admits no great doubt.
APPENDIX 11.A  BRIEF RATIONALES FOR THE SITUATIONAL HYPOTHESES

Hypothesis 1
QWL tends to be less successful when initiated in an existing organization than when started from scratch (e.g., Sashkin et al., 1985).

In existing sites, it often is unfeasible—and certainly is more difficult—to revamp all elements of the organization, such as staffing or existing norms and practices. This does not hold for new plants (Walton, 1982), continues this rationale. Start-up organizations require less combatting of entrenched values and methods (Connor and Lake, 1988), and they can plan a congruent total system—selecting people who complement one another, developing policies and systems to encourage and reward desired behavior, and often involving employees in intense ways in building the organizational infrastructure consistent with QWL values (e.g., Perkins, Nieva, and Lawler, 1983). Moreover, these systematic possibilities avoid complications with a “participative island” within a larger non-participative host (e.g., Lawler, 1982).

Hypothesis 2
It is easier to develop a successful QWL program in a small organization than in a medium-sized or large organization (e.g., Weisbord, 1997, pp. 63–68).

Few counter-rationales exist in the literature concerning this second situational feature, so we can be bold here. In a small organization, the employees are often said to have the benefit of features—e.g., optimum communication, a sense of all personalities, fuller participation, and an open atmosphere—that facilitate trust. These factors are all conducive to the success of QWL efforts.

Oppositely, most observers see large organizations as inhibiting the kind of openness and informality which can exist “naturally” in a small organization. Why? Large size can create problems of coordination and communication, with the consequences that multiple functions and numerous hierarchical levels distance action-taking from decision-making, and especially in traditional structures.

Hypothesis 3
QWL efforts in private organizations are generally more effective than those in bureaucratized or public-sector organizations (e.g., Devine, 1991; Weisbord, 1997).

The classic argument derives from von Mises, and gets contemporary practical expression in Reaganesque rhetoric and policies. von Mises (1969) argues that public bureaucracies by their nature are less responsive to change than private organizations because (for example) the former follow rigid rules established by
law, lack economic incentives, and suffer from a low-energy “habit background” (Golembiewski, 1985). Indeed, some observers propose that planned change is all but impossible in the public sector (e.g., Giblin, 1976).

A similar convincing case has been made against bureaucratized organizations of all kinds—whose major subunits are seen as separate “smokestacks” of their several functions (e.g., Ashkenas, Ulrich, Jick, and Kerr, 1995).

In hypothetical contrast, the common view holds that businesses have dual advantages. They can respond more flexibly to changes in market-driven conditions, and they have a clear motivation in doing so.

Note that a respectable counter-emphasis has been developing. Basically, it proposes that public agencies are not that peas-in-a-pod (e.g., Kettl, 1993; Wilson, 1987).

**Hypothesis 4**

QWL efforts in economic or profit-based organizations will be more successful than efforts carried out in other types of organizations (e.g., Devine, 1991).

Hypothesis 4 is both broader and more specific than its predecessor, and distinguishes four types of organizations: economic, service, commonweal, and mutual benefit. Economic organizations include industrial and manufacturing firms, retail establishments, insurance companies, banks, and other commercial organizations providing goods and services to clients for money or payment in kind. Service organizations—e.g., public hospitals, schools, voluntary agencies—provide services to clients without direct monetary payment or payment in hand. Commonweal organizations are government agencies (e.g., national, state, and local government departments) that provide service to the public-at-large, usually without charge. Finally, mutual benefit organizations (e.g., political parties, unions, professional associations, churches, and clubs) provide members with opportunities for social interaction, mutual support, and symbolic identification without direct monetary payment.

Typical expectations encourage the view that QWL outcomes will vary with the market sensitivity of the four types of organizations. Thus, economic organizations are often said to be most receptive to change (e.g., Porter, Lawler, and Hackman, 1975). Service, commonweal, and mutual benefit organizations should host QWL efforts having increasingly positive outcomes, if the common wisdom applies.

**Hypothesis 5**

QWL programs have a greater likelihood of success in nonunion settings (e.g., Ashkenas and Jick, 1982).
This fifth hypothesis reflects a dominant but arguable rationale. Unions complicate change, goes the majority view—as by requiring additional consensus-seeking efforts, affecting the preconditions for change, or limiting the nature of change. The minority proposes that unions encourage productive operations, as by putting pressure on inefficient managements (e.g., Driscoll, 1981).

This much having been said, research on questions related to OD applications in unions remains rare (e.g., Bolton, 1986). Although the professional talk is generally uninhibited, the state of research clearly encourages tentativeness.

**Hypothesis 6**

Multiple QWL interventions are more likely to increase productivity and employee well-being than a single intervention (e.g., Quinn and McGrath, 1982, Robertson, Roberts and Porras, 1993).

Directly, this hypothesis proposes that QWL efforts employing only a single intervention cannot address systemic elements. Where inappropriate role definition as well as poor interpersonal dynamics exist, for instance. Consequently, a change program using interaction-centered team-building probably will be less effective than one using team-building plus a sociotechnical orientation.

This is an imprecise notion, obviously, for it leaves open the issue of whether multiple interventions are well matched or not, poorly or well implemented, and so on. Nonetheless, the essence of the notion frequently appears in the literature.

**Hypothesis 7**

QWL efforts that focus on more than one level in an organization will be more successful than those which focus on one level only (e.g., Robertson, Roberts, and Porras, 1993).

QWL ideation reflects a systemic orientation, on definite balance. Consequently, although some researchers insist that change championed at the highest management level has the greatest chance of success, most argue that organization interdependencies give a definite priority to multi-level designs.

Hence, “multiple level” codings on Hypothesis 7 are expected to be associated with the most positive QWL outcomes and, beyond that, the four other coding categories in Table 11.2 are arrayed in hierarchical order—managers, staff, first-line supervisors, and line employees. This pattern of assignment reflects the general advice in planned change: that programs should “begin from the top,” so as to avoid situations in which learning or change at lower levels is punished by organizational superiors who are not on board.
Hypothesis 8

QWL programs that are introduced throughout the whole organization are more successful than programs limited to part of the organization (e.g., Bullock, 1986; Robertson, Roberts, and Porras, 1993).

From a somewhat different perspective than its immediate predecessor, Hypothesis 8 also relates to the common wisdom that interdependent relationships among organizational elements make it unlikely that narrow-targeted change efforts will be effective (Bennis, 1983). For example, subsystemic approaches can encourage interference by people from parts of the organization who are excluded from the “participative island” (e.g., Walton, 1975, 1982), and are resentful about it.

Hypothesis 9

QWL programs that involve high levels of participation generally lead to better performance than those involving low levels of participation (e.g., Burden, 1975; Crawford, Thomas, and Fink, 1980; Dunn and Swierczek, 1977).

This reflects the basic OD and QWL gospel, and the growth of worker participation in recent years evidences its common benefits for both employer and employee. Directly, participation can positively impact a person’s cognitive, affective, and behavioral responses to task requirements (e.g., Latham, Steele, and Saari, 1982). Hence, substantial participation can improve morale, productivity, financial performance, and quality, as well as lower costs (e.g., Levine, 1984).

Hypothesis 10

QWL programs introduced by both internal and external consultations are more likely to generate intended effects than QWL programs introduced by either one or the other (e.g., Gluckstern and Packard, 1977).

French and Bell (1978, and in all subsequent editions through 1995) note that using a change agent is essential, and much attention has been directed at the role. “Internal” change-agents are employees of the host organization; and “externals” are under contract from consulting firms, universities, and so on.

The general opinion in QWL circles seems quite definite. The underlying rationale, reinforced by some research, emphasizes that internal and external consulting roles contain potentially reinforcing dissimilarities. Internals specialize in knowledge about “fit” of QWL activities to the local culture (Huse, 1975). And external consultants can inhibit the reasonable tendency of internals to “go native” (e.g., Browne, Cotton, and Golembiewski, 1977).
Hypothesis 11
QWL programs introduced with the help of internal consultants are more likely to be successful than QWL programs introduced by outside consultants (e.g., Franklin, 1976).

Hypothesis 11 is a subclass of its predecessor, and discretely focuses on how the two different role types impact the success of QWL programs, if the choice is either/or.

The dominant opinion adds the qualification that two phases of QWL interventions need to be distinguished—early and late. In early phases, observers often note that organizations resist change and hence internal consultants can facilitate entry and serve as a bridge. Therefore, early reliance on the internal change-agent as the local “champion” seems indicated (e.g., Gavin and McPhail, 1978). Beyond these important early stages, continues the rationale qualifying Hypothesis 11, external change-agents can bring objectivity and neutrality to a situation where vested interests tend to dominate (Connor and Lake, 1988).

Hypothesis 12
The longer the system exposure to QWL, the greater the change in intended directions (e.g., Beckhard, 1975).

Virtually all QWL theorists believe that significant change takes substantial time (e.g., French and Bell, 1995), although “substantial” has different meanings. Thus, Maslow (1965) argues that “real change” in organizations may take up to 50 years, whereas Likert (1967) requires only 5 years. In order to test Hypothesis 12, this study codes all 314 cases in one of five categories assessing the duration of direct QWL activities: 1–6 months, 7–12 months, 13–18 months, 19–24 months, and longer than 24 months. This terminal category is convenient only, and especially because the present population contains only a small minority of cases that extend their coverage of effects even that long.

REFERENCES


Illustrating Large-System Change in Business
Detailing a Design for Strategic Planning and its Effects

From many salient perspectives—perhaps from the most salient perspective—the contributions of OD are perhaps best reckoned in the effects involving large-system change. Let us accept this proposition for the sake of the present argument, although that proposition is too simple by half. In any case, to honor the proposition, four chapters—12 through 15—are devoted to different aspects of large-system change, starting with the present emphasis on strategic change in business.

Giving the considerable concession above, however, is no indication that the store will be given away. That is, virtually all of the OD literature agrees that large-system change is rare (e.g., Bunker, 1997), but even that consensus is seriously overstated. That is, literally from the start of OD, large-system change has been central in both conceptualization as well as application. Thus, perhaps the landmark book in planned change (Bradford, Gibb, and Benne, 1964) took great pains to lodge its basic conceptual targets in very large-system dynamics. And early practical applications took those conceptual directions to heart (e.g., Marrow, Bowers, and Seashore, 1967; Golembiewski, 1972; Harmon, 1975).

To be sure, there was an interregnum decade or two following this start-up period of 10 years plus. So if you were not a clear and close reader of OD’s early history, you might be convinced that OD was all small-groups dynamics—team-building, various limited applications of sensitivity training, and so on.

Although the opinion about the neglect of large-system dynamics rested on a shaky foundation, it tended to persist into the 1990s. Since the Organization
Development Institute inaugurated an award for large-scale applications, worldwide, there has been a steady supply of candidates for the awards—often of sufficient scope and quality to recommend multiple awards for multi-year projects in macro-contexts. The awards during the first four years of awards went not only to applications in typical North American contexts but also to large-system change in business and government in loci such as Zimbabwe, India, and Germany plus multinational settings.

The ODI criteria are so strict that qualifying applications can be taken to be a small proportion of the relevant work. Among other features, qualifying programs of change have to be supported by credible evaluations of reliability and validity, have strong public executive support, an explicit rooting in OD values (Gellermann, Frankel, and Ladenson, 1991), and evidence of acceptance or diffusion beyond the site of original application.

**A SPECIFIC TARGET**

This chapter deals with a strategic planning (SP) exercise guided by Organization Development (OD) values and approaches. Three initial sections are conceptual, and introduce the uses of SP ↔ OD in integration: one section details the advantages of the SP concept as well as its common limitations in practice; a second section focuses on the details and theoretical justification of a worldwide SP effort in a multinational firm, including a sketch of methods for estimating effects of the design; and the third major section presents a rationale for linking OD with SP in the present fashion. A concluding theme in this chapter deals with the results of an application of the design and its underlying rationale.

This précis profits from some early elaboration.* The values of effective strategic planning (SP) have received stereophonic support. Decades ago, organizational growth motivated SP’s development and diffusion to help direct what threatened to get out of hand due to the growth of government and business. More recently, the usefulness of effective strategic planning has been reinforced by its relevance for coping with downsizing and doing more with less, as well as for moving into the era of the worldwide marketplace (e.g., Jelinek and Litterer, 1988).

Both now and then, however, problems bedeviled applications of SP concepts. On the demand side, expectations often were lofty—indeed, often far beyond reasonable reach (e.g., Golembiewski and Scott, 1990). On the supply side, SP presented an attractive package of analytic tools and orientations but, as Ansoff notes (1984), often with only vague appreciation of the centrality of implementation.

---

So this chapter seeks to blend a strategic model with tactical means and guiding values. In essence, OD values and approaches are presented in two aspects—as useful for maximizing the advantages of strategic planning (SP) concepts, and as capable of minimizing their limitations. The first two major sections evaluate SP conceptually and practically, and then review SP/OD linkages; and a third section describes an OD design for strategic planning. A final section evaluates the consequences of the design at both macro and micro levels.

A CONCEPTUAL CONTEXT FOR IRONY V

How does SP contribute to OD at its present stage of development? Consider only a few factors. First, SP processes are central in most organizations, or should be, in all sectors—business, government, and not-for-profit. A long tradition of attention in managerial studies leaves no doubt about this very broad SP relevance (e.g., Ansoff, 1964; Bryson, 1986; and Chandler, 1962). Particular lines of SP development have been faulted, of course, but a useful SP integration seems to be emerging (e.g., Golembiewski, 2000).

Second, as many observers emphasize ever more insistently, SP is critical to OD (e.g., Jelinek and Litterer, 1988) and the crucial connections are easy to illustrate. SP deals with conceptually linking today’s organizations with tomorrow’s needs and aspirations (e.g., Beckhard, 1966). And OD will be intimately involved in developing the actual in response to the ideal. No more need be said of this crucial linkage of the conceptual to the operational.

Third, and curiously, OD has paid little attention to SP, and vice versa, as has been documented in detail (e.g., Golembiewski, 2000). But this condition is not an unalterable given, as this chapter illustrates. And now seems an appropriate time for mutual attention-paying.

Such factors promise great leverage in moving beyond Irony V. It pairs substantial OD success rates with a myopia—generally avoiding elements of large-system change like SP processes that can help guide and reinforce the culture-building and institutions-building that OD can facilitate.

As an important value-added, this chapter also permits an in situ demonstration of trinitarian change, and especially of gamma change. The distinction about types of change is a central issue in the behavioral sciences, and blends fundamental status with substantial acceptance in principle and yet general neglect in practice. See also Chapter 24 for additional details.

Overall, the discussion below details a successful application, and seeks in part to summarize what was learned in one form that might have broader applicability. At its base, however, the case tells a familiar story—how human wit and will can fashion a system of regenerative interaction that can marshall substantial forces in support of effective action planning.

Much available literature provides a context for this chapter, and it places
much of the responsibility for strategic planning “whoopses!” in the traditional
type of organization—the bureaucratic model. For a convenient overview, see
Golembiewski (2000). Although interaction gets the prime share of attention be-
low, the reader also will find much evidence of structural and policy effects. A
reorganization was effected in the firm, but the situation did not permit a detailed
study of the intended changes. In general, these changes followed the thrust of
the flow-of-work model dealt with at several points in this volume (e.g., in Chap-
ter 14).

A PRIMER ON STRATEGIC PLANNING

Strategic planning is straightforward in its essentials, diverse in its several mani-
festations, and typically byzantine in applications. This fullness cannot be ac-
corded proper treatment here; however, two emphases provide an initial scan of
the associated conceptual and practical territory. Even a circumscription of the
concept reveals important SP advantages. A following emphasis introduces sev-
eral SP features that have only of late been taken into sufficient conceptual and
practical account.

Advantages Inherent in SP Concept

Few will argue about the SP concept, in general. A typical list of SP activities
(e.g., Tichy, 1983; Hurst, 1986) implies multiple advantage of successful imple-
mentations:

   Perfecting information-generating capacities that can assess organization
   strengths/weakness as well as environmental conditions;
   Running an “environmental scan” to profile an organization’s several are-
   nas—social and political as well as economic and technological;
   Generating an “organizational scan” to assess internal characteristics—cul-
   ture, values, strengths, and weaknesses;
   Developing objectives that ideally fit organizational to environmental fea-
   tures;
   Evaluating alternatives for approaching objectives, in planned transitions;
   Settling on strategic choices;
   Organizing to implement strategic choices; and
   Mounting a review and update to “roll forward” the package of strategic
decisions.

   So compelling are these conceptual essentials that SP diffused rapidly, and
   with apparently potent effects. Thus Hurst proposed (1986, p. 5) that SP is “the
dominant management paradigm in North America . . ., the model of the manage-
ment process” (my emphasis). In addition, its diffusion is fueled by more than
Large-System Change in Business

EXHIBIT 12.1  Developments in SP Practices, or Motivators of SP ↔ OD Linkages

1. Technique or Technology ↔ Hearts and Minds
2. Balance between Experts ↔ Collaboration
3. Analysis ↔ Appropriate Interaction
4. Vague Model of Intervening Dynamics ↔ Clearer Model
5. Analytic Elegance ↔ Psychological Ownership

and attractive logic. At least in business, for example, SP organizations seem to outperform their SP-less counterparts (e.g., Dess and Davis, 1984; Robison and Pearce, 1983), or at least this was a major article of faith until quite recent years.

This SP wave was not uniform. Public-sector diffusion was delayed, in general, although an early SP form appeared there as PPBS, or Planning, Programming, Budgeting System (e.g., Golembiewski and Scott, 1990). More recently, the literature agrees that “leaders and managers of public and nonprofit organizations must be effective strategists if their organizations are to fulfill their missions and satisfy their constituents in the years ahead” (Bryson, 1988, p. xi).

Limitations of SP Practices

Over time, however, consciousness has grown about significant SP down-sides. Six examples illustrate the fuller catalog of limitations, which Exhibit 12.1 characterizes as developmental continua.

In general, SP practice had its major shortfalls in not exploiting the higher state of both extreme of each continuum. To illustrate, techniques or technology got more telling attention than men’s hearts and minds, with the latter nonetheless being the more central to SP success.

SP as Technique or Technology

In large part because virtually all of its key practitioners were specialists in economics and finance, as well as because SP was propelled by the development of various statistical techniques and was associated with early high-speed computers, early SP approaches had a definite rational-technical bias. Indeed, that bias was all but overwhelming. In the federal service, multiple millions were spent for training in the techniques of PPBS, but only peanuts went into implementation—for example, in building skills required by applications, and in developing appropriate socio-emotional contexts for specific applications (Golembiewski and...
Specifically, regenerative interaction can be critical in SP start-ups, but SP seldom received the necessary training, especially in the earlier days (e.g., Golembiewski and Scott, 1990).

However, the realization has grown that, in basic ways, SP is not “on the shelf” but in the hearts, minds, and motivations of people and groups (e.g., Ansoff, 1984). After the first blushes of enthusiasm as to what technology or technique could do, recognition grew that effective strategic planning rests fundamentally on insight, intuition, the building and maintenance of coalitions in trusting relationships, and even leaps of faith.

Extreme reactions to this gap often have a Luddite character (e.g., Peters and Waterman, 1982, esp. pp. 29–54), but the point can be expressed in balanced ways. For example, Vaill (1981) proposes that “policy analysis” needs to be counterbalanced by a “new field”—he labels it “analysis policy,” or the “study of when and why you model.” He explains (1981, p. 85): “In our analytical computer-and-model-oriented world, it is understandable that strategic planners should search constantly for techniques which reduce the ambiguity of the data, which will make the options more comparable and the risks more easily assessed. Yet there is little attention given to the nature of the ‘impulse to model and to quantify,’ by Vaill or anyone else.

SP and Balance Between Experts/Collaboration

Relatedly, SP ideology has come to realize that reliance on experts—planners, mathematical modelers, and so on—must be balanced by the development of collaborative relationships among numerous organization members. The balance at first inclined definitely toward experts, but often at great costs. Keller articulated the main point in a direct way: “You know, the SP process was a technical success, but the plan never got acted on” (e.g., Keller, 1983, pp. 121-139).

A growing appreciation now exists of the primacy of collaboration over the technical sophistication of planning processes and techniques. SP should be “collaborative” in kaleidoscopic senses, but perhaps the dominant one relates to the quality of interaction. As Vaill observes (1981, p. 86): “Issues of communication, trust, ‘politics,’ leader-member relations, etc., arise constantly. Openness is an especially delicate matter for the players are usually powerful, articulate individuals who may feel ‘baring one’s soul’ is unseemly for one in their position.”

The points of access for OD into SP processes are both obvious and numerous.

SP and Appropriate Interaction

Appropriate interaction in SP applications legitimizes and facilitates an aroused awareness, in effect, and also serves to buffer the powerful centrifugal forces almost always unleashed. This interaction can help release information and build
commitment, as well as inspire the ownership of the SP as well as the consequences and exigencies that will support or assail it. This constitutes a crucial willingness “to see it through.” Rough and unanticipated patches are all but inevitable and, for example, even modest trust can be very helpful in practice.

SP processes not only typically fail to induce this special regenerative quality of interaction but, worse still, often induce powerful contrary forces. This does not exactly happen by design, but common practices reinforce such a paradoxical consequence. Space here permits attention to only three such common orientations or practices that often contribute to sub optimization in SP processes which, of course, SPers propose to maximize.

Paramountly, perhaps, SP approaches in practice typically reflect a top-down mentality—as the way an executive “we” can suitably constrain a recalcitrant “they.” This can induce determined attack/defend postures, nowhere better illustrated than in the abortive experience of President Johnson’s efforts to impose PPBS on a culturally unprepared and often suspicious federal bureaucracy. PPBS provided a costly learning experience about SP imitations, and in some very curious ways (Golembiewski and Scott, 1990). In this chapter, also, OD values and approaches are shown to powerfully reinforce SP processes.

In addition, SP processes often pursue consensus-building by what may be called enemy-making, for want of a better label. This balkanization can occur overtly, as SP processes feature divide-and-conquer strategies that encourage subunits to develop cohesiveness in opposition to one another. Most often, the effect results from early SP processes being confined in parallel but separate tracks, especially those within the several functional departments of a traditional structure. Suborganization consensuses develop early in the planning as data percolate up several separate hierarchies. Reinforced by practical or convenient tendencies for each functional unit to circle its own wagons, the stage is set for zero-sum competition between units when executives later bargain to resolve local preferences-in-opposition after they have hardened. Blunting these dour tendencies was the core motivator of President Johnson’s commitment to PPBS, for example, because he apparently interpreted the approach as permitting a settling of such issues at the very top as well as their centralized imposition at lower levels before subconsensuses developed.

Such subunit loyalties and identifications have long been recognized as unintended consequences of the very act of structuring work and delegating in bureaucratic ways (e.g., Gouldner, 1954), and more broadly as major awkward consequences of the bureaucratic model (March and Simon, 1958, pp. 36–47.) Subunit loyalties can make a mockery of SP, as when its final product becomes the simple aggregate of the “wish lists” of the several major units of an organization. This attempts to “have” a strategic plan while finessing its essence—the difficult and deliberate choice-making between desired or desirable options, which also preserves some workable coalition.
As a final perspective on these suborganizational dynamics, they seem even more powerful (if subtler) when once-bureaucratic organizations adopt nontraditional structures such as the divisional model (Chandler, 1962). Functional structures departmentalize around the “parts” of work—human resources, research, manufacturing, and so on—and their subsystemic fragmentation is well known. Absent structural change, the common remedy is some substantial degree of centralization, for only integrative management—at the top of an organization, or very near it—can make reasonable decisions about the “whole.” This centralization is consistent with early SP reliance on the experts in central staff capacities advising top management, and in effect seeks to overcome we/they dynamics. Of course, the sub units are self-interested in making decisions for their particularistic good.

A divisional model gives each operating subunit all or many of the resources necessary to “run its own business,” and the primary departmentalization is organized around more or less total flows of work—around products, for example (e.g., Chandler, 1962). This solves some problems but creates generally unacknowledged others, as far as SP is concerned. For example, divisional executives have both the motivation and means to urge overhead executives: “You hired me to run it, so let me run it.” Hence the frustration of senior executives monitoring clusters of divisions. They know they need to let divisions run in their own ways, in general, yet they may wish with every fiber of their being to jerk on the reins in specific cases (e.g., Mastenbroeck, 2000). Here in this chapter, and virtually everywhere, SP faces special challenges because very subtle and significant dynamics are engaged.

SP AND ITS BLACK BOX FOR IMPLEMENTATION

Put more directly, the power of SP techniques and technology is typically not connected to the proposed attractive outcomes by any reasonably discrete model of intervening dynamics. A kind of black box is allowed to make do, in general, in the absence of a model specifying appropriate interpersonal and group dynamics.

This characterization of SP provides an obvious set of linkages, both strategic and tactical, for SP process. Two points of elaboration add weight to this conclusion.

SP and Neglect of Psychological Ownership

As one specific consequence of the general lack of attention to the dynamics of implementation, early SP developed a kind of paper fetish—too much focus on technique and an attractively packaged plan, and too little concern for implementation. This bias proved consequential. Michael Naylor, then General Director of Corporate Strategic Planning at General Motors, briefly articulates the common
consequence of this bias: “Strategy fails because of lack of ownership” (quoted in Poza, 1987, p. 7); the concomitant neglect of psychological commitment often led to undercut implementation.

Patently, a strategic plan that does not get reflected in action is neither a useful plan nor strategic. An effective SP is not a document or an offsite meeting, but a strong commitment by all relevant stakeholders to a short game plan, and especially to rolling adjustments to realities over the life of the plan. SP processes should be seen, then, in the critical sense as involving the mobilization of supporting attitudes and the wills of multiple stakeholders at several levels of authority and responsibility. Given that support, there is almost no such thing as a “bad strategic plan.” SP is one game not decided by points for style.

**SP and Closed Processes**

Typical SP processes do little to mobilize a collective massing of opinion and will, especially early in the game; and the failure to legitimate open processes misses some real opportunities. Norms and attitudes developed or reinforced early can serve as templates for later decision-making, and, in turn, can help avoid later wasted motion as well as conflict. In terms introduced earlier, SP application seldom target the fateful movement degenerative → regenerative interaction.

In fact, basic SP processes can also intentionally forfeit such advantages. SP processes often percolate in isolation—either within subunits or reserved to top levels—and this can obfuscate with narrow self-interests what clear systemic objectives could help avoid. Moreover, more mystery than knowledge often surrounds the making of the key SP decisions, which helps energize the rumor mill and can have varied (but usually troublesome) effects on subsequent implementation. Executives may even make a fetish of secrecy as a competitive advantage. This exacerbates “normal” tendencies toward closedness, and legitimizes them.

The combined impact on implementation of these closed processes can be severe. Thus key decisions tend to get made late and narrowly—in one-on-one encounters, or in small gaggles of the elite—and the network is widened only in subsequent rounds of attempted “selling.” This serves executive needs for secrecy, and perhaps creates a useful sense of localized in-ness, but at great systemic cost. Poza (1987, p. 7) highlights this central cost: for example, “middle managers, supervisory personnel and salespeople . . . report having less influence . . . They also find themselves less sure of how appropriately they are dealing with the competition. They feel less ownership in the business and its future, which often results in inaction and extreme caution.”

Public forums are seldom used, although examples exist of the value of diffusing relevant knowledge and building commitment in this way (e.g., Keller, 1983, pp. 44–58). Perhaps, basically, an open approach reduces the coercion or manipulation often necessary to “sell” an SP package to a broad “them.” Appro-
priately, Vaill notes (1981, p. 89) that a “strategic consciousness,” perhaps above all else, “is social.” SP is “done with others, not alone on a mountaintop or in a closet.”

**OD LINKAGES WITH STRATEGIC PLANNING**

Here, let us get specific about what has been more suggested than established in earlier discussion. Recently, growing numbers of observers emphasize the need to relate OD activities with strategic planning, both conceptually and in operating terms. The proposed integration rests on a core insight. OD values and approaches can help moderate major limitations in SP practices and orientations such as those just previewed. This increases the chances of exploiting the advantages of strategic planning and its patent relevance to contemporary organization challenges.

The SP ↔ OD literature is sampled here. Attention is first directed at several senses in which OD can, or more often why it should, become strategic. Later, attention is directed at alternative OD designs for moving toward that goal. Immediately attention goes to three aspects of OD’s readiness for SP. The focus, in turn, falls on:

- OD as maturing;
- OD as technical helper; and
- OD as strategic actor and action.

**OD as Maturing**

A first variant emphasizes SP’s developmental or evolutionary value for OD. Directly, OD has focused on the small group and interpersonal relationships—not exclusively, of course, but decisively. In this view, OD’s maturation requires encompassing more macro-features, and strategic planning constitutes a convenient target for early inclusion as OD extends its comprehensiveness to include organizational goals and structures as well as regenerative face-to-face interaction (e.g., Golembiewski, 1979, vol. 2).

**OD as Technical Helper**

A second variant stresses OD’s facilitative role in SP. For example, Buller (1988, p. 42) proposes OD as the answer to these common criticisms of SP. He observes: “Strategic planning does little to change the organization’s direction because, too often, its implementation ignores the many structural and behavioral forces that resist change. As a result, many firms with apparently well-conceived strategies fail to budge from their time-honored courses.” Hence the appropriateness of a marriage of OD values and approaches with
SP’s analytic technology. Buller (1988, p. 43) sees six primary sets of activities that can enhance SP. In sum, OD can help:

- Assess and develop an organization’s readiness for change;
- Facilitate the SP process;
- Implement strategy;
- Create the conditions for successful mergers and acquisitions;
- Manage organizational decline; and
- Develop leadership skills.

In these several senses, OD values and approaches relate to critical managerial challenges, skills, and competencies.

Two factors especially reinforce Buller’s recommendation for an OD ↔ SP blend. To explain, not only do SP activities burgeon—perhaps most visibly in higher education (e.g., Keller, 1983)—but SP practice today definitely seeks collaborative models. As several observers note (Johnson, Pogoncheff, and Sherrett, 1989, p. 69):

Rather than have the plan for the organization developed by a small group of planning experts, current models have emphasized broad participation by members of the organization. This broader participation has the advantage of utilizing the expertise as well as the commitment of those persons who will be charged with the implementing of the plan.

Hence the relevance of OD values and approaches for SP. OD is the major game in town when it comes to inducing participation, involvement, and commitment. For a long time, indeed, it was the only game, but success has induced some competition as to labels and claims about parenthood (e.g., Golembiewski, 1989, pp. 9–11).

At times, this second variant appears tinged with regret that OD has been its own worst enemy. Thus, Jelinek and Litterer (1988, p. 135) emphasize that “facilitation skills and change management [not only are] at the heart of OD” but are central not only in SP but also vital when it comes to survival in our global economic marketplace. Managers, they conclude, “have never needed OD skills more.” Echoing the first variant while transcending it, these observers also note pointedly that OD’s traditional focus on

- discrete small groups and its . . . indifference to organizational goals
- has often alienated those who need OD most. In the context of survival struggles and urgent needs for integration across organizational units,
- OD’s focuses and limitations damage its legitimacy. (p. 135)

**OD as Strategic Actor and Action**

The last of the present short list of three extends OD’s reach-and-grasp in proposing that OD is strategic actor and action rather than a mere helper. Two versions
will be isolated for present purposes, although a longer list could be developed with little effort.

Culture and Influencing Values

For some, OD is strategic actor and action in critically linked senses associated with infusing organizations with values and with other cultural products. To these observers, that is, OD constitutes the major carrier of values providing the basic content for the organizational cultures that will permit us to do more and better with less, which seems our near-term heritage and perhaps our distant future as well. Numerous observers (e.g., Deal and Kennedy, 1982; Peters and Waterman, 1982) have drawn attention to the criticality of “tight cultures” which direct but do not stifle people in organizations; and in their more perceptive moments such observers realize (e.g., Peters and Waterman, 1982, pp. 77–81) that some “tight cultures” can be more problem than solution. The question is not: Tight culture or no? The issue relates to the specific values infused into any tight culture.

In this sense, OD has a strong claim to strategic status. Not only does OD embody a set of values for guiding interaction and structure that are widely accepted (e.g., Likert, 1967) but OD also provides a technology for approaching those values in operating situations, and with high success rates to boot (e.g., Golembiewski, Proehl, and Sink, 1981, 1982; Nicholas, 1982).

Some observers attach strings to OD’s potential as both strategic actor and action. For example, Jelinek and Litterer (1988, pp. 137) allow that “sweeping opportunities” do exist, but exploiting them will require a “fundamental revitalization of OD.” This revitalization cannot be detailed here, but it includes in their eyes a new statement of professional mission, an updated understanding of how that mission will be served, and an appropriate vocabulary for communicating rooted in client concerns. Wisely, Jelinek and Litterer advise no change in basic OD values: instead, they simply recommend a new turf on which OD should be applied.

OD DESIGNS FOR SP

Not surprisingly, then, OD has produced designs appropriate for SP. The resulting effort is substantial but not directly relevant to the arena of special interest here—a multi-product line, high tech, divisionalized, worldwide business organization. In turn, two sections will sketch evidence supporting this lack of relevance.

Selected OD Designs for Strategic Planning

The literature contains three major varieties of OD designs for strategic planning. First, several designs have been developed to guide short-cycle SP efforts via OD values and approaches. For example, Poza (1987) sketches a 7-hour design
Large-System Change in Business

for “strategy setting” in organizations. The design has been used for aggregates of 30 to 100 managers having limited time and resources for planning. Ad hoc groups—each a “diagonal slice” of levels and functions—develop individual “strategic statements,” based on summaries of “strategic information” announced to them by central authorities. These statements are shared with all managers, and are amalgamated into what amounts to a consensual “strategic template” that subsequently guides action-planning by homogeneous departmental subgroups, and plans are then made for informing those not present at the 7-hour session.

Second, OD perspectives in many cases are fitted to traditional SP models, as in the detailed workbook of Pfeiffer, Goodstein, and Nolan (1986). In addition to the prescription that all interaction be guided by OD values—openness, owning, and trust—two OD add-ons distinguish their model. Thus the design moves away from the traditional reliance on expert consultants and high-level planners. Rather, in this model, all managers can be candidates for the planning team. This clearly gives a “collaborative tilt” to an SP process, as Johnson and his colleagues assert (1989, pp. 70-71), but the selection of the planning team is crucial. It may be seen as co-optation by organization members; or, worse still, those selected may be influential whose power derives from bases antithetical to OD values. Given sufficient care, however, one can select the planning team in ways that encourage collaboration (e.g., Carrigan, 1973). In addition, Pfeiffer and his associates (1986) provide for a “values audit”—a rigorous examination of an organization’s culture along with its associated values, supplemented by an analysis of the degrees-of-fit with the values of major stakeholders. Of course, these emphasis on the fit between the cultures and values of multiple stakeholders constitute staples of the OD approach.

Third, OD values and approaches are reflected in a family of short-term SP models tailored to specific arenas. Thus the Schindler-Rainman and Lippitt model (1972, 1975) focuses on community development and change in arenas featuring schools, public and private agencies, and business organizations, which can profit from cooperation in community affairs. Similar models exist for not-for-profit helping agencies (United Way of America, 1958), for religious or church groups (e.g., Harms, 1983), for educational settings (e.g., Keller, 1983), and so on. Commonly, such designs adopt the traditional SP stages or phases, and add OD features—collaboration in open systems for planning, participation by all major stakeholders, and involvement throughout the polling process. In a few words, OD perspectives and methods seek to vivify the rational-technical logic of standard SP stages or phases.

Some Design Limitations

These illustrations suggest a flurry of activity following early criticisms of OD’s puny role in SP (e.g., Vaill, 1981, esp. p. 89), and although the largest part of
the task remains undone, the achievements have been substantial. Put another way, the work sampled above serves as a narrow beachhead for more-targeted work and also highlights these important limitations of existing OD forays into strategic planning:

Almost all of the available applications of OD values and methods deal with short-cycle planning. Exceptions exist, both early (e.g., Beckhard, 1966; Marrow, Bowers and Seashore, 1967), as well as late (e.g., United Way of America, 1985), but these exceptions need supplementing. Why? Examples abound. To illustrate, the shortest full research plus development cycles in the pharmaceutical industry approximate a decade. Relatedly, the exemplars seem to incline toward what might be called “less mature” SP applications. Two earlier points encourage such a conclusion. To be specific, many applications seem basically interested in heightening commitment to a more-or-less firm plan, as contrasted with building participation and involvement in a full planning process that subsequently generates an SP product. Poza’s (1987) work has such an orientation which—although it may be both useful and desirable—differs from full-fledged applications.

The literature tends to emphasize organizations that are variously “in trouble” and hence in more-or-less desperate need for strategic planning (e.g., Jelinek and Litterer, 1988). Certainly, such organizations need caring attention. However, this condition may load the dice in favor of failure, or at least SP tends to take on the character of a “last chance.” In any case, it seems instructive to consider organizations in less-vulnerable conditions as well as those in extremis.

The available literature concerning OD in SP contains only gentle and sporadic touches of empirical analysis. Directly, much of the literature is avowedly conceptual (e.g., Pfeiffer, Goodstein, and Nolan, 1986), or even speculative. Moreover, even when a specific SP/OD design has been applied many times, outcomes get summarized in only general terms (e.g., Poza, 1987).

Relatively, care is necessary that OD not become inbred, and consequently collapse unto itself. Hence, the centrality of practical and theoretical extensions of the OD core. Such extensions there must be, for good or ill, lest OD become (or be seen as) stagnant and dated.

Even in the small minority of SP/OD applications that estimate effects in some reliable way, the typical focus is on macro-effects alone. Certainly, such estimates are consequential, and perhaps even primary, when dollars-and-cents are involved (e.g., Beckhard, 1966). Practically, this myopia is also convenient, if the applications “really work.” Theoretically,
cally, in either case, the consistent neglect of micro-dynamics is a recipe for atrophy.

Existing SP/OD applications seldom seem relevant to the “new organizational challenges” that appear all around us. Consider, for example, the rarity of OD guidance of Information Technology applications. Exceptions exist to this generalization, of course, considering the full history of OD (e.g., Beckhard, 1966; Lippitt, Langseth, and Mossop, 1985), but they remain rare exceptions.

What are these “new challenges”? Jelinek and Litterer (1989) see today’s businesses as operating in global marketplaces, involving increasingly sophisticated and powerful technologies, requiring longer-range perspectives, as well as shrinking half-lives, and facing massive and continuous retraining of work forces, among others factors. No doubt, successful organizations will be either unprecedentedly large or exceptionally aggressive niche-fillers. Both cases require much of managers and employees—greater and changing skills, higher motivation that derives from self-discipline, and enhanced commitment to programs and products whose major constant is change.

Finally, most SP/OD efforts relate to functionally structured organizations.

There, the basic departmentalization emphasizes the “pieces”—the so-called “management functions” such as manufacturing, research and development, human resources, and so on. In such case, the strategic plan becomes one of the managerial tools for encouraging some integration of the “pieces” into some reasonable “whole.” Such systems tend to be centralized, and the expert model of planning fits nicely with the line/staff model of structuring work usually prescribed for functional systems. Generally, however, such conditions were tolerable only in early SP applications. Most of today’s larger organizations are variously beyond those conditions (Golembiewski, 1989).

Life becomes far more complicated in post-bureaucratic structures, which the SP/OD literature tends to neglect, but that is the price of doing something reasonable about the issues that face bureaucratic organizations. Here, in addition to lean central staff, departmentalization focuses on “wholes”—as in the divisional model whose primary units include the full range of activities needed to research, develop, make, and market some product or service (e.g., Chandler, 1962). Peters and Waterman (1982, pp. 126–134) refer to this basic structural reform as “chunking” large systems into several smaller ones—to build commitment and loyalties, to reduce the length of communication chains, and so on. Each “chunk” can “do a whole thing.” For details, again, the interested reader also might well consult Chapter 14.
Strategic planning in post-bureaucratic structures takes on subtle features, although the realization often has come slowly or even not at all (e.g., Ansoff, 1984). Domination of planning processes by headquarters experts is reduced, in large part because determined efforts typically have been made to reassign all but lean-and-mean central cadres of staff to each division to permit them greater control over their own flow or work; the opportunities for hide-and-seek games burgeon in central/local relationships; the prevailing philosophy is to let the separate units “do their own thing,” and yet numerous exigencies require quicksilver mixtures of both independence and collaboration from the units designed to be capable of “doing it all.”

Organizing around total flows of work at once seeks to set lower-level units of organization generally free to serve their own interests, and yet requires their efficient collaboration in numerous specifics. In short, the dynamics of influence in such systems combine “let go” and “hold on,” from the perspective of top-level executives. This complicates SP in numerous particulars, patently.

ONE OD APPROACH TO STRATEGIC PLANNING

The present approach to SP ↔ OD seeks to respond to a dynamic situation in a successful multinational firm in health care, while applying to it lessons learned from the experiences sketched above. The firm approaches $1 billion in annual sales, has maintained a very high rate of growth over the previous decade, and plans on extending this trend line through the foreseeable future. The firm has in recent years adopted a basic divisional structure, with activities aggregated into several Strategic Operating Areas (or SOAs). Earlier, a functional structure dominated and line/staff principles typically applied.

Three emphases will provide guiding detail. In turn, sections below will introduce the objectives of this SP ↔ OD design, then provide details of the design, and finally outline methods for evaluating major consequences of the design.

Design Objectives

The firm’s strategic planning had been unsatisfactory to all, and the preceding year’s effort was so abrasive that the Chief Operating Officer widely vowed “that never will happen again.” Basically, mutual withholding resulted in an SP process of several escalating fits-and-starts. A few statements convey the simplified essence of this serious hide-and-seek between the two central “line” executives and their subordinates in the several SOAs, which went through several unpleasant iterations:

We need the maximum in before-tax profits you can give us, so let’s not play games.
This is absolutely the most we can do. That’s not only not enough, but you have obviously not taken us at our word.

OK, we can do more (after much posturing and many protestations). Why did we have to push so hard to get you to come up with more? And, remember, we asked for the maximum, up front.

Avoiding the associated unpleasantness motivated the SP ↔ OD design, which also was a logical extension of other OD work in the firm—with a culture statement, team-building at several levels, and so on. Exhibit 12.2 places the SP ↔ OD effort in the context of major milestone in the firm’s immediate history of moving from degenerative → regenerative interaction.

Overall, this application seeks to rely on OD values and technology to avoid the SP limitations detailed earlier. Put in positive terms, this design seeks to:

- Create a substantially public arena for SP dynamics, from a very early stage;
- Build a supportive quality of interaction, there (as well as here) called “regenerative;”
- Rely to a substantial degree on group settings—to build confidence in regenerative interaction, to provide a degree of safety (especially initially) via providing multiple forums to test attitudes and concepts, as well as to help build social support for OD values and attitudes in problem-solving;
- Avoid cohesiveness-building in one’s organization unit via making enemies in other units;

**EXHIBIT 12.2 Milestones in Host Firm**

<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>Begin what became numerous cycles to develop regenerative relationships and cultures</td>
</tr>
<tr>
<td>1986</td>
<td>“Chunk” structure into strategic operating areas (SOAs) from functional model</td>
</tr>
<tr>
<td>1987</td>
<td>Surpass $500 million in sales</td>
</tr>
<tr>
<td>April 1989</td>
<td>Worldwide strategic planning exercise, Year 1 (3 days), part-way through normal SP developmental cycle, to provide a test of, as well as reinforcement for, a supportive quality of interaction</td>
</tr>
<tr>
<td>Dec 1989</td>
<td>Planned follow-up, Year 1</td>
</tr>
<tr>
<td>Last quarter 1989 and beyond</td>
<td>Surpass $700 million in sales and get spun-out of parent corporation as separate entity with own board, etc.</td>
</tr>
<tr>
<td>April 1990</td>
<td>Worldwide strategic planning exercise, Year 2</td>
</tr>
</tbody>
</table>
Minimize reliance on planning experts, and seek to approach a consultative model; and

Build commitment for a strategic plan seen not as a cookbook but as a basic template for adapting to an unfolding reality, guided by shared values and basic operating commitments.

The “supportive quality of interaction” requires a fuller statement. This application seeks to approach regenerative interaction, and to minimize degenerative aspects where they exist or develop (Golembiewski, 1979, Vol. 1, esp. pp. 1–132). These opposed models have been referred to at several points in this volume, especially in the Introduction and Chapter 4. This reemphasis is useful here even though it risks some repetition. The two systems of interaction are combinations of four variables, viewed as a set of multiply reinforcing linkages, as Figure 12.1 shows. The central point is that effects at any point in the model will, in general, induce reinforcing changes at all points. Thus, heightened trust

![Figure 12.1](image.png)

**Figure 12.1** Two models for interaction.
will tend to reduce risk as well as heighten openness and owning. Conceptually, risk refers to the degree of objective threat, while trust estimates the degrees of confidence that one’s associates can be relied on to generate positive effects and that “things will turn out OK.” In addition, one can be open about ideas or feelings without owning them. This combination often appears in statements such as: “They, but I can’t tell you who, don’t like you.” Figure 12.1 also details some of the dour consequences that motivate avoidance of degenerative interaction, in SP as well as in life generally.

How to build toward regenerative interaction in SP processes, acknowledging that the firm had multiple experiences with that model in various settings (e.g., Golembiewski, Hilles, and Daly, 1987)? Exhibit 12.2 presents the schedule of major milestones in this SP/OD linkage, and reflects two major conclusions drawn from the OD literature. Thus, a worldwide collection of managers, assembled for the first time for any purpose, had reviewed an SP draft and other materials before the Phase I meeting. This was both perceived and advertised by top management as “an unusual act of trust,” which conceptually sought to increase the openness and owning by others and hoped to minimize the felt-risk of participating openly in SP processes. Three intents dominated the Phase I design: to provide grist for the mill in successive drafts of the plan, to provide a model for a Phase II meeting in about six months to review a fine-tuned SP draft, as well as to generate experience where fuller involvement in SP processes might be possible in a subsequent team. The wide sharing of the SP draft was envisioned as a kick-start toward inducing regenerative interaction in limited time-frames in which problems can be substantial, even formidable (e.g., Golembiewski, Proehl, and Sink, 1981, 1982). This generalization assumes that a “crisis of agreement” does not exist (Golembiewski, 1990, esp. chap. 6). In addition, fade-out effects can occur, especially if organizational realities conflict with training site designs. Hence, based on long experience (e.g., Golembiewski and Carrigan, 1970a, 1970b), Exhibit 12.2 provides for a booster shot during Phase II to minimize fade-out effects, as well as for a succeeding full SP cycle. (Massive legal and organizational changes resulted in a postponement of Phase II).

**Design Details**

These considerations motivated a two-phase SP design, whose general flow is reflected in Exhibit 12.3. The decision to implement the design was made after beginning a traditional SP cycle, which featured inputs from the divisions as well as from corporate that were assembled by a central planner. His first draft was then circulated to some 70 key officials worldwide, to begin a collaborative effort.

Exhibit 12.3 reflects the detailed senses in which this SP/OD application seeks to approach the regenerative and to avoid the degenerative. The trigger-
event is management’s unprecedented willingness to share the first-draft of a strategic plan with the full complement of key management personnel in a regenerative setting.

Operationally, Exhibit 12.4 may be viewed more specifically as having four major thrusts. These may be summarized as:

- Ice-breaking, especially for the upper management groups (item 2 in Exhibit 12.4)
- Variously testing the SP first draft (items 3, 4, 5, and 6 in Exhibit 12.4)
- Relating SP efforts to the firm’s ongoing approach to an ideal vision and its associated culture (item 7 in Exhibit 12.4)
- Emphasizing realistic market and environmental forces

Note that earlier OD efforts began inducing a detailed sense of the firm’s vision, missions, culture, and philosophy as well as tested for effects. The firm’s normative aspirations guided the development of 43 survey items dealing with nine arenas: competition, communication, balance of life, people orientation, quality, understanding the business, entrepreneurship, action orientation, and anticipated trends. Data from several administrations of the survey to limited target populations were fed back to SP participants to remind them of the public values of the firm, to provide estimates of the present realization of the vision and culture, and to permit focus on how SP processes can serve to reduce ideal/real gaps.

Two basic kinds of groupings are utilized in the design reflected in Exhibits 12.4 and 12.5, with all activities occurring within a single large ballroom so that “all the action” was, by deliberate choice, in the direct view of all. The Operating Committee (Op Com) is an intact team, consisting of: the chief executive officer,

<table>
<thead>
<tr>
<th><strong>Exhibit 12.3</strong> Schedule of Major Events</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Day 1</strong></td>
</tr>
<tr>
<td><strong>Days 15–17</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Days 197–200 (approx.)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Days 201–500 (approx.)</strong></td>
</tr>
</tbody>
</table>
**EXHIBIT 12.4** Major April Design Elements

1. Pre-work for individuals: read SP document for key issues, clarifications, and recommendations for change/refinement
2. Unfreezing: each of the *ad hoc* Upper Management (UM) groups develops a name, logo, and motto.
3. Eight UM groups discuss and prioritize item 1; 8 spokespersons report to the Operating Committee (Op Com); general discussion follows.
4. Each UM group assumes identity of a specific competitor: devises a “creative yet achievable” strategy to drive own firm “into the ground” within 5 years
   - Reports to Op Com, which chooses best strategy
5. Each UM group does a role-reversal: develop strategy to defeat competitor
   - All report to Op Com, which chooses best strategy
6. Each UM group synthesizes and prioritizes 5 changes/refinements to SP draft, based especially on items 4 and 5
   - Reports and discussion
7. Can host “culture statement and philosophy” support SP?
   - Survey data regarding ideal vs. actual gaps emphasizes:
     - Causes of gaps
     - How to reduce the gaps
     - Assignments for follow-up
8. Review of key priorities
   - CEO and COO do so publicly, for first time ever, with focus on fit of their key priorities with data from number 1–7
   - Pairs of managers from cross-UM groups do similar exercise

The chief operating officer, the heads of the several SOAs, and corporate staff officials. Eight heterogeneous *ad hoc* groups of key worldwide executives also were formed, deliberately mixing functions and SOAs. Indeed, many members of these upper management groups previously knew one another only by reputation or telephonic acquaintance.

Op Com in Phase I played several roles. At times (e.g., item 2 in Exhibit 12.5, its members had exactly the same charge as the upper management groups. At other times (e.g., item 3), Op Com worked on its own issues. Then again, Op Com at times served in an evaluative role (e.g., item 6). Also on occasion, each Op Com member sat in on the discussion of a single upper management group, later reporting back to Op Com while all others observed a fishbowl exercise (item 7).

Two motives explain the heterogeneous mix of the upper management groups. They were seen as the primary arenas in which OD values and attitudes would be tested; and mixed composition was seen as a greater challenge to group formation than intact work teams or groups representing the same SOA or staff
## Exhibit 12.5  Major Design Elements, Rationales, and Time Allocations for Phase I of Worldwide Strategic Planning

<table>
<thead>
<tr>
<th>Activities</th>
<th>Rationale</th>
<th>Time allotted (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pre-work for individuals: Read first draft of 1989 SP Identify any omitted key issues Note any needed clarifications Develop suggestions for change or refinement of SP</td>
<td>1. To share an SP with unprecedented numbers of key managers, and to legitimate by this act of trust an active, creative posture by many toward SP processes</td>
<td>1. Individual choice</td>
</tr>
<tr>
<td>2. Each of 9 working groups, chooses a name, logo, and publicly reports them (N.B. suitable baseball hats provided each team incorporates these identifiers!)</td>
<td>2. For ice-breaking and tension reduction, as well as to build early sense of identification in 8 Upper Management groups</td>
<td>2. 2.0 hours</td>
</tr>
<tr>
<td>3. 8 Upper Management groups prioritize omitted key issues, needed clarifications, and suggested changes for refinement for SP (see #1 above) Spokespersons report priorities to Op Com, in general, followed by open discussion</td>
<td>3a. To provide a first-cut at generating common understanding of SP, and to begin the process of refining the plan 3b. To provide an opportunity for Op Com to be influenced by any shared priorities 3c. To prepare for activities #4 and 5</td>
<td>3. 2 hours</td>
</tr>
<tr>
<td>4. Each Upper Management group assumes the identity of a competitor and devises a &quot;creative yet achievable&quot; strategy by which that competitor can drive the host organizations &quot;into the ground&quot; within five years 10-minute presentations in general session are made by Op Com, which acts as &quot;board of directors&quot; and selects the best strategy</td>
<td>4a. To provide stereophonic perspective on the SP—to test its assumptions and expose any vulnerabilities 4b. To assess Upper Management knowledge about competitors and markets 4c. To enhance skill and experience in taking the role of the competitive other</td>
<td>4. 3.5 hours schedule but several teams worked through a &quot;business lunch&quot;</td>
</tr>
<tr>
<td>5. Each Upper Management group does a role-reversal—it now develops a strategic attack on the competitor whose identity it had assumed in #4. 5-minute presentations in general session are made to Op Com, which selects the best strategy for attack</td>
<td>5a. To use the first draft SP against a competitor, where possible, and to invent enhancements or new strategic elements where necessary or appropriate for a specific competitor 5b. To enhance familiarity with SP in an action setting, to further test its assumptions and vulnerabilities, and to further legitimate the search for SP clarifications, refinements, and changes</td>
<td>5. 3.5 hours</td>
</tr>
</tbody>
</table>
6. Each of 8 Upper Management groups prioritize five changes or refinements to SP
   Spokespersons report in general session and discussion follows
   Op Com reacts to proposed changes or refinements

7. Estimate ideal vs. actual gaps between host “culture statements and philosophy” and existing management practices, and develop ideas for an action plan to reduce the “gap”
   Resource person summarizes the results of two recent surveys assessing the gap as perceived by executives and middle managers
   8 Upper Management teams discuss “What causes this gap?” each with an OP Com member sitting in as an observer who is coached on what should be reported to Op Com
   Op Com members meet in a “fishbowl,” hearing reports about team views of causes of the gaps, surrounded by Upper Management teams
   Each Upper Management team explores a different issue related to: “What we should do differently to reduce the gaps?”
   Teams report recommendations, along with persons or functions assigned to follow-up

8. Review of individual key priorities and objectives for 1989 by all participants
   CEO and COO share their 1989 priorities and objectives publicly with Upper Management, for the first time
   All other individuals review their 1989 objectives with a partner, how they might be aligned better with the emerging SP
   8a. To relate today with the unfolding tomorrow
   8b. To move toward a greater articulation of action with strategy
   8c. To begin demystifying key priorities and objectives, which had been narrowly shared in the past, if at all
   8d. To have two executives lead-by-example
   8e. To encourage helping relationships between peers
function. Moreover, heterogeneous groups sought to avoid the kind of suboptimization that occurs in strategic planning that features organizationally homogeneous groupings, which tend to fixate on their individual self-interests.

**Methods for Evaluating Effects**

The focus here shifts from the character to the consequences of Phase I of the SP/OD application. A pre-test and a short post-test via survey methods provides some sense of both macro- and micro-effects, and the survey content needs to be introduced, even if briefly here.

*Estimating Macro-Effects.* Thirteen items test reactions of about 70 key managers to SP dynamics—those of last year contrasted with the unfolding processes of the current year. For present purposes, six categories can be distinguished:

- Rating SP processes and outcomes
- Corporate trust level
- Success in surfacing and dealing with conflictual issues
- Success in uniting individual SOAs or functional units in support of collective goals
- Degree of knowledge of, and comfort with, key players
- Degree of personal commitment and comfort

Directly, the items tap conventional targets of SP exercises. Given the present scoring conventions, a successful Phase I would be signaled by decreases in the two items dealing with we/they dynamics between departments as well as with personal frustration and discomfort. Increases in the other eleven items also will signal a successful Phase I.

*Measuring Micro-Effects.* This study also estimates micro-effects in the nine groupings featured in the learning design, which provide some perspective on significant SP dynamics. Friedlander’s Group Behavior Inventory (1968) serves in this particular, with attention going to five of his GBI dimensions:

- Group Effectiveness
- Mutual Influence
- Personal Involvement and Participation
- Intergroup Trust
- General Evaluation of Meetings

Additional details appear in technical additions to this chapter (e.g., Golembiewski et al., 1992). However, in general, increases are expected to result in all five dimensions from a successful OD ↔ SP design.
MAJOR ELEMENTS IN A THEORETICAL RATIONALE

A kind of spiral of specificity characterizes this effort to sketch the nexus of theory underlying the design reflected above, with the major motivation coming from a surfeit of success. That is, applications of group-oriented designs tend to be so powerful as to discourage elaboration of their fundamentals. Moreover, some things have an importance that justifies their being repeated, even excessively. The present spiral of specificity has three major components.

Groups as Mediating Powerful Forces

Basically, this design uses group settings to counterbalance subsystemic tendencies, which can proliferate, if not dominate; the heterogeneous *ad hoc* upper management (UM) units are intended as systemic analogs to create and sustain personal and organizational forces toward integration in SP dynamics. This is a difficult effort, because the tendencies toward subsystemic outcomes in organizations are deeply rooted in basic organization processes. For example, see the unintended consequences of the primordial organization activities involved in delegation (e.g., March and Simon, 1958, esp. pp. 44–46).

The contest between long-entrenched, subsystemic tendencies and the *ad hoc* UM collectivities may not seem fair, but it is unavoidable and can marshal substantial counterforces. Specifically, and especially in the public setting of the design, the *ad hoc* collective needs can influence behavior both at the learning site and afterward. In sum, the behavior-regulating forces derive from such important individual and collective needs met by small groups (e.g., Golembewski, 1962, pp. 92–97; Jedlicka, 1977, pp. 26–67):

- Small groups serve as agencies through which members obtain and evaluate information about themselves and their environment.
- Small groups provide both a motivation and environment for conflict resolution and compromise via open communication about feelings and ideas.
- Small groups create aspects of reality that affect the behavior of members and even nonmembers, as in cultures and norms that provide evaluative criteria for judging aspects of reality as well as in defining what is taken to exist.
- Small groups reduce anxiety and induce senses of belonging and self-worth, as in providing protection, affection, and affiliation.
- Small groups may powerfully mediate, if not determine, a person’s self-esteem, status, and perceived prestige.

The learning design seeks to engage such forces, consistently and insistently. Initially, group significance is raised gently, even jocularly, in the choice as well as public presentation of a group name, logo, and motto. Later in the
design, the level of group development is reflected for all to see in the rounds of presentations regarding competitive strategies.

**Facilitating Group Development**

Much is known about how to facilitate group development (e.g., Zander, 1982), and the design in Exhibit 12.5, as well as supportive consulting activities, variously sought to vivify this theory-cum-experience. Consider only that the planned heightening of the cohesiveness of the UM groups proceeded on three tracks, as it were.

First, it seems that much of the variation in cohesiveness is due to prestige of task, prestige of membership, and liking of members (e.g., Golembiewski, 1962, pp. 149–170). Attendance at the SP activated the first two contributors, virtually by definition, and heightened liking often follows the intensive interaction legitimated by the SP design.

Deviant cases can exist, of course. Thus intensive interaction also can induce dislike, although apparently in a minority of cases. In addition, some past histories imply low interpersonal compatibility. All *ad hoc* groups were constituted so as to separate known antagonists. Low-trust combinations between strangers also can develop quickly (e.g., Friedlander, 1970; Zand, 1972), but this apparently occurs only in a minority of cases.

Second, as Exhibit 12.6 sketches, group cohesiveness can be heightened in several ways related to the meeting of individual needs. Paramountly, perhaps, the experience with regenerative interaction was intended to highlight the need-

---

**EXHIBIT 12.6 Guides for Facilitating Two Aspects of Group Development**

A. **Making a Group Cohesive**
   - Help members identify individual needs they may satisfy in the group.
   - Increase the number and strength of these needs.
   - Increase the group’s ability to meet these needs.
   - Determine how well the group meets those needs.
   - Help members understand how their needs are being met in the group.
   - Give members a chance to make sacrifices for the group.

B. **Creating an Active Concern About the Group’s Fate**
   - Develop standards for the group’s purposes, activities, goals, procedures, and beliefs.
   - Encourage each member to abide by these standards.
   - Stimulate a disposition among participants to experience pride and satisfaction with the group if it successfully accomplishes a challenging group task.

Large-System Change in Business

serving potential of the UM groups. Thus, the ad hoc groups provided contexts in which new information was generated and agreements developed. Norms also encouraged the public sharing of materials that individuals probably would not have released, if left to their own devices.

This experience was not simply one of having things done unto the UM groups by the design. The Operating Committee members attempted to trigger heightening cycles of regenerative interaction in several ways—by sanctioning the design, as well as revealing/representing themselves in action, perhaps especially so for the CEO and COO (e.g., item 8 in Exhibit 12.5).

Basically, high cohesiveness refers to meeting member needs—via liking of other members, prestige of membership, or relevance of group task for members’ needs. The process is reinforcing and circular. Participation in groups can meet individual needs—affective as well as instrumental; and this need-satisfying potential reinforces involvement in maintaining the group. Regenerative interaction seems to have similar consequences, on balance.

These self-heightening tendencies have direct and profound consequences. High agreement about norms will tend to develop along with high member satisfaction. Both of these outcomes are consistent with a high degree of control over member behavior, which can lead to one of two conditions—high or low performance, as in productivity.

Ideal Track of Consequences

In sum, the design attempts to create a specific pattern of associations, with “cohesiveness” being especially useful in explaining the ubiquity of group effects. “Cohesiveness” may be defined as the resultant of the forces attracting individuals to membership in a specific group minus those forces encouraging individuals to leave that group. Convenient reviews of this research (e.g., Golembiewski, 1962, pp. 149–170; Zander, 1982, pp. 4–10, 45–56) emphasize these major linkages (see Figure 12.2).

The low-performance outcome has troubled some observers, even though it seems substantially less probable, but this condition reflects the uniformity of behavior required for both high and low performance. By hypothesis, low

![Figure 12.2](image-url)
performance under high cohesiveness usually occurs as a defense—for example, as an aggressive reaction to an autocratic management. OD seeks to minimize the defensive reaction by high levels of participation and involvement, as well as by creating regenerative interaction.

The high cohesiveness induction could fail, of course, but this seemed a low going-in probability and was a consciously-taken risk. For example, Op Com’s decision to run the design may have been seen as untrustworthy—as an outright ploy, or perhaps as good intentions that would fade. Or something about the specific mixes of UM group membership might have generally inhibited the formation of high trust that would facilitate group development. As noted, care was exercised that individual UM groups be free of members who had known personal or organizational problems with one another. Beyond this precaution, maximum-mix rules prevailed: genders, functions, geographic locations, and so on were distributed throughout the ad hoc groups, as evenly as possible. More deliberate approaches to assembling high compatibility UM groups—as by the use of Schutz’s (1966) FIRO-B—were considered but not employed.

Data presented in a technical report suggest one UM group had major developmental problems, and could be seen as “resisting change” or as refractory to the design. Major contributors to this outcome seem clear. As fate had it, that collection’s membership included a high proportion of late arrivals and—perhaps relatedly—their performances as a group in the two design loops dealing with marketplace competition were not distinguished. “Real losers” in such design elements are a possibility and constitute one of the design’s risks (Golembiewski, et al., 1992).

AN IN-PROCESS SUMMARY

Thus, the stage is set for an empirical test of the SP ↔ OD linkage. This test rests on a demonstration of how OD values and approaches can contribute to several ongoing SD developmental tendencies. One design for SP/OD linkage is detailed above, and the theoretic and experiential rationale for the design is illustrated. An approach to measurement also is sketched and results are detailed.

The description of action research above introduces the design and rationale guiding the participation in strategic planning by the top 70-plus worldwide executives of a consumer products firm. The design reflects Organization Development (or OD) values and approaches with the rationale being firmly nested in group analysis and theory. The design calls for two phases. Phase I results are reported here, and various reinforcing efforts were scheduled for nine to ten months after the initial intervention. Phase II gets only brief mention here; it was delayed by major legal and organizational changes in the firm that discouraged longer-range treatment than below.

The effort is both novel and conventional, in different aspects. Thus, the
design and rationale are conventional and have numerous near-analogs whose success rates are substantial; see especially Chapter 1. Simultaneously, the use made of that design is rare, if not unique. Many observers have emphasized the need and opportunity to apply OD to strategic planning (e.g., Jelinek and Litterer, 1988), but no reports exist of efforts with the sweep and scope of the present one.

So what were the consequences of this effort to integrate OD perspectives and strategic planning (OD ↔ SP)? Four approaches to analysis contribute to a general conclusion: on balance, the design had the intended effects, and often markedly so.

The four approaches to analysis may be outlined, by way of introductory preview. First, thirteen items estimated the macro-properties of SP processes and consequences—for example, to what degree did SP deal productively with disagreements or conflicts?—for 1989 versus the previous year, which used a conventional, top-down approach. Almost all comparisons fall in the expected direction, and many also achieve statistical significance.

Second, the Group Behavior Inventory (Friedlander, 1968) assesses the micro-properties of the nine individual learning contexts in the SP design – one intact Operating Committee of twelve executives, and eight Upper Management groups created ad hoc for the design from among the other approximately 70 assembled key actors worldwide. Change is greatest for the Upper Management groups, as expected. In addition, most observations show intended change, and a substantial proportion of the intended differences attain or surpass $P < .05$.

Third, several comparisons focus on the eight Upper Management groups, to sharpen the overall analysis of micro-effects. Design expectations are substantially met.

Fourth, analysis indicates a profound change in how participants viewed their world, before and after the SP design, considering both the thirteen macro-feature items and those seventeen items assessing the micro-properties of the nine group contexts employed in the SP design. In short, gamma change occurs (e.g., Golembiewski, 1990), and this implies fundamental changes in perceptions as a result of the SP design. This effect suggests the power of OD values and approaches, even in the limited time-frame involved here.

One aspect of the design requires elaboration. Note a significant difference in the history underlying the Group Behavior Inventory (GBI) responses coming from the two categories of groups involved in this design. Op Com was an intact team, and its members consequently had a substantial base of experience to make judgments. In contrast, the eight Upper Management groups are ad hoc collectivities whose members had variable prior contact as individuals and no history as a group. Their pretest came after about an hour of interaction on a task (see item 2 in Exhibit 12.4), when respondents were instructed to provide “your first impression” on the GBI pre-test. In both cases, the GBI post-test came as a terminal activity of the 3-day SP design.
The thirteen items assessing SP features and consequences were obtained at the same two points in time. The pretest instructs respondents to look back and provide “then” estimates of the 1988 SP cycle, while the post-meeting responses deal with “now,” or 1989.

ESTIMATING SP→OD CONSEQUENCES

In general—and in the case of macro-outcomes, almost without exception—the findings imply a successful SP experience. This conclusion is supported by reviewing results of the four approaches to macro- and micro-effects sketched above.

Effects on Macro-Properties

This research distinguishes six broad classes of OD→SP outcomes, whose thirteen constituent items are detailed in Table 12.1. The broad classes include:

Rating SP processes and outcomes (items 1, 8, 11 in Table 12.1)
Corporate trust level (2)
Success in surfacing and dealing with conflictful issues (3, 10, 13).
Success in uniting individual SOAs or functional units in support of collective goals rather than opposing one another (4, 6)
Degree of knowledge of, and comfort with, key players (5)
Degree of personal commitment and comfort (7, 9, 12)

Self-reports on eleven of the thirteen stimulus-items, given a successful design, are expected to be higher under the 1989 collaborative format than under the traditional format that dominated in 1988 and earlier. Hence, successful changes for these items are scored as “/>” or “Signif.>” in Table 12.1. Two items should decrease with a successful design—item 6 assesses whether unification of one’s department came as a result of we/they dynamics, and item 9 estimates the degree of personal frustration and discomfort with SP processes. In these cases, changes falling in the intended direction are scored “<” or “Signif.<” in Table 12.1.

Table 12.1’s preview of changes in macro-features indicates a successful experience, with modest qualifications. Thus twenty-three of the twenty-six comparisons (or 88.5 percent) fall in the intended directions, with over 34 percent of them (8 of 23) attaining $P = .05$ and an additional 13 percent (3 of 23) closely approaching that usually-accepted level of statistical significance. Of the three cases falling in unexpected directions, one achieves the .05 level and a second case approaches it.
### TABLE 12.1  Macro-Properties of Worldwide Strategic Planning, 1989 vs. 1988

<table>
<thead>
<tr>
<th>SP Macro-Features</th>
<th>Participative 1989 vs. Traditional 1988</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Upper Management</td>
</tr>
<tr>
<td></td>
<td>N varies*</td>
</tr>
<tr>
<td>1. Rating of SP processes—in encompassing the broadest range of issues, developing a coherent plan, and building support</td>
<td>Signif. (&gt;b)</td>
</tr>
<tr>
<td>2. Corporate trust level</td>
<td>Signif. &gt;</td>
</tr>
<tr>
<td>3. Surfaced necessary, even if unpopular, issues</td>
<td>Signif. &gt;</td>
</tr>
<tr>
<td>5. Knowledge of, and comfort with, major players</td>
<td>&gt; (.082)</td>
</tr>
<tr>
<td>6. United my department, but as a consequence of we/they dynamics with other departments</td>
<td>Signif. &gt;</td>
</tr>
<tr>
<td>7. Personal commitment</td>
<td>&gt;</td>
</tr>
<tr>
<td>8. Dealing with business strategies of competitors</td>
<td>&gt;</td>
</tr>
<tr>
<td>9. Personal frustration and discomfort</td>
<td>&lt;</td>
</tr>
<tr>
<td>10. Pleasant surprises—e.g., in isolating unexpected issues, data, and solutions</td>
<td>Signif. &gt;</td>
</tr>
<tr>
<td>11. Clarity of goals and purposes</td>
<td>&gt;</td>
</tr>
<tr>
<td>12. Personal estimates of average commitment of all members</td>
<td>&gt; (.091)</td>
</tr>
<tr>
<td>13. Dealt productively with disagreements or conflicts</td>
<td>&gt;</td>
</tr>
</tbody>
</table>

* Because of major acquisitions and minor turnover, various sub-Ns responded to different items. Only paired-comparisons are included in this analysis.

b “Signif.” indicates a difference that achieves or surpasses .05 level, by one-tailed t-test. The P values of cases falling between .05 and .10 are given in parentheses.

c Designates differences contrary to expectations.
Effects on Micro-Properties

The going-in expectation that the design will induce changes in macro-properties rests, basically, on the design’s ability to induce quite specific effects in each of the eight *ad hoc* groups. Simply, positive changes in micro-properties—here estimated by the seventeen items used to measure the five GBI dimensions—are taken to reflect movement toward regenerative interaction. In turn, such effects will tend to support a positive profile on the macro-features among the participants in the SP↔OD design. The five GBI dimensions include the following (Friedlander, 1968):

- **Group Effectiveness** measures the degree of creative, realistic team effort in sharing responsibilities and solving problems.
- **Mutual Influence** assesses the degree to which leaders and members influence one another.
- **Personal Involvement and Participation** estimates the degree to which members want, expect, and attain active roles in group meetings that generate a fulfillment reflected in a desire to continue group meetings.
- **Intergroup Trust versus Intragroup Competitiveness** depicts a group whose members have a confidence in one another, and who can alter personal opinions for the sake of progress toward a working consensus.
- **General Evaluation of Meetings** as good, valuable, pleasant, or bad, worthless, and unpleasant.

The first four scales are estimated by several items along 5-point intervals ranging from “strongly agree” to “strongly disagree.” The last dimension is estimated by seven equal-appearing intervals—good vs. bad, pleasant vs. unpleasant, and so on.

A successful Phase I experience implies increases in the five GBI dimensions, comparing the short post-test with the pretest. GBI items vary the direction of individual response stems to inhibit response set, but all dimensions are scored so that high scores indicate more attractive levels.

The rationale is direct. Traditional strategic planning has a percolate-up character, which means that substantial early consensus tends to develop within each of an organization’s component major units, as particularistic recommendations wend their separate ways toward the top. Then, broadly political processes of exchange and log-rolling occur at the most senior levels as bargains are struck and trade-offs made. In the present case, basic reliance is placed on the ability of heterogeneous Upper Management groupings to move toward early publicly-shared agreements about SP descriptions and prescriptions. The explicit intent seeks to counterweight any build-up of particularistic consensus in monofunctional or different SOA groupings, both during and after the SP design.

In a few words, the present design seeks to give an unprecedentedly large number of key worldwide actors broad and early opportunities for contributing to an emerging SP. The central assumption is that any integrative agreements
developed at the offsite will not only influence decisions there, but also will remain as counterbalances to the several particularisms that permeate back-home—different locations, diverse functions, variable ambitions, and so on.

The desired micro-effects require increases on all five GBI dimensions, in each of the eight ad hoc groups as well as in the Operating Committee. Together, such increases imply heightening trust and cohesiveness in the ad hoc groups, which imply greater control over member behavior as well as shared attitudes and norms (e.g., Golembiewski, 1962).

Does the evidence suggest that the nine groupings in the design developed the intended micro-properties? Table 12.2 presents relevant data for Op Com as well as eight Upper Management aggregates and, overall, eight of the ten means move in the expected direction, excepting Mutual Influence and Personal Involvement for Upper Management. For example, Group Effectiveness increases significantly for Upper Management, and trends in that direction for the Operating Committee. In addition, three of the eight changes in the expected direction far surpass the .05 level. The pattern for Upper Management clearly is sharper, which has to be interpreted in the context of the risks Op Com perceived in this fundamental change in their operating style—for example, the sharing of data with an unprecedentedly large group of executives and managers, which implied the sharing of influence.

Technical reasons imply that such comparisons can be interpreted at face value. Measurement properties of the five dimensions of the Group Behavior Inventory seem to support quite direct interpretations, except in the case of Personal Involvement.*

Macro- and Micro-Effects in Individual Learning Groups

The data also can be arrayed in another useful form. The data above are aggregated for all Upper Management groups, but useful comparisons here are made between the eight small ad hoc units in which direct learning occurred. The voluminous data are not reported here, to conserve space, and because their pattern is clear and consistent, in general.

This disaggregation supports the conclusion that the overall pattern of change is in the intended direction, as three points demonstrate. First, on very definite balance, the eight Upper Management groups do not differ on the pre-test estimates of the eighteen measures—thirteen measures of the macro-environment, and the five Group Behavior Inventory dimensions. In sum, less than 8 percent of the 504 total comparisons differ significantly by t-test, two-tailed. Hence, pre- vs. post-comparisons permit uncomplicated interpretations, overall.

Second, in general, the eight Upper Management groups generate a pattern

## Table 12.2 Scores on Group Behavior Inventory Dimensions, Pre-vs. Post-Test

<table>
<thead>
<tr>
<th></th>
<th>Upper Management, total means</th>
<th>Operating Committee means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Post-test</td>
</tr>
<tr>
<td>Group effectiveness</td>
<td>3.66</td>
<td>3.82</td>
</tr>
<tr>
<td>Mutual influence</td>
<td>3.87</td>
<td>3.83</td>
</tr>
<tr>
<td>Personal involvement</td>
<td>3.91</td>
<td>3.89</td>
</tr>
<tr>
<td>Intergroup trust</td>
<td>3.47</td>
<td>3.72</td>
</tr>
<tr>
<td>General evaluation of meetings</td>
<td>6.16</td>
<td>6.49</td>
</tr>
</tbody>
</table>
Large-System Change in Business 315

of pre- vs. post-test changes in the expected directions. Specifically, 74.3 percent of the cases fall in expected directions, and 29.9 percent either achieve $P \leq .05$ or closely approach it ($P \leq .1$). No contrary cases attain the usually accepted level of statistical significance, and less than 3.5 percent of these cases approach significance. In seven of the eight Upper Management groups, in fact, this summary understates the strength of the pattern, as is suggested by a concluding point.

Third, one of the eight Upper Management groups constitutes a clear outlier. Over 55 percent of its pre- vs. post-test comparisons fall in an unexpected direction, and 11 percent of them closely approach significance. Not surprisingly, this ad hoc group had unusual difficulties getting started, especially due to travel delays and unexpected absences of individuals assigned to it. This seems accidental, since group assignments were revealed only at the initial meeting.

The analysis thus far considers only the most direct form of change—alpha, where incremental differences only are observed, and the psychological space as well as measuring intervals are substantially the same. Here, gamma change receives attention: the interventions are so potent as to change the very dimensions being measured as well as their measurement intervals. See also Chapter 24.

Gamma Change

In addition, evidence implies that the SP→OD intervention has effects far more powerful than those estimated above. Directly, gamma change seems to have occurred—not a change in degree, but a fundamental change in the state of the psychological space in terms of which respondents structure their pre- vs. post-test responses to the thirteen macro-environmental items and the seventeen GBI items (Golembiewski, 1986). Thus, the brief interventions seem to have had the powerful effect of restructuring the way respondents view the psychological space encompassed by the items relied on here.

Chapter 24 provides other conceptual detail on trinitarian change. To sketch the statistical procedure, the congruences of pre- and post-test factor analyses are compared, and the four estimates of the shared variance in all paired-comparisons range from 42 to 48 percent. Separate factor analyses are run for total respondents as well as for Upper Management only, on both pre- and post-test, but Op Com had too few members to permit this procedure. Ahmavaara’s (1954) procedure generates two correlation coefficients for each paired-comparison of factorial structures. In sum, the several estimates of the congruence of pattern as well as of pattern and magnitude for each pair of factorial structures covers a narrow range of coefficients: .6486 to .6896.

The convention has been to consider 50 percent or less shared variance as the indicator of gamma change (Golembiewski, Billingsley, and Yeager, 1976), and this seems conservative. Such massive incongruence is not easy to explain as one variety or another of measurement error. The preferred hypothesis here
is that individuals generally and consistently restructured the psychological space encompassing the 30 items, in the short interval between the pre- and post-test. Here these structures are isolated by factor analyses, whose congruences in terms of both magnitude and pattern are estimated by Ahmavaara’s (1954) technique.

In its present form, however, the operational definition of gamma change does not specify direction, and therefore the strict interpretation of gamma change does not here permit a judgment whether the change was in the intended direction. Other operational definitions of gamma change overcome this major limitation (e.g., Golembiewski, 1990, pp. 241–272), but they are not applicable here. Interviews with participants, as well as the reactions of the consulting staff, support the view that the conceptual restructuring here trends in the desired direction.

Conventional analysis focuses on alpha change only—that is, on changes in degree on stable dimensions, as estimated along relatively constant intervals. From the conventional viewpoint, gamma change can be dismissed as measurement error, and unqualified judgment is appropriate only when alpha-alone has been established. In the present case, since gamma seems to have occurred, the several low Cronbach alphas in question cannot be viewed as simple measurement error. (See the Note on the next page.)

The three preceding tests of design effects assume only alpha effects, or differences in degree along relatively stable dimensions of reality. The fourth kind of analysis reinforces their sense of general change, that its effects were sufficiently powerful to qualify as gamma change.

Conceptual details about the difference between alpha and gamma change are available conveniently (e.g., Golembiewski, Billingsley, and Yeager, 1976; Golembiewski, 1990). The research status of gamma has been reviewed recently (Tennis, 1989), and that summary reports the broad acceptance of this conceptualization of plural change and emphasizes its centrality in OD.

**DISCUSSION**

In brief, this assessment of an in-process SP↔OD effect implies that intended effects occurred. The interpretation is the same when the data are reviewed for alpha change as well as for gamma.

As for alpha, the generalization about intended effects holds especially for the ad hoc Upper Management groups, on both the macro- and micro-effects, as subsections 1 through 3 establish. All eighteen of the expectations for the Upper Management aggregate are confirmed in comparisons of pre-test vs. short post-test. Eight cases attain $P < .05$, and an additional three cases closely approach that level. For Op Com, fifteen of the eighteen comparisons fall in the expected directions, but only three attain $P < .05$.

In general, the difference between Op Com and Upper Management do not appear surprising. Thus, one can understand why the 3-day session seems to have
been only marginally (if positively) impactful on assessments of GBI dimensions by long-standing Op Com members. Moreover, on the macro-effects, Op Com sees itself as stretching to depart from traditional practices, and this may induce unrealistically high expectations in Operating Committee members about effects on Upper Management. In sum, Upper Management personnel may have been positively impacted by changes from the traditional SP pattern, while their behavior still reflects too much of a wait-and-see attitude for Op Com tastes.

As for viewing the data as gamma change, both Upper Management and total populations suggest a major conceptual restructuring between T1 and T2. Close observers report that this basic restructuring is in the intended direction, in general.

Phase II of this design was intended to provide useful guidance for these and other possible interpretations of effects in at least three senses. First, close attention would have been given to the three macro-effects on which Op Com members report effects contrary to expectations.

Second, the pretest involving 1988 SP processes involves a “then” estimate of how things were in the last cycle. Some evidence suggests the reliability of such “then” estimates (e.g., Terborg, Howard, and Maxwell, 1980), and Phase II data would have permitted more confident real-time tracking of how participants respond to SP dynamics set in motion by the present design.

Third, later interviews and financial data would have permitted detailed cross-checking of the self-report data reported here.

Phase II did not occur. Major changes in corporate ownership occurred before the scheduled extension of the design, and not only did the resulting turbulence delay the planned follow-up for such an extended period but a new pattern of ownership also encouraged a fundamental reorientation of corporate management and discourages extension of Phase I discussion.

NOTE

Cronbach’s alpha indicates that four of the five GBI dimensions have reliabilities appropriate for research purposes. In sum, for the pretest:

<table>
<thead>
<tr>
<th>GBI</th>
<th>Group Effectiveness</th>
<th>Mutual Influence</th>
<th>Personal Involvement</th>
<th>Intergroup Trust</th>
<th>General Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Respondents</td>
<td>.8335</td>
<td>.8334</td>
<td>.5885</td>
<td>.2947</td>
<td>.6134</td>
</tr>
<tr>
<td>Upper Management</td>
<td>.8709</td>
<td>.8612</td>
<td>.4216</td>
<td>.2041</td>
<td>.3918</td>
</tr>
<tr>
<td>Operating Committee</td>
<td>.7804</td>
<td>.7744</td>
<td>.6193</td>
<td>.3334</td>
<td>.5801</td>
</tr>
</tbody>
</table>

The consistently low alphas for Personal Involvement encourage interpretive caution, of course. See also the discussion of gamma change.
REFERENCES


Large-System Change in Business


Illustrating Large-System Change in Government
Examining Some Surprises in Labor/Management Cooperation

This chapter adds to the sense and substance of OD applications, this time in the public sector. This is, in fact, the second in a quartet of chapters used to illustrate the character and potential of large-system interventions. Chapters 12 through 15 detail these interventions—in business, government, health care, and at the politics/administration “interface.”

This chapter presents a mini-case study of an organizational change effort in the Employment and Training Administration (ETA) within the U. S. Department of Labor. The text describes the agency and its roles, sketches the planning effort, and dwells on the factors that influenced the initial blockage of the effort. That is, poor fit existed between situational factors and change requirements toward the end of the scheduled effort, as exacerbated by the inadequate melding of professional interests in ETA and Labor.

Nonetheless, matters in Labor were resolved in favor of the effort at planned change, as a concluding section notes but does not even attempt to explain here. That full demonstration remains for another time and place.

These four chapters relating to Irony V are at one in the sense that the reach-and-grasp of OD can be increased by greater attention to large system change, but the quartet also reflects a general specialization of roles. Thus, Chapter 12 deals with strategic planning (SP) in a business context, which is an arena that is flagging somewhat because SP applications have been insufficiently informed by
theory and experience of change (e.g., Golembiewski, 2000). Chapter 14 provides a detailed view of the change process in a hospital setting, and emphasizes the range of measurements in a complex setting that relate to establishing success rates. The record in Chapter 14 is very encouraging. Chapter 15 provides a summary of OD applications around the politics/administration interface and, while neglecting details in the mode of Chapter 4, this concluding Chapter 15 provides summary optimism that can be attached to change at macro-levels, usually in the public sector, where change is alleged to be particularly difficult. And, now, we can target Chapter 13’s focus on an OD application in the U. S. Department of Labor. Deliberately, this chapter seeks to illustrate the welter of temporary inventions in a program of planned change at a systemic level, as well as to extract a few generalizations about change suggested by Labor’s experience.

INTERVENING IN THE U.S. DEPARTMENT OF LABOR*
VECTORS TOWARD SUCCESS/FAILURE IN PLANNED CHANGE

Since your author and a colleague, C. F. Miller (C.M.) have in another place reported quite favorable success rates of planned change at the politics/administration interface (Golembiewski and Miller, 1997, 1998; Golembiewski, 1999), it is only fitting to take a close look at a set of interventions that got off to a fine start but later took a twisting trajectory that left the final fate in doubt until the last moments.

How did this Labor outcome occur, given the quite positive record in general at the interface, as well as the favorable initial prognosis and progress in Labor? This compound question preoccupies this chapter, and your author and his colleague represent very different personal perspectives on those dynamics. C.M. was a Labor employee, a union official for field employees of the department, and a prominent actor in the various twists-and-turns of the processes of this planned change. R.T.G. was an external consultant who came to Labor relatively late in the scenario. How do these hopefully observant participants see the Labor dynamics? In essential senses, a summary explanation follows the prescription for a wedding: something old, something new, something borrowed, and something blue.

The metaphor closing the preceding paragraph will get only a little elaboration here. The “something borrowed” came out of two partially overlapping but congruent sources. For C.M., his guiding ideation came out of a strong sense of the need to transcend adversarial bargaining between labor and management,

* The original version was a paper presented at the Fifth National Public Management Research Conference, George Bush School of Government, Texas A & M University, College Station, TX, December 4, 1999.
which had its proximate roots in his personal preferences. These were reinforced by Labor’s immediate history of support for cooperative arrangements between labor and management, as well as his own career progress in union leadership. For R.T.G., a background in conventional OD provided his lenses for viewing the world (e.g., Golembiewski, 1993). These two basic identifications implied much that was relatively new, and borrowed to reduce resistance in public-sector planned change; and the apparent outcome in ETA reflected much of the old and expected that could frustrate even careful attempts at planned change (e.g., Golembiewski, 1967).

And what of the blue? That closely describes the emotional tone of R.T.G. and C.M. around the time of the apparent failure of the effort. Moreover, for many in Labor, disappointment also expressed their reaction to what seemed the final outcome, if it did not deepen a cynicism that even the good struggle often comes to naught.

Finally, and also like many weddings, this one had its rocky stages that led to a separation. But hope of a reconciliation existed as of the day that the writing of the initial draft of this chapter was begun, and they prevailed. Quite soon, Assistant Secretary Ray Bramucci announced an ETA restructuring that was recommended by a Joint Labor-Management team acting on working materials whose development will be detailed below.

Expressed more prosaically, attention below falls under three heads. Immediately, the text describes the context in which the change effort was rooted. Then, some unexpected events that undercut the application are introduced. Finally, positive recent events conclude this chapter.

IMMEDIATE CONTEXT OF CHANGE EFFORT

To launch this mini-case study of the early development of the Labor experience, the context is described in which several unexpected effects occurred as the project underwent its early implementation stages. Specifically, this involves two basic elements: a description of the components of the host for the change effort; and several common features of that host as well as of its milieu.

Components of Employment Training Administration

The focal agency, henceforth ETA for convenience, is one of seven major units within the United States Department of Labor. ETA is responsible for the funding and regulation of training programs administered by state and local agencies, and it also operates the United States Employment Service as well as the Unemployment Insurance system. Other functions of the ETA include: the establishment of standards for apprenticeship programs and the promotion of related programs among individuals and private industry; the provision of technical assistance and
advisory services to private industry; the establishment and registration of quality high-skill training programs; the protection of the welfare of apprentices; the assurance of equal employment opportunity; and credentialing of training programs and participants (Goldinger, 1990).

If not a suitcase agency, then, pre-change ETA could make a reasonable claim for heterogeneity. Somewhat more specifically, the idealized thrust toward ETA integration involves three clusters of activities. In outline, they are:

1. Mainline ETA, under the direction of program units of the Office of Regional Management in Washington D.C. and ten regional administrators, had approximately 552 employees in Washington and 558 employees in the regions.
2. Job Corps (JC), under the direction of a national director and ten district directors, had approximately 76 employees in the national office and 150 employees in ten district offices.
3. The Bureau of Apprenticeship and Training (BAT), headed by a national director, with ten regional directors, and a state director in each of the fifty states. BAT included approximately 32 employees in the national office, and 72 employees in field offices outside Washington, D.C.

Some operating details provide an added sense of ETA as context for change. Figure 13.1 assists the reader in further appreciation of this context.

Mainline ETA

One cluster of ETA activities is responsible for the direction of employment and training programs under the Job Training and Partnership Act of 1982 (JPTA), the U.S. Employment Service (ES), and a Financial and Administrative Management Section. JC administers programs for special targeted groups, with the basic authorizing legislation going back to the late 1960s. BAT provides administration of the National Apprenticeship Act of 1937.

Office of Job Training Programs

Operating through a block grant funding process, this Office provided the majority of its funding to job training programs. It worked with private industry along with both state and local governments to plan, monitor, and assess these programs. It also operated a displaced worker program which, pursuant to the Job Training Partnership Act, offered retraining and other services to workers who are unlikely to return to their former jobs.

U.S. Employment Service

Under the provisions of the Wagner-Peyser Act, ES provides assistance to states in establishing and maintaining a system of local public employment offices.
Through this network of state-operated job service offices, ES helps jobless workers find employment and assists employers in finding workers. ES provides state and local labor market information to facilitate the working of the various labor markets. ES also aids workers who have been displaced by foreign competition, and it ensures that the admission of aliens into the nation's workplace does not adversely affect U.S. workers. ES has responsibility for certifying hard-to-employ workers, enabling employers to take advantage of Targeted Jobs Tax Credits (TJTC). In most metropolitan areas, ES maintains a computerized job bank listing local, intrastate, and interstate job openings. ES also encourages local job service employer committees to suggest ways field offices can more effectively serve local industry and community needs.

Unemployment Insurance

ETA also serves those temporarily out of the workforce. The UI section provides leadership and policy guidance to state employment security agencies, monitors state unemployment security programs to ensure compliance with federal laws and regulations, assists states with the administration of unemployment programs,
and establishes formulae to determine the amount of money needed to administer
state unemployment services (Goldinger, 1990).

Financial and Administrative Management

This administrative section provides budget, accounting, and data analysis ser-

vices for the agency. It also plans, develops, promulgates, and executes policies,

and guidelines governing management information, budget, account-
ing, and financial procurement systems (Girard, 1993).

Job Corps

Under the provisions of JTPA, Title V, the Job Corps is a federally administered,
national program providing education, training, and support services to youths
in 109 centers throughout the country. JC Centers furnish a comprehensive array
of services designed to overcome the multiple barriers to employment faced by
disadvantaged youth. The Departments of Agriculture and Interior operate Job
Corps Civilian Centers on public land, while major corporations and nonprofits
manage and operate the rest under Job Corps auspices (U.S. Department of Labor,

Bureau of Apprenticeship and Training

Under the National Apprenticeship Act, BAT is responsible for safeguarding the
welfare of apprentices, which ties ETA into a long tradition of craft unionism. This
includes the formation, promotion, and furtherance of labor, necessary stan-
dards, cooperation with the states in the promotion of such standards, and the
bringing together of employers and labor to formulate programs of apprentice-
ship. Programs must meet standards established by the Bureau or a recognized
State Apprenticeship Council. BAT works closely with employers, labor unions,
vocational schools, community planning groups, and other organizations con-
cerned with apprenticeship. BAT also conducts field compliance reviews to deter-
mine conformity with federal equal employment opportunity laws as well as with
other standards for apprenticeship and training (Girard, 1993).

Some Central Commonalities

These (and other) multiple activities and dates of inauguration contribute to a
pattern, or at least to some common features of a pattern. Even these sparse details
suggest the potential for differentiation and even fragmentation along the lines
of the formal structure sketched in Figure 13.1. This potential is exacerbated by
differences in missions, as well as conflicts inherent in such differences; differing
dates of start-up, which suggest different key coalitions among the congress and
other stakeholders; and the implied multiple webs of contacts between ETA com-
ponents and their “little groups” of policy neighbors—individual legislators, different legislative committees and subcommittees, national and local interests, and so on.

The “common features of a pattern” can be sketched, if with some caution. Each of the three major ETA units operates independently and has established its own cultural identity. Although located in the same ten cities, illustratively, the JC and BAT regional directors report to their respective national directors, while mainline regional directors report to the Director of the Office of Regional Administrators located in the national office. All program heads and the three national directors report to the Assistant Secretary of Labor for ETA. Overall, then, ETA was a traditional bureaucratic structure with prominent “stovepipe effects” in which there was too little integrative communication between organizational units, and employees concentrated on performing the separate missions of their respective organizational units, in Washington and multiple field locations.

Cultural Change in ETA

This pattern of differentiation notwithstanding, an integrative recognition began to grow—primarily among Labor political appointees, but also among many professional or career employees. In 1991, Secretary of Labor Lynn Martin introduced a “Serving the Customer” exercise in which all ETA employees were required to identify their direct internal and external customers, and then to develop specific methods to improve service to them. This was a radical departure from the traditional top-down approach in ETA. Centrally, employee participation was solicited in the development of ways to identify and serve customers better.

The exercise proved catalytic. As one effect, two champions linked up in the early 1990s: Secretary Martin and Jesse Rios, president of the National Council of Field Labor Locals (NCFLL), the union which is the exclusive representative of all regional or field non-management employees. Among the products of this linkage were an Employee Involvement Quality Improvement (EIQI) initiative. A parallel agreement, the Quality Partnership (QP), was also signed in the national office by Local 12, a separate union which represents all non-management employees in the national office.

Under both of these agreements, to simplify more complex realities, a Union Management Pair (UMP) was designated for each national office and each mainline ETA regional office. Each UMP was responsible for establishing a quality improvement program in their members’ respective offices. A modified version of Deming’s (1982) Total Quality Management (TQM) process was selected as the model for the program. All ETA employees received training in TQM quality improvement methods, as well as experiential learning in how to work more effectively in teams. The evidence seems convincing that these two training
efforts, and the broad rationales underlying them, had major support in Labor and ETA. See the Reinvention Office, in Figure 13.1, which clearly reflects a substantial interest to seek new ways-and-means for doing business in Labor.

In 1993, with the change in presidential administrations as well as the inauguration of the National Performance Review (NPR), the emphasis in Labor switched from improved customer service and broadened to adapting TQM to more effective and efficient operations, as well as to better customer service through entrepreneurial government, or “reinvention.” Secretary Robert Reich selected Doug Ross as the Assistant Secretary to spearhead this effort in ETA. Assistant Secretary Ross had initiated the same type of effort while head of the Michigan Department of Commerce in the 1980s. He emphasized a mission-driven organization characterized by “chunking and hiving,” to structurally reinforce the decentralization of authority seen as necessary to respond more quickly to customer needs (Osborne, 1993).

Immediately after he was appointed, Assistant Secretary Ross began to develop statements—for mission, vision, and goals. He solicited input from all employees through a series of national and regional office meetings sponsored by EIQI and Quality Partnership groups, then standing for some time. Several draft documents were created and shared with all employees within ETA, through an all-employee “town hall meeting” as well as via electronic mail. All employees had the opportunity to comment on the final version before it was shared with Secretary Reich.

The final products included the following key statement: “ETA’s mission is to ensure that all Americans have access to the resources they require to successfully manage their job lives, and that U.S. businesses have access to the skilled workers and training resources they need to successfully compete in a global marketplace” (ETA, 1994, p. 19).

The value statement is too long to quote extensively here, but brief excerpts provide a useful sense of the Ross-championed proposal to Secretary Reich. The statement emphasized the terms “Customer service, quality, equity, personal responsibility, partnership, and public trust.” They are to be used to distinguish the right from the wrong way to do things within an organization where employees are empowered to respond to situations where the ‘one size fits all’ approach is no longer satisfactory because situations and solutions vary too greatly” (ETA, 1994, p. 21).

To apply such value stances, further, goals of ETA were “phrased as outcomes or descriptions of situations that must be achieved if ETA is to fulfill its mission” (ETA, 1994, p. 16). The goals included:

1) Increased individual and business demand for direct information, employment services, and training to provide the skilled workforce
and high performance businesses required for vigorous job growth and rising standard of living in an intensely competitive international economy.

2) Widely available learning systems for high school students and young adults that enable them to master the academic competencies and occupational preparation necessary to meet employers’ expectations and find a good first job without a four-year college education.

3) Customer-driven, high quality local labor markets that provide individuals with the information, employment services, career planning, training opportunities, and insurance systems to take effective control of their job lives and that provide businesses with the information and services they require to recruit an educated and competitive workforce.

4) Learning and labor market systems that are accessible and affordable to all American businesses and individuals, including individuals at risk economically, the working poor, and those outside the job market. ETA has specific historic commitment to the disadvantaged that must be enhanced, not diminished, by a broader mission to serve all Americans (ETA, 1994, p. 16–17).

The ideals embodied in these statements were intended to guide the cultural change necessary to undergird the reinvention of ETA. The statements took several months to develop and reflected the definite balance of employee attitudes and preferences. The essential sense of these dynamic and involving activities were reflected in Executive Order 12863.

A Reinvention Team (RIT) was also established during this time. This steering committee consisted of twelve members—including Assistant Secretary Ross, two regional managers, and six union representatives—four national office representatives from Local 12, and two regional representatives appointed by the NCFLL. RIT soon was expanded to include two regional mainline union representatives, with co-author Miller being one of them, a JC regional manager and union representative, and a regional BAT manager and union representative.

As a clear reflection of its preferred mode of operation, RIT established “construction crews” comprising management and bargaining unit volunteers, who were empowered to identify stakeholders and partners, to conduct required surveys, and create customer standards and plans that would comply with the requirements of Executive Order 12863. RIT also established a “Steam Team,” which was also a joint labor-management team, to develop a plan for “streamlining” the operations of ETA. Their final recommendations included the establishment of customer clusters focusing on the different external customer groups ETA served, a reduction in staffing, and increased investment in the three T’s—travel, technology, and training.
Threatened Gutting of ETA

No sooner had ETA personnel begun to settle in to such cultural changes, in early September 1995, than events seriously tested Labor and ETA, as well as the employee involvement and resolve invested in the cultural change. Directly, the Republican-controlled Congress proposed two bills to consolidate over 120 education and job training programs administered by two departments—Education and Labor.

This threat went substantially beyond mere talk. The House of Representatives passed the CAREERS Act; and the Senate passed the Workforce Development Act of 1995. Both versions called for a 40 to 60 percent staff reduction in ETA, and a transfer of ETA’s major programs and functions. This transfer could occur in one of two ways—either directly to the states through block grants to the governors, or to a newly created Workforce Development Board including several business leaders, two state governors, and the Secretaries of Education and Labor.

Other changes added to this environmental turbulence, and the positive experience with cultural change was soon tested in a serious way. Timothy Barnacle, a former ETA Regional Administrator who had been active as a lobbyist, replaced Doug Ross as Assistant Secretary for ETA. Barnacle immediately asked the RIT to develop ETA’s response to these legislatively proposed changes and threats.

The RIT quickly created an umbrella taskforce—Organization Design and Transition Team (ODTT). It consisted of a Deputy Assistant Secretary, one RA, one Local 12 representative and one NCFLL representative, all of whom met on a weekly basis.

ODTT soon established it would not be satisfied with an analytic role. Team members directly begat several mandated entities: a Federal Role Team (FRT) to define ETA’s new role; a Structural Design Team (SDT) to make recommendations for a new organizational structure for the agency; an Operations and Transition Team (OTT) to provide guidelines intended to ease transition to the agency’s new role and structure; and an Internal Communications Team (ICT) to keep all ETA associates informed about the change efforts. On September 29, 1995, RIT sent a memo to all ETA associates (read “employees”) informing them of the creation of the ODTT, and soliciting their involvement in the process of “building a new organization.” Organization design work was expected to begin in January 1996 (ETA, September 29, 1995 memo).

After a brainstorming session—attended by approximately 35 management and bargaining unit employees representing all national office program units and regional components of Mainline ETA, JC, and BAT—the ODTT developed a plan to implement the reorganization initiative. Specifically, the plan provided that:
FRT would develop a Federal Role paper.  
SDT would develop a structural design model that would be used to achieve their proposals.  
OTT would develop a plan for the continued operations of the agency during the transition from the current structure to the structure proposed by the SDT.  
ICT would provide information to all associates so that they could be aware of what was transpiring at all times, and also would have opportunities to comment during all phases of implementation.

On October 2, 1996, conveners—selected by management and the unions—were selected to head each team, after they were approved by the RIT. Volunteers also were selected for each team, after approval by both management and the unions.  
After a three-month delay caused by a dispute between management and the unions concerning the selection and roles of team members, the teams were finalized. In sum,

FRT was composed of three management conveners, nineteen national office associates, nine regional managers, and sixteen bargaining unit employees.  
SDT had three national office managers and an NCFLL representative as conveners, nineteen national office associates, six regional managers, and ten regional bargaining employees.  
OTT was composed of one national office manager and two regional managers who served as the conveners, eight national office associates, five regional managers, and six regional bargaining unit employees.  
ICT had a Local 12 representative, a national office manager plus an NCFLL representative as conveners, seven national office associates, four regional managers, and ten regional bargaining unit employees.

All four teams met in April 1997, four months after the original target date, with a mandate from the ODTT to move quickly because various Congressional subcommittees were holding hearings on the proposed consolidation of ETA programs. The reorganization plan had to be ready to go when and if the legislation passed and was ready to implement.

Overall, evidence implies that the teams moved carefully, even ponderously, and certainly not at a forced march. Suggestively, all four teams responded by establishing ground rules and developing charters. Developmental details vary, but a general pattern seems to have applied. For example, FRT created three subteams to (1) examine federal responsibility under the current legislative authority, as well as (2) under changes anticipated in the proposed legislation to
analyze what customers seek from ETA; and (3) to revise mission, vision, and values statements.

Similarly, SDT established four subteams to look at ETA’s current organizational structure; to catalog previous restructuring efforts in ETA; to survey other agency efforts at restructuring; and to review research on structural design theory.

SDT ran on a parallel track, simultaneously with FRT, while members developed their products. After establishing their charter, OTT suspended activities until SDT completed their proposal. ICT operations surfaced earliest: its members developed a communication plan and inaugurated a weekly newsletter, *The Grapevine*. The newsletter was sent via electronic mail to all ETA associates and kept them informed of all ODTT-related activities.

Although FRT was supposed to complete its report in four to six weeks, their initial draft paper was not sent out for comment until July, or approximately three months after the first meeting of the ODTT creations. The final version was submitted in November, which approximates 7 months after the initial April 1996 meeting.

Arguably, FRT occupied the key pathway on most relevant PERT charts, and FRT’s measured pace was widely advertised and noted. The FRT conveners continually advised ODTT and RIT that their report would be delayed because members of its third subgroup were unable to complete their activities due to competing workloads, compounded by an inability to reach consensus on the report. Crucially, also, SDT could not complete their final proposals without this input.

The overall case caused some discomfort, especially but not exclusively among political officialdom. The Assistant Secretary, RIT, and ODTT stressed the need to expedite the process due to continuing legislative activity. Although the FRT report was not finalized, SDT was eventually told to proceed with its draft version.

These efforts to force-draft the several streams of activity proved unnecessary, although the effort no doubt was prudent, just in case. Although both houses of Congress passed bills calling for the consolidation of ETA’s programs, legislators were unable to come to an agreement when it came to the conference committee’s attempted reconciliation of the two bills. Both bills died.

After much discussion among the political appointees, upper level management, and the unions, RIT decided to continue ODTT’s activities, focusing on restructuring the organization because the last restructuring effort occurred 15 years earlier when ETA had 4,500 employees, very much larger than 1996 employment. Most commentators concluded there were many gaps in the Labor structure that needed to be filled or eliminated, no matter what the congressional impasse.

All these efforts encouraged action among the ETA committee structure. After a series of telephone conference calls and face-to-face meetings in Washington-
Large-System Change in Government 333
ton, SDT issued its report in November 1996. The report contained twelve structural design options. An abridged version was sent to all employees through *The Grapevine*, and the full report was offered to any associate who requested a copy. Two members of the team held face-to-face meetings with all employees in all 10 regional cities, and four town hall meetings were held in the national office during November and December. ODTT’s members explained the highlights of each option, answered questions about them, and all associates were given until the end of December to provide written comments and suggestions to the conveners of SDT. A final recommendation was scheduled for January 1997.

**A Little Help From Friends and Neighbors**

While these dynamics internal to ETA were under way, related events in Labor suggested that a supportive context existed for change. Broadly, the ETA effort could count on inheriting managerial currency from the National Performance Review, launched with so much ballyhoo—if at times a cross-pressured hype (e.g., Golembiewski, Vigoda, and Sun, 1999)—under the direct championing of Vice President Gore. In addition, three Secretaries of Labor in a row—Elizabeth Dole, Lynn Martin, and Robert Reich—were helpful in efforts to achieve change in Labor. And, proof of the pudding, OSHA—one of ETA’s neighbor agencies in Labor—had successfully completed a change project focused on the customer/client via a horizontal structuring of work, with substantial public notice (e.g., Minner, 1993).

**SOME UNEXPECTED EVENTS**

All this concern for participation and buy-in—in largest part probably idealistic, but in critical cases apparently motivated by hedged self-interests rather than humanism—ran afoul of changing situational features. More or less, the major events can be labeled “unexpected,” in one or both of two senses. Thus, some of the events were low-probability possibilities that nonetheless occurred, and simultaneously. And other events were unexpected in the sense that they were not clearly consistent with rules-of-thumb usually associated with successful organization change, and in cases stood flatly inconsistent with accepted guides for praxis. Six cases of the “unexpected” are distinguished below, in the context of usually accepted notions about organization change. Broadly, the reviews below may be seen as qualitative complements to the quantitative analysis in Chapter 3.

**The Common Wisdom Is Virtually Unanimous on One Point: Planned Change Often Founders Because Efforts Lack Support from the Top**

ETA seemed well prepared in regard to this aspect of the common wisdom, but several unexpected simultaneities occurred. Consider that Secretary Reich an-
nounced his resignation during the first week of the SDT presentations and Assistant Secretary Barnacle announced his resignation two weeks later but indicated he would serve until Secretary Reich was replaced; and before any action could be taken on the large number of suggestions made by ETA employees in review meetings about the proposed structural changes, Assistant Secretary Barnacle advised RIT that all reorganization efforts should be put on hold. And it seems possible, even probable, that any previous presidential support had been withdrawn.

No one can definitely explain all of these unexpected events, although some have left extensive records on the matter (e.g., Reich, 1997). As a partial explanation, for example, Barnacle told Labor associates that some members of ETA not only opposed the restructuring effort, but also had gone to their Congressional contacts. These legislators were threatening to hold hostage the already-troubled nomination of Alexis Herman as the new Secretary of Labor as long as ETA interests persevered in opposition.

After several days of discussion and debate, RIT abandoned the restructuring initiative. Here, or so it seems, the common wisdom proved serviceable, as was reflected in a string of events whose simultaneity was unexpected.

Not Only Is “Support from the Top” Often Preferred and Usually Considered Necessary, There Seems No Such Thing as Too Much of This Kind of Support

Indeed, this aspect of the common wisdom usually seems to be seen as face-valid. Who can question “support”? And where better to have such champions than “at the top”?

At least at times, however, this advice can be limited and limiting, and even counterproductive. And so it was in ETA.

Precision is not possible, but clearly the essentials of this prescription relate to the critical role of “champions” or “idea champions” (e.g., Peters and Waterman, 1982). As one popularizer explains, such champions “provide the time and energy to make things happen. They fight to overcome natural resistance to change and to convince others of the merit of a new idea” (Daft, 1995, p. 273). This prescription certainly must be granted, in some part. But how big a part? Hence, organizational development theorists have emphasized the need to get the commitment and involvement of top management to accomplish successful organizational change (Goodstein and Burke, 1991).

To a substantial degree, and especially early on, ETA was well situated in this particular. Doug Ross had a reputation for having been a transformational leader in Michigan, and he clearly attempted to instill the need for a sense of vision, goals, and values in ETA. A typical cultural transformation model was introduced during his tenure, and employees were encouraged to think and act
“outside the box,” to become involved in developing ways to strengthen the organization and to work with others rather than on their own program in their own “silo.” Ross attempted to mobilize the commitment of the agency to his vision, and also to institutionalize the changes by establishing RIT and using it as a vehicle to involve all levels of ETA associates in the change process. The unions and management representatives were called on to become the champions of the process. All of these actions are consistent with the need to have transformational leaders, rather than transactional leaders, revitalize organizations (Tichy and Ulrich, 1984).

ETA was in the process of accomplishing the transformation from a traditional top-down bureaucracy to an entrepreneurial organization envisioned in the National Performance Review, among other places. However, as is typical of many public management efforts, additional external forces intervened. Specifically, as noted above, legislation was introduced that would dramatically reduce the size and roles of ETA.

Championing in ETA waned some at this crucial point, due to the expected short tenure of federal political appointees. The Clinton Administration replaced the transformational leader with an experienced political actor who was sympathetic to ETA’s historic programs and operations. To be sure, early on, Tim Barnacle attempted to respond to his own sense of the challenges without overturning the reinvention efforts begun by his predecessor. However, other classic obstacles to organizational change in ETA surfaced, and the overall effort was put on hold even as it had made significant progress in particulars.

Basically, attempts to unblock resistance through training and involvement in the reinvention of ETA did not suffice. Perhaps this was due to inadequate time for any learning to take hold before Congress attempted to deconstruct the agency, if not demolish it. This is often seen as a severe constraint inhibiting change in the public sector.

But could things have been different? No one will ever know but, just possibly, there was too much support from the top at one point, as well as too little later in the process. Consider the experience of one of the co-authors. As he made the rounds of political and career officials in the agency, he tried to encourage public commitments from those who expressed clear preferences. “Why waste it?” was the typical comment. “The big hitters are on the same page, and maybe even the President. Who needs me as a champion?”

In sum, the rule-of-thumb about “support from the top” may have led to an unexpected effect. Thus, that support toward the end may have been insufficient to marshal internal support against external attack; but that support from the top concerning ETA, unexpectedly, may have been great enough at earlier stages to discourage others to take their stand as public champions.

In any case, efforts to encourage champions to surface continued even when they seemed unnecessary. Later events justified the effort.
No Such Thing as a Bad Program for Change

When in their most expansive mood, change agents often approach, if they do not actually propose, the Panglossian view that there is no such thing as a bad program for change; many programs will work; and the only thing certain to kill a program of planned change is inadequate participation, involvement, and buy-in by stakeholders. Indeed, the present author almost accepts such a view.

This rule-of-thumb is almost certainly insufficiently nuanced, however. Granted, too few opportunities for participation, involvement, and buy-in can emasculate any program, but one can argue that ETA overdid it, leaving proponents of change vulnerable to mere delays and obfuscation by a dyed-in-the-wool opposition.

At the very least, proponents of the change went to great lengths to act consistently with this third guideline and, arguably, this led to a strategic vulnerability. At a minimum, the program could have been stone-walled by those demanding ever-greater roles and looser schedules.

Certainly, the details above suggest many such opportunities for sand-bagging the change program. For example, during the process of reviewing structural recommendations, BAT managers complained that many of their associates were not located in regional cities and they were therefore unable to attend the briefings on the key SDT reports because they did not have sufficient travel funds. Additional funds were made available and several more face-to-face meetings were held in Atlanta, Chicago, Philadelphia, and San Francisco in early January. The comment period extended until the first week in February, and the attendees did not waste their opportunity. Over 150 individual and group comments were sent to the conveners. Not so coincidentally, prominent BAT employees seem to have remained opponents of the change initiative.

Controlling Bureaucracies by Keeping Careerists Underfunded and Understaffed

In much of the public-sector literature, the career service is pictured as technologically aware and interested in change, whereas political appointees are usually portrayed as short-timers of limited program knowledge, if not actually committed to the view that the best way to control bureaucracies is to keep careerists underfunded and sufficiently understaffed so as to be unable to perform at excellent levels. This is the cost, some propose, of inhibiting bureaucracies from restricting the liberties of the citizenry. For an overview, see Golembiewski (1985, pp. 150–185).

Adherents to this viewpoint will find major surprises in ETA’s political appointees—see Figure 13.1, including the Deputy Assistant Secretaries. They were at times too busy with their own political responsibilities to be of help to the
Large-System Change in Government

careerists, to be sure; but at many critical points they supplied the key impetuses supporting changes in ETA’s culture and structure. In addition, the basic internal opposition came from several career bureaucrats, in key positions, who did not accept the proposed changes and were able to create enough active and passive resistance to delay the proposed changes long enough to cause trouble. Careerists may be largely driven by needs to maximize program performance, but this may leave them feeling that a specific change will reduce their own performance even though it may be the case that overall performance would be improved.

Relatedly, competing interests of individual careerists often overrode public and organizational interests. Certainly, not all evidence supports this point, but it seems instructive that JC and BAT careerists were the most vocal in their opposition to the restructuring effort. Although they had representatives on the FRT and SDT, the majority of the written comments came from JC. Most comment was virtually identical, emphasizing two messages: that the current structure “works fine,” and “if it ain’t broke, don’t fix it.” Several BAT associates also claimed they had complained to their congressional representatives, after Assistant Secretary Barnacle announced his resignation. These associates not-so-covertly took “credit” for stopping the proposed changes, at the point where they felt a zone of relief if not protection from overhead controls.

It Is Convenient, Perhaps Even Ideal, That Any Major Adverse Personnel Actions Take Place Before Any Participative Program of Change Begins

For relatively simple change efforts, the rationale is direct (e.g., Golembiewski, 1985, pp. 235–251). Basically, such action-taking clarifies goals, avoids the possibility of building opposition for persons or projects already ticketed for other fates, does not burden the change program with adverse actions that were fully justified on their own, and may reduce subversion to developmental processes to serve ends already decided but unannounced.

For projects like ETA, the advice can be more limiting, however, even counterproductive. Thus, the present purpose might have had elements of moderating or precluding further congressional cutbacks by a strong justification of present staffing levels, more or less. Moreover, from another point of view, waiting for Congress to specify adverse personnel actions could have contributed to forfeiting the possibility that ETA could impact Congress rather than the other way around. At the same time, however, being proactive—or even “entrepreneurial”—had its costs, real or imagined. “Getting out in front” can be dangerous, for example, and might encourage the more active participants to variously “play both ends against the middle”—to temporize, be facilitative in narrow senses.
only, to adopt weak roles in interpreting information, and in generating paralysis-by-analysis—to present a numbing array of alternatives rather than synthesizing one or a few alternatives.

Nobody can say for certain, but evidences in ETA suggest overcaution, if not conscious delay, that might not have existed had Congress “dropped the other shoe,” before the restructuring began gearing up. Suggestively, the report of the Structural Design team, itself chartered by the Reinvention Team, saw itself as “not a decision-making team, but rather an options development body” (ETA, 1996, p. 1). Moreover, the SDT Report contained twelve options, with one of them reproduced in Appendix 13.A to give the reader a sense of tone and coverage. Please note that readers should not be too hard on themselves if not all the elements in Appendix 13.A’s example are crystal clear. That is understandable and may have been intended by the mischievous. Of course, even the apparently arcane if not arch may reflect ETA shorthand for cultural and operating features and here are not sufficient to identify how SDT approached its difficult task. The full report has seventy-one pages, and Appendix 13.A presents six of them.

The ETA processes did not respect guideline 5, and that exposed the dynamics in ETA to complex pushes and pulls—some expected consequences and others that were unexpected.

**Cultural Change Is a Long-Range Business**

The shoptalk on one final point is all-but-unanimous: cultural change is a long-range business, and there is no point in trying to rush any aspect of a total effort. You cannot take too much time, then; but you can easily hurry a project and create avoidable resistance. The guides for action are direct: bend over backward to give everyone multiple opportunities to participate, gain information, contribute, and buy-in. This serves OD values, and may deprive some opponents of a rationale for resistance; and it minimizes the risks of creating new opposition. This view seldom stands as a front-and-center rule-of-thumb for planned change, but represents more of a common attitude in the literature. It gets voice in such injunctions: If you can take the time to do something over, as often happens, why not just take the time to do it correctly the first time? “Correctly” here refers to a high degree of participation, involvement, and buy-in, all of which can take time that may seem to some to threaten timely action.

In the public sector, however, such a view can have unexpected consequences. Especially given the typically short tenure of political appointees, as was the case in ETA, simply “taking all the time necessary” may miss an open window of opportunity. In ETA, pretty clearly, several political appointees were not able to accomplish their goals because of brief tenures. This case study supports the point that some public administrationists see as a central problem in
Large-System Change in Government

public management (e.g., Golembiewski, 1985, esp. pp. 318–334). That is, cultural change requires a long-term commitment that often takes several years to accomplish. However, the tenure of political appointees in federal government is often fleeting, which makes any substantial change a more difficult challenge than in most business settings. Short window of opportunity may exist in business contexts, of course, but they seem to most observers less frequent and less dramatic there, in general.

If this line of argument is correct, the prime implication supports high careerist championing. The concomitant danger is that the careerists “get ahead” of the political appointees, which raises practical problems as well as difficulties in principle of the political control of management.

RECENT DEVELOPMENTS

Such unexpected events aside, however, ETA’s effort at change was successful in gaining congressional approval, at the eleventh hour, as it were, if not at 11:55 or even 11:59. The full description of this outcome suggests that the design of planned change in ETA had developed foundations sufficient to withstand even the major unexpected events described above, given a bit of ingenuity by agency officials, building on the foundations of what to some (and your author!) was a fine OD design that was a victim of common serendipities associated with delays that in the aggregate were extensive enough to allow unfortunate events to overcome a carefully developed process of participation and consensus-building.

It may be helpful to review matters before moving on to developments during the first three months of 1999. There is no better day to start than August 7, 1998, when the Workforce Investment Act of 1998, Public Law 105-200 (WIA), was enacted. This law abolished the Job Training Partnership Act and created State and Local Workforce Development Boards to provide for the administration of various job training programs previously administered by ETA. WIA also contained this section:

Reorganization: Not later than 1 year after the date of enactment of the Act, the Secretary of Labor shall reorganize and align functions within the Department of Labor and within the Employment Training Administration in order to carry out the duties and responsibilities required by this Act (and related laws) in an efficient and effective manner.

This section was written by a senior ETA manager in the “failed reorganization,” in the context of the six lessons above. Your author saw this paragraph as strategic because of its underlying belief that a public agency probably will be more likely to make change when a legislative mandate exists. In effect, WIA settled for a broad legislative mandate, to work out the details in the context sketched above, which had focused on highlighting alternatives.
Upon enactment of WIA, ETA’s management contacted the National Council of Field Labor Locals (NCFLL), which is the union that represents all regional bargaining-unit employees. Local 12 represents all national office bargaining unit employees. They reactivated the two dormant Partnership Councils in order to accomplish the ETA reorganization in a collaborative mode. The two councils established a National Office Consultation Team as well as a Regional Office Reorganization Team to design a new organizational structure. The National Office group consisted of managers and bargaining unit volunteers from each program office in the national office. The Regional Reorganization Team was composed of six NCFLL representatives on the Partnership Council, and also six regional managers from mainline ETA, Jobs Corps, and BAT. Both teams completed their reports and submitted them to the Assistant Secretary for ETA on 10/31/98, as detailed above.

In early January 1999, Assistant Secretary Ray Bramucci announced the acceptance of the recommendations of both groups to reorganize ETA into four product lines: Office of Job Corps and Youth Services (JC/YS), Office of Adult Services (OAS), Office of Apprentice Training and Employer and Labor Services (ATELS), and Office of Workforce Security (OWS). In addition, the number of regional offices was reduced from ten to six, with four affiliates. Each region had a Regional Directorate overseeing the day-to-day operations of both regional and affiliate offices. The product line concept had many inspirations, no doubt, but the two authors were delighted with the choice. Emphasizing product line theory was a major element in the rationale that added RTG to the OD effort, at CM’s recommendation. In fact, the latter’s major presentation to the change team highlighted attractions of product or division structures like those introduced in Chapter 5.

A National Office Consultation Team was established to define the new roles of each of the national program offices, as well as to sketch their interactions. Six regional reorganization teams were created to determine the functions and staffing patterns necessary to accomplish those functions for each of the four product lines, the Regional Directorates, and the administrative functions in each region. Each of the teams consisted of management and regional bargaining unit volunteers. Each team was a self-directed work team that worked for two weeks in various parts of the country exclusively on developing a reorganization plan for each of the product lines.

The recommendations of each team were compiled into a “blue book.” The blue book was reviewed and accepted by the two Partnership Councils and a final recommendation was made to the Assistant Secretary, who accepted the recommendations in the blue book in their entirety. The final plan was negotiated with the NCFLL and was presented to Congress during the first week of August 1999. Congress accepted the final negotiated reorganization plan, and ETA is in
the process of implementing that plan with completion scheduled no later than the end of March 2000.

Overall, then, the Labor experience can be accounted a mixed success. Paramountly, the earlier flurries of temporary structures reflected a regenerative character that seemed worth saving in end-of-game play. At the same time, the products of that process did not seem to prevail and, indeed, at times seem intended to have been deleterious efforts to keep the processes of change open by more or less deliberate reticence to take or recommend specific options. The catalog of twelve structural options presented toward the latter stages of the case perhaps best reflects this emphasis on useful processes rather than one or a few products of choice, for good or ill. At the end, development of an alternative dominated.

APPENDIX A  OPTION 6 FROM STRUCTURAL DESIGN TEAM REPORT

This option (Structural Design Team Report, p. 48–52) focuses on restructuring around two major groups—adults and youth—and assumes the differences in the service needs of adults and youth will shape the organizational structure and choices. This option consolidates into a new *Office of Adult Workforce Development Services* programs previously operated under JTPA Titles II-A and III, TAA, NAFRA-TAA, MSFW, INA, BAT and those functions under ES, UI, Worker Profiling and Reemployment Services, and the One-Stop and LMI initiatives as these programs apply to adults. Programs previously operated under JTPA Titles II-B and II-C, Job Corps and those appropriate portions of the ES, UI, WPRS and the One-Stop, LMI and STW initiatives applicable to youth would become components of a new *Office of Youth Workforce Development Services*.

This approach recognizes the need for subject matter specialization in order to build the capacity of National Office to be an “expert resource” for the regions on a broad range of workforce development issues and it also enables regions to be a valued resource to States on policy and operational issues. There are two organization charts (Diagrams A and B) for the customer-based model, distinguished as follows. Diagram A assumes that national policy functions are organized discretely and are housed in an Office of Policy that directly supports and reports to the Assistant Secretary (see Figures 13.A–13.C). This variant also puts Foreign Labor Certification in the Office of Regional Operations. The other customer-based model locates national policy functions in each major office (e.g., Office of Adult Workforce Development Services). Foreign labor certification is placed in the Office of Adult Workforce Development Services.

In the National Office, two Deputy Assistant Secretaries (DASS) would head Adult Workforce Development and Youth Workforce Development offices,
FIGURE 13.A National office organization customer-based integration strategy (discrete policy function).

FIGURE 13.C Regional organization.

respectively. These two offices would be established to focus on the needs of the two groups served by the Workforce Development Administration—adults and youth. In addition, this option, at the national level, would:

Centralize policy development, legislative affairs, R&D, performance management and administrative support (e.g., personnel, budget and accounting, grants and contract management support, financial and management information systems and records management); and

Include specialists at the national level—training, adult, youth, reemployment, actuarial, administrative-related and workforce development equity—knowledgeable in the best reemployment and training practices for their respective groups for true program integration.

The regional administrator’s role would be enriched with the RA’s authority enabling them to deal in an empowered way with the States and other partners.

Regional Offices would consist of three organizational units:

A Technical Support Unit internally organized around specialty or areas of expertise, i.e., youth, adults, reemployment, training, employer services, etc. Unit staff would assist any State in the region as consultants on workforce development policy and operations. This unit would house area specialists similar to those at the national level on adults, youth, reemployment services, training, etc., but not necessarily would each
region have specialists for all topics and some specialists could “cover” more than one region.

A **Workforce Development Unit** would provide technical services on administrative, programmatic and legislative matters and be the primary point of contact with one or more specific States. Customer satisfaction feedback (survey design, execution and analysis) would be located in the workforce development unit. And, since this unit would house the primary points of contact with the States, it would also possess the following responsibilities: operational decision-making, waiver authority and grant/contract officer authority.

A Foreign Labor Certification Unit.

**Strengths**

Matches with the evolving delivery system (it “resonates”).

Focuses on the ultimate customer

Divides workload into generalist, State-focused tasks and specialist technical tasks.

De-emphasizes program distinctions.

Integrates functions around the needs specific to each major customer group.

Aligns priorities around customer-specific needs.

Aligns with the Assistant Secretary’s office by putting RA’s in a leadership position, enabling them to deal effectively with Governors.

Matches the State-level organizational approach to workforce development services in many cases.

Decisions tailored to reflect State frameworks and grounded in recognition of customer needs.

**Weaknesses**

Increases the difficulty of the federal partner’s work as long as separate programs exist.

May encounter strong political head-winds that resist the integration of BAT, Job Corps and VETS services.

Initial “capacity gaps” are likely.

**REFERENCES**


ETA (September 29, 1995). Memorandum for all ETA employees, “Laying the Foundation for a New ETA Organization.”
Skills and spirit must be blended in nursing, and much evidence indicates that today’s challenges threaten to overwhelm coping competencies and resources. Can efficiency be increased, and can care-giving costs be stabilized or decreased in a patient-centered mode, without diminishing the quality of nursing services and also without rubbing nurses emotionally raw? This chapter justifies optimism about an affirmative working answer to this question by responding to ubiquitous role problems in nursing, while reducing costs and enhancing the quality of nursing performance as perceived by major stakeholders—patients, their families, nurses, and physicians.* The major interventions are well known in Organization Development, or OD: large-scale surveys; action-planning with a design group; developing supportive norms and interpersonal processes; and redesigning the structure and policies/procedures as well as interaction in a specific worksite. Such emphases qualify the interventions as broadly socio-technical. The positive results are especially attractive because they derive from a setting that was initially in decent shape before interventions began.

The details available here make this chapter virtually unique in the literature from the earliest days (e.g., Marrow, Bowers, and Seashore, 1957); that is to say, similarly intentioned applications have existed, but without the conceptual and operating details available here. The detail here is justified in two senses: the critical character of the health care focus and the rarity of analyses that detail the new pattern’s activities—and their reinforcing cultures and structures—necessary to describe the designs and their effects. For recent examples of the few demonstrations of the instrumental changes required of normative reeducative strategies like OD, see Hennestad (2000) as well as Liphity and Popper (2000).

**SYSTEM REDESIGN IN NURSING: ACTION PLANNING IN A MEDICAL-SURGICAL WARD**

The first part of this chapter has four emphases: it outlines some common conditions and concerns that plague nursing; it sketches measures and methods for assessing the degree of movement from the initial to a desired condition; this chapter also details the interventions for change along with values and theory underlying them, and it concludes with a number of pre-test features and comparisons that motivate change in the 5-East ward at Sutter Hospital in Sacramento, CA.

The rest of this chapter will detail the results of a change in nursing care. Here, attention is directed at the ways the intervention unfolded, especially concerning redesign of 5-East Medical-Surgical Oncology. Moreover, pre- vs. post-test comparisons occupy the bulk of the latter portions of this chapter. Both self-report and archival data are used, and three major stakeholders are represented—patients, nurses, and doctors.

**Nursing’s Common Conditions and Concerns**

Without doubt, nursing deserves attention, as four points briefly establish. First, nursing seems beset with major role problems, nationwide (e.g., Hay Group, Inc., 1989) and perhaps worldwide. An executive of the hospital hosting this action research isolates a central issue in this role ambiguity in her own health organization:

. . . Clearly the staff registered nurse is one of our most gifted resources in providing patient care. Indeed, many of our allied health professions are outgrowths and/or extensions of nursing care. Unfortunately, as we’ve had to address issues of regulation, cost-containment and competition, we inadvertently over-loaded the registered nurse with ‘tasks’. Once ‘task bound,’ the ‘work itself’ inhibited innovation and job satisfaction . . . while simultaneously increasing frustration. This seemed to get progressively worse each year. Somehow the cycle had to be broken.
The executive also had a remedy in mind: “We need to support the nursing unit staff in examining those phenomena and [to] empower them to initiate the needed changes.”

Second, reasonably, role issues will be most pressing in areas like surgical oncology, feared by patients as well as trying for nurses. The latter are often in the position of “just taking” the stressors generated by the former, for obvious reasons.

Third, major problems complicate each stage of supplying medical-surgical nurses. Thus, training presents special difficulties, as do recruiting and retention of competent nurses.

Fourth, in conventional hospital settings, surgical nurses often feel isolated and alienated from fellow professionals, reinforced by strong one-to-one linkages with patients. Thus, nurse ↔ physician relationships may be difficult; and communication may be uneven between various cohorts of health-care providers—for example, registered nurses (RNs), licensed vocational nurses (LVNs), or nursing assistants (NAs). Perhaps paramountly, nursing professionals tend to get socialized in ways that emphasize the individual and his or her skills. As one RN noted, “. . . to a large degree I function as an independent practitioner.”

**Philosophy, Basic Model, and Worksite Features**

There are numerous possible approaches to these common conditions and concerns, and more than one may “work.” This effort has dual roots. It is firmly nestled in a specific work setting, with its own features and causal textures, and with a conviction that—in the words of one local health care provider—“Providing an environment in which nursing staff are involved in professional practice issues will not only enhance employee/manager/physician satisfaction, but should have a direct positive impact on the quality of patient care.” In addition, this effort is rooted in Organization Development, or OD. Both rootings encourage an eclecticism in the service of specific needs and, where appropriate, major borrowings will be acknowledged.

To be more specific, this approach to improving worksite conditions for one hospital’s nurses will be detailed in three ways. In turn, the focus shifts from overall philosophy, to underlying model, and finally to the specific setting for intervention—a medical-surgical oncology unit, 5-East, at Sutter Hospital, Sacramento, Calif.

**A CONCEPTUAL NICHE FOR IRONY V, PART 1**

Some historical perspective here will help. If the interventions of interest here had been made in the 1960s and 1970s, there seems no doubt that the emphasis would have been on such guiding models as Argyris’ (1957) dimensions for self-
actualization. In sum, as Exhibit 14.1 implies, individuals tend to be equipped with a specific kind of guidance system—if you will, with a gyroscope that steers people toward the desired and desirable, when the individual has a meaningful choice in deciding on direction.

Nowadays, such conceptual groundings receive less explicit notice, but this writer sees them as still useful. Rigorous research has not been able to define and refine such models, for purposes of high validity and reliability. At the same time, however, the dimensions provide clear guidance about which ways the winds of change are blowing, and why. The reader can usefully occupy self with tracing the effects of how the design is fitted to the diagnosed situation.

OVERALL PHILOSOPHY

A first way in which the Argyrian dimensions can be said to underlay the present intervention was birthed in a corporate-level task force at Sutter Health, which later became the “steering group” for the project. Building on the findings of the several surveys of nurses (e.g., Rountree and Weber, 1991a), task force members sought approaches to enhancing the working experiences of bedside care-providers. Their basic focus was on a bottom-up approach to planned change, as these propositions reflect:

Work redesign should be highly decentralized to the manager and nursing staff at the unit or bedside level.
Nursing participation in the process is critical to success.
Specifics and priorities to address, as well as the approach for pursuing identified goals, should be determined by the nursing staff, in accordance with their assessment of their unique needs and environment.
The manager and staff should be encouraged and supported in experimenting with the pursuit of identified goals that are consistent with the broad objectives detailed above.

---

**EXHIBIT 14.1** Argyrian Dimensions for Self-Actualization

<table>
<thead>
<tr>
<th>Passivity</th>
<th>Increasing activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependence</td>
<td>Growing independence and interdependence</td>
</tr>
<tr>
<td>Limited behaviors</td>
<td>Widening range of behaviors</td>
</tr>
<tr>
<td>Short time perspective</td>
<td>Lengthening time perspective</td>
</tr>
<tr>
<td>Subordinate position</td>
<td>Peer or superordinate position</td>
</tr>
<tr>
<td>Lack of awareness and control</td>
<td>Growing awareness and control</td>
</tr>
</tbody>
</table>

*Source: Based on Argyris, 1957.*
Organizational learning should be shared, as appropriate, consistent with maintaining patient confidentiality.
Changes impacting budget will need administrative approval.

These philosophic underpinnings have clear roots in OD, seen as having three basic foci, all value-loaded: on interaction, on structure and roles, as well as on policies, procedures, and practices (e.g., Golembiewski, 1979, Vol. 1; 1992). Consequently, the approach clearly represents a broadly sociotechnical perspective. Basically, performance problems develop more from ineffective systems than from individual performances.

The model underlying the present intervention is sketched in Figure 14.1. To make the obvious connection, the overall philosophy and values are reflected in five central aspects of the bedside environment: the work; the physical context within which the work is done; the cultures and climates of relevant work groups; the broader organization, especially in the character and quality of overhead supervision; and the various applicable policies, practices, and procedures, which will include organizational inputs as well as those directed as a profession by way of law and traditions. Figure 14.1 also distinguishes three panels of outcomes.

A METHODOLOGICAL CONTEXT FOR IRONY V
A few comments serve to locate this chapter in a methodological context for Irony V—the broad success of OD without a corresponding emphasis on large
system change. This chapter adds to the impetus on this theme as it impacts on OD. First, this chapter stands with Lewin in his observation that only a good theory will lead to the generation of useful analogs in practice.

Second, the nursing context is an important setting for such a demonstration, probably a uniquely salient one. This is no easy piece, but its challenges are attractive because many direct spin-offs will be possible.

Third, this chapter reflects the strategic usefulness of combining attention to interaction, policy, and procedures, as well as to structure. “Success” in less-demanding settings often can result from narrower interventions.

Fourth, the potential for increasing success rates, as well as expanding the reach-and-grasps of OD, seem substantial. Its gross and subtle features not only might reduce failures in similar settings, but also could increase the capacity of the target population for impactful OD. Just such double-plays will enhance OD and its consequences.

Fifth, this chapter illustrates which intervention skills and cultures can be applied in technologically demanding work settings, and how.

MEASURES AND METHODS

Three emphases provide sufficient detail to allow the reader to evaluate what was measured and, to a degree, why; and these measurement activities provide a sense of what OD values and why. The first emphasis relates to data about seven of the domains in Exhibit 14.2, and the second deals with an eighth domain—Employee Performance. The third emphasis below deals with analytic methods.

These multiple measures serve several uses. Thus, all Sutter General nursing staff were surveyed, which provides a benchmark of the state of local nursing. Later, as the discussion surrounding Exhibit 14.3 will detail, these initial results provide a pretest for estimating the effects of the present interventions in a selective target.

Worksite Features

The 19-scale ADVENT questionnaire developed by the Hay Group, plus several other self-reports, broadly characterize nursing worksites. In all cases, the items and scales focus on relatively specific arenas amenable to positive influence via personal development, training, or coaching. Consequently, data from them can help assess the need for specific interventions, as well as to estimate the effects of interventions. These 31 measures relate to the first seven domains in Exhibit 14.2 and are labeled there in ways that should identify their general content.

Performance Measures

Eight estimates are relied upon to assess this crucial domain. They relate to: aspects of nurse burnout; relationships of nurses with physicians; patient satisfaction;
**EXHIBIT 14.2 Multiple Measures in Eight Domains**

**The Work Itself**
- Autonomy
- Job content
- Meaningfulness
- Motivation potential
- Responsibility for outcomes of work
- Role clarity
- Satisfies needs that motivated one to become nurse

**Workload**

**Physical Environment**
- Physical working conditions
- Unit layout/design

**Work Group Environment**
- Communication within immediate workgroup
- Cooperation
- Openness
- Psychological sense of community
- Trust
- Support from other departments

**Organizational Environment**
- Availability of supplies/equipment
- Communication between departments
- Communication between shifts
- Facility images
- Management style
- Nursing leadership (in understanding problems/needs of caregiver)

**Policies, Practices, and Procedures**
- Career advancement opportunity
- Career earnings potential
- Compensation
- Growth opportunities
- Pay-for-performance
- Promotional opportunity

**Employee Satisfaction**
- General satisfaction

**Employee Motivation**
- Attitude about quitting
- Opinion of self

**Employee Performance**
- Work sampling
- Utilization of professional knowledge/skills
- Emotional exhaustion as burnout surrogate
- Nurse/physician relationships
- Patient satisfaction (QCI Index)
- Physician satisfaction (TMPE Index)
- Use of call light by patients
- Costs of nursing services


**EXHIBIT 14.3** Major Pre-Experimental Activities, Days 1–159

<table>
<thead>
<tr>
<th>Approximate Dates</th>
<th>Description</th>
</tr>
</thead>
</table>
| Days 1–7          | Pretest, Basic Diagnostic Data  
|                   | Work Sampling  
|                   | Survey of Worksite Features (all domains in Exhibit 14.1) |
| Day 40            | Feedback of Diagnostic Data  
|                   | Feedback to all staff at Sutter 5-East, E and C (4 hours), in three aggregates assembled to be convenient to nursing staff |
| Days 41–159       | Action Planning  
|                   | Cognitive/skills input to E participants (4 hours) |
| Days 160–189      | Experimental Period |

physician satisfaction; use of call lights by patients; costs of providing nursing services. They deserve some explanation.

**Work Sampling: Utilization of Professional Knowledge/Skills**

Work sampling quantifies the amount of nursing time in the average workday spent on specific activities. Data for this categorization of activities were collected over a 7-day period in November 1990 and come from self-reports whose reliability was supported by on-site observation. The categories sought to balance two factors. The information value of the categorizations increases as the categories become more specific, but increases in specificity and in the number of categories can be costly, as in the difficulties of recall as well as of recording data. The present system was developed by an external consultant, and can be schematized in these terms:

- Direct care—to the patient (29 activities)
- Indirect care—for the patient (18 activities)
- General—for the unit/staff (20 activities)

Of special usefulness, work sampling permits an estimate of the degree to which work done by specific kinds of nurses permits them to utilize their differential training and knowledge/skills. In effect, these categorizations help profile role performances by nursing staff, and permit perspective on possible role conflicts.

**Emotional Exhaustion and Burnout**

Both normal nursing as well as stressors generated by interventions could generate strain, hence the interest in burnout. To explain, in part, the implementation of any action plan was seen as having real potential for such effects:
Increases in cognitive complexity of some nursing jobs, especially for RNs; and

The need for major increases in social-emotional skills for all nursing staff—e.g., in early surfacing of potentially conflictual issues versus traditional avoidance of conflict.

Hence, the reliance here on items tapping “emotional exhaustion,” or EE, as one estimate of employee performance. Items from the Maslach Burnout Inventory (Maslach and Jackson, 1982, 1986) are used here to estimate EE.

Most observers (e.g., Maslach and Jackson, 1982, 1986) agree that emotional exhaustion is a key component of burnout. Measures like the phase model of burnout would have been more inclusive (e.g., Golembiewski and Munzenrider, 1988), but much is known about the measurement properties of Maslach’s EE subdomain, and no doubt exists that advanced burnout is heavily loaded by it.

Nurse and Physician Relationships

Several items tap this significant domain. They were chosen to tap troublesome dynamics, and interventions are targeted to improve matters. Consider patient satisfaction. “Quality of care” is much emphasized nowadays, but definitions tend to differ, often profoundly. Here, the basic measure of nursing care derives from the self-reports of patients concerning the degree to which nursing services are perceived as:

Responsive to patient needs; and

Provided by staff who are caring, considerate, concerned, and competent professionals

The 6-item Quality of Caring Index, or QCI (Rountree and Deckard, 1984), estimates patient satisfaction. For the two months prior to the present intervention, a pretest benchmark is provided by a representative sample of patients who were telephoned within twenty-four hours of discharge from the hospital and asked to respond to QCI items. This pretest sampling was devised to extract a sample of responses with a 95 percent probability of representing the true perceptions of quality in the total patient population.

For the post-test, all discharged patients from both Experimental and Control conditions were telephoned and asked to respond to the 6-item QCI. Immediately following QCI responses, newly discharged patients also made the open-ended comments and observations about their care.

In addition, the reactions of physicians were estimated by self-reports to The Medical Practice Experience, or TMPE (Rountree and Davis, 1990). The pretest was administered one week prior to the experiment, and post-test data were obtained in the week following the 30-day experimental period. The response rate by physicians was high for both pretest and post-test. For example,
responding physicians accounted for more than 85 percent of all admissions to 5-East during the experimental period.

An unobtrusive, objective measure of patient care also was provided by use of call lights by patients. Patient call lights were monitored electronically, and they provided another assessment of the quality of services reflected in the unmediated responses to patients.

Costs of Nursing

Finally, costs of nursing services are estimated, pre- versus post-test. Corporate archives permit a dollars-and-cents estimate of comparative nursing costs under experimental and control conditions. Proprietary considerations preclude sharing direct dollar costs, but ratios of annualized costs will provide a revealing picture of relative costs.

Analytic Methods

Overall, analysis often will focus on standard scores, and for obvious reasons. Primarily, standard scores on all measures—arrayed from +1 to −1—facilitate multiple comparisons: between E and C conditions at 5-East; between various nursing units and the national surveys conducted by the Hay Group (e.g., Rountree and Weber, 1991a); and between 5-East and other locations within Sutter General and other locations.

One-way ANOVA usually will be relied upon, despite the obvious ineligance. As is well known, many of the present measures are intercorrelated and multidimensional. Rather than attempting to overpower such features of the data set with other methods, the approach here will be threefold. The reliability of all self-reports will be established by Cronbach’s alpha; several objective or “hard” measures are utilized; and the consistency of multiple measures will be central in estimating effects.

EARLY ACTION-PLANNING AND SOME DYNAMICS

Exhibit 14.3 outlines the pre-experimental design for 5-East, focusing on the gathering and sharing of data that will be taken into account in action-planning. Three emphases below elaborate on these pre-experimental activities: design issues; behavioral guides for the processes of action-planning; and a sense of the dynamics of action-planning and their early products. Details about the interventions that resulted from later action-planning also are provided. Together, both early and later planning—to expand on Figure 14.1 and Exhibit 14.2—intend to foster change in the values in a specific work setting, which would then impact on the Work Itself and the Work Group Environment. In turn, these value-loaded effects would positively impact on Employee Satisfaction and Employee Motivation. Then, in turn, various facets of Employee Performance would improve.
In this approach, note that three of the domains in Exhibit 14.1 do not get extensive intervention, especially because of narrow time-constraints. These domains are Physical Environment, Organizational Environment, as well as Policies, Practices, and Procedures. Pre- versus post-test comparisons in these domains can serve two purposes: to help test for a “halo effect” in arenas beyond the intervention and to provide confidence that experimental improvements did not produce awkward effects—for example, poorer communication between shifts.

**Design Issues**

The design involves a simple pre-test versus post-test, with a control (C) that is a duplicate of the experimental (E) unit, in a medical-surgical ward. 5-East is a rectangular ward with about 40 beds, and was divided lengthwise and equally into E and C conditions. Management agreed to keep the patient census at equal and moderate levels in both E and C subunits during the full period of the experiment, and also committed to balancing the acuity of patients assigned to the two conditions.

The lengthwise division helped keep important features more or less equal in both E and C, more importantly, but also allowed open observation by all parties interested in what was going on, as well as avoided the encapsulation of an “experiment” that often occurs. Absence of such openness can become an obstacle in subsequent diffusion of learning. “That may have happened there,” goes this resistive refrain, “but it can’t happen here” (e.g., Walton, 1977). This commitment to publicness appears in other forms. For example, both E and C were exposed to some parts of the treatment—e.g., diagnostic data were fed back to all 5-East staff.

5-East became the experimental site in a direct way. Interest in change existed there; and management considered that unit, if not necessarily representative of other units in Sutter or elsewhere, as constituting neither the easiest nor the most recalcitrant encountered in nursing settings (e.g., Hay Group, 1989). Right or wrong, the concern at 5-East was in facilitating subsequent diffusion rather than inhibiting contamination. Moreover, on balance, these arrangements permit robust interpretation of any observed differences. For example, any “contamination” would tend to reduce E versus C differences.

A basic part of the treatment also can be viewed as “contamination.” Consistent with the operating philosophy detailed above, volunteerism determined E selections. This constitutes intended contamination, as it were, and is one of the major ways in which “action research” differs from “straight science” (e.g., Susman and Evered, 1978). The values guiding the latter seek to preclude any “contamination,” as by random assignments to treatments, compartmentalizing data, and so on.
EXHIBIT 14.4  Confrontation Ground Rules

Purpose is to “fix” problems, not to blame or retaliate.
Acceptable timing for all parties.
Place: neutral, comfortable, and private.
Descriptive rather than evaluative.
Specific about behaviors.
Both parties are committed to stick to it—to find a solution acceptable to both.
Listen, listen, listen to others (paraphrase to check understanding).
Confidential, in the sense that all parties agree to discuss only agreements with others.
No badmouthing, before or after.
All parties decide whether to use a mediator, and, if so, who.
Parties commit to record and review specific time.

Some Agreements for Guiding Process

At several stages in the feedback and action planning, participants agreed to normative guides for their behavior. Consistent with the goal of regenerative interaction (Golembiewski, 1993; 1979), participants agreed to “confrontation ground rules” to induce more effective communication. In brief, change would be facilitated by direct communication, and might be impossible without it. Exhibit 14.4 reflects both techniques and policies for appropriate interaction. The term “all parties” usually refers to a pair, but more actors can be involved.

E participants also entered into a broad partnership agreement, participatively developed. Exhibit 14.5 details this important normative compact to guide action planning and its implementation. Exhibit 14.5 reflects another way in which the E group sought to give a specific direction to the work culture via norms that were intended to drive interaction. In a variety of ways, Exhibit 14.5 seeks to approach regenerative interactions, and to anchor it in a relatively formal “partnership agreement.”

Basically, Exhibits 14.4 and 14.5 signal the end of “business as usual.” Perhaps the most overt sign of this intent was the dramatic publicness about E effects. To reinforce effects for themselves and information for others, for example, E participants typically posted products of their interaction on posterboards in prominent locations visible to C patients and personnel.

Dynamics and Products of Early Action-Planning

As noted in connection with Exhibit 14.3, the basic diagnostic data from the pre-test survey were fed back to all staff in three separate meetings that lasted about 4 hours each. The feedback sessions were arranged at convenient times for the
EXHIBIT 14.5 5-East Experimental Team Partnership Agreement

To support a positive work environment I pledge that:
I will not expect people to be perfect.
I recognize that each individual is unique with varying resources, abilities and demands.
I will talk directly to the person with whom I share a problem.
I will avoid the three Bs—Bickering, Backbiting, and Bitching.
I will give and expect strong peer support.
I will believe in me and I will forgive myself when I am not perfect.
I will try to separate the problem or issue from the person.
Everybody has the right to support his/her values. I value your disagreement as much as I value your agreement.
I won't displace responsibility, blame, or be overly critical.
I will challenge only with an intent to improve. I will use compassion in my confrontations.
If you do anything that puzzles or offends me, I will tell you when it happens. I ask you to do the same for me.
I believe in having fun at work. Do not confuse my lack of soberness with not taking what I am doing seriously. I just do not take myself seriously.


staff, and each session was facilitated by a Sutter consultant. Overall, the purposes centered around inducing a climate featuring regenerative interaction—high trust, low risk, high openness, and high owning (e.g., Golembiewski, 1993, 1979, Vol. 1, esp. pp. 61–65;). More specific goals are generic to survey/feedback (e.g., Golembiewski and Hilles, 1979), and include:

Creating a common pool of information about work experiences, much of which had never been publicly shared or organizationally sanctioned
Collaborative diagnosis: e.g., identifying problem situations; uncovering conflict and its causes; and so on
Identifying possible sources of resistance to any changes
Surfacing additional data at operating levels
Providing benchmarks against which to estimate effects
Empowering action-planning and implementation at operating levels
Meeting short-term needs while enriching longer-term potentials—for personal and professional growth, promotion opportunities, and so on.

These feedback sessions involved all staff in public exchanges, and two emphases dominated: comparisons of 5-East with other Sutter General units and a consistent pattern of different perceptions by roles, especially Registered Nurse (RN) versus Licensed Vocational Nurse (LVN). The first emphasis derives from
the pretest job diagnostic data, with standard scores facilitating multiple comparisons. The latter emphasis rests on Work Sampling data.

**SOME COMPARATIVE PERSPECTIVES ON Es VERSUS Cs**

Overall, 5-East staff saw their workplace *more* favorably than the average nursing unit in Sutter General which, basically, makes significant change more difficult. On average, to illustrate, the 5-East nursing staff sees its own situation as surpassing other Sutter nursing units in that 5-East reports:

- More consistent support and assistance at the bedside from other hospital departments;
- Better working relationships and communication with staff on the other shift;
- Having a more participative supervisor, who communicates clearly and makes fair and equitable decisions;
- Nursing managers and executives who are in touch with the problems, needs, and expectations of bedside care-givers; and
- Pleasant, professional, and moderately effective relationships with physicians

Only Sense of Community (SOC) breaks this pattern. 5-East nursing staff saw themselves as less “tight-knit, interdependent, and dependable” than other nursing units in their hospital. For reasons both local and cosmopolitan—in the reactions of 5-East staff, as well as in available theory (Golembiewski, 1962b, pp. 149–170, 265–270)—the low levels of SOC attracted attention. Thus, three subscales are involved (Rountree and Deckard, 1984):

- Primary Work Group Cohesion, whose items relate to the degree of pride, cohesiveness, and tightness of bonds;
- Supervisor Social and Emotional Support, whose items relate to supervisor’s role in energizing communication and support; and
- Coworker Social and Emotional Support, whose items relate to worker relationships—their caring for each other, and their availability when help is needed.

These three qualities are critical in assessing work group effectiveness, in general, and probably have a special salience in work settings like 5-East. Hence, they received intensive attention.

**Different Views of RNs and LVNs**

Along with such commonalities in perceptions, consistent differences emerged around the RN and LVN roles. The point could be established in many ways
Large-System Change in Health Care

but, overall, whether 5-East saw its work as more or less attractive than other nursing units, RN perceptions usually differed from those of LVNs. Thus, both categories of nurses saw SOC as a real problem, but LVNs were significantly more concerned. This situation seems common in nursing settings (e.g., Hay Group, 1989).

Three emphases usefully organize these interrole differences. Thus, some central themes get attention; then the focus shifts to the blurring of roles, and this focus on RNs and LVNs concludes by highlighting role devaluation.

Some Central Themes

As the several feedback sessions processed the survey data, several central themes emerged. These words of participants illustrate the emerging points of agreement:

“. . . for the most part I function as an independent practitioner. Some [times] there is a team [and some times] not.”
“I think that if an individual cannot count on all of those they work with, and we can’t or don’t, that’s a bad problem on a unit like this.”
“RNs feel like they can depend on one another and others in the unit and LVNs definitely don’t feel that way.”
“There are cliques on this unit, like on every unit. The LVNs don’t feel like they fit in the cliques at all.”
“A lot of the time an LVN won’t ask an RN for help . . . they think the RNs should know when they need help.”

Blurring of Roles

As in the nationwide Hay Group surveys, and in common with other Sutter locations, the 5-East nursing staff were experiencing cross-pressures about roles. The details are burdensome, but the bottom lines are clear. Work Sampling demonstrated clearly that RNs and LVNs often performed much the same activities, despite the greater educational attainments of the RNs. Specifically, the two role-sets were differentiated largely by the RNs’ performance of only five of the total number of activities, basically those reserved for RNs by the California Nurse Practice Act. These distinguishing activities accounted for only some 4 percent of RN total working time. Similarly, at 5-East as well as in other populations, LVNs spent much of their time doing activities that Nurses Aides are trained to perform (Rountree and Weber, 1991a, p. 9). See also Table 14.1 below. This role-blurring had many awkward consequences (Rountree and Weber, 1991a, p. 9) the more prominent of which observers note:

. . . because of stretched schedules and staffing, an LVN might have been responsible for all the routine care received by a patient, only to
find that when the doctor arrived, he or she asked to see “a real nurse”—[an RN] who doubtless knew little about the patient in question.

One manager summarized this failure to distinguish RN and LVN roles—in education, experience, and accountability. “Little wonder,” this manager concluded, “that the physician has difficulty dealing with the issue.”

Hospital policies exacerbated the obvious potential for dissatisfaction, and even for conflict. Thus, hospital executives announced a policy of upgrading the content of the LVN role, so as to expand its level of practice closer to the limits of the state Nurse Practice Act. This at once resulted from, as well as contributed to, a “role compression.”

Role Devaluation

What may be called “role devaluation” also was a prime concomitant of this widespread blurring. Specifically, nurses in the feedback sessions differentiated the various activities performed into two bundles—whether or not they required different levels of basic knowledge and skills. In sum, as Table 14.1 demonstrates, substantial proportions of both RN and LVN time were spent doing activities for which their specialized training/experience was not required.

The consensus view was that the pattern represented neither an efficient nor an effective use of valuable and limited human resources, and this is consequential. For example, the Hay Group studies conclude that utilization of knowledge/skills was highly and directly correlated with work satisfaction.

IN-PROCESS OBSERVATIONS

In sum, this sketches the situation targeted for action-planning by the nursing staff. The problems were not critical, but neither was the situation ideal. What remains? The emphases below are on the redesign of nursing work.

Overall, the intervention had multiple and positive effects in a nursing situation that is frequently encountered. Broadly, the cost of nursing services was re-

---

**Table 14.1** Activities Performed for Which Knowledge/Skills Are Required/Not Required

<table>
<thead>
<tr>
<th>Jobs</th>
<th>Required</th>
<th>Not required</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>53%</td>
<td>47%</td>
</tr>
<tr>
<td>LVN</td>
<td>60%</td>
<td>40%</td>
</tr>
</tbody>
</table>
duced by making better use of the skills/abilities in a “Shared Care” mode. In addition, the quality of the experimental (E) worksite was maintained or enhanced, in the views of all major stakeholders, on all measures; and the control (C) condition generated only random differences in the 30-day experimental period.

MAJOR CONSEQUENCES OF ACTION PLANNING

To become more specific, Exhibit 14.6 reviews the major activities in the design for change. The narrative below concentrates on the action planning and its evaluation, which extended from days 41 to 200.

Nothing was “broken” in E or C, but much could be improved upon. The building of shared information among nurses from a survey about the character of their work setting induced forces to reduce the gap between the actual and their ideal condition. As one registered nurse (RN) observed about the feedback of data from the survey:

I learned a lot about those I work with and [about] myself that day. I had no idea. Because we [the nursing staff] haven’t communicated as openly or as regularly as we need to . . . and because we don’t see one another much sometimes . . . I was astonished to hear—first hand—how some of the others experience their jobs on the unit so differently than I do . . . especially the LVNs [licensed vocational nurses].

Overall, the survey/feedback sessions highlighted such themes:

Concern for the low sense of community, the lack of enjoyment derived by many from working in the group, the heavy emphasis on individual practice, the lack of clarity between roles, and the presence of cliques;

**Exhibit 14.6**  Major Activities Associated with Intervention

<table>
<thead>
<tr>
<th>Approximate Dates</th>
<th>Pretest, Basic Diagnostic Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days 1–7</td>
<td>Feedback of Diagnostic Data</td>
</tr>
<tr>
<td>Days 40–45</td>
<td>Action Planning</td>
</tr>
<tr>
<td>Days 41–159</td>
<td>Pretest Data, E and C</td>
</tr>
<tr>
<td>Days 58–159</td>
<td>Patient Satisfaction</td>
</tr>
<tr>
<td></td>
<td>Physician Satisfaction</td>
</tr>
<tr>
<td>Day 160</td>
<td>Begin Experimental Period</td>
</tr>
<tr>
<td></td>
<td>Post-test, Patient Satisfaction, within 24 hours of discharge</td>
</tr>
<tr>
<td>Day 190</td>
<td>End Experimental Period</td>
</tr>
<tr>
<td>Days 191–200</td>
<td>Post-Test Data, E and C</td>
</tr>
</tbody>
</table>
Desire to enrich the work itself, which underutilized the knowledge, skills, and thus the capabilities of RNs; and

Need to enhance the relationship between nurses and physicians through improved availability of RNs, better communication, and taking care of business while physicians are “on the floor,” thus reducing the need for follow-up phone calls.

An action planning team met over some three months to detail appropriate interventions, building on the newly regenerative character of their interaction in the survey/feedback session. Twelve volunteers became an action-planning team—the Special Quality Action Team. SQAT met on hospital time, and had a broad missions to:

- Generate in-depth analysis of survey data;
- Develop a prioritized list of problems as goals for continuous quality improvement;
- Differentiate causes from symptoms;
- Formulate strategies for change that target priority goals;
- Implement or experiment with strategies; and
- Evaluate the impact of change strategies on all stakeholders, including patients, physicians, staff, and the employing organization.

The Special Quality Action Team featured four concerns: elaborating an alternative mode of service delivery, detailing appropriate role definitions, designing a new structure; and agreeing on a re-educative intervention. Consistent with their goal of increasing regenerative interaction (Golembiewski, 1993), their proposals sought to lower risk, and raise the levels of openness, owning, and trust. Detailed notes of meetings were kept and the products of discussion often were posted around the E site in full view of C. Not only did this approach share information, but the publicness also helped reinforce commitment to decisions and actions.

Beyond Interaction

Let me take the opportunity here to introduce a new emphasis at Sutter—basically, an emphasis on new structure and policy/procedures. Again, usefully, Argyris’ model for self-actualization—of tendencies that individuals prefer when they have a real choice—provides useful counterpoint to the description of effects below. Indeed, beyond the scheme in Exhibit 14.1, the Argyrian dimensions also imply specific organizational ways-and-means that can give concrete organizational force to the desired/desirables inherent in that model. Exhibit 14.7 details such ways and means that, in effect, constitute for the organization mission the mandates that people must accomplish in order to achieve what they deeply need to do.
EXHIBIT 14.7  Organization Guidelines Associated With Argyrian Dimensions for Self-Actualization

<table>
<thead>
<tr>
<th>Argyrian dimensiona</th>
<th>Organization guidelinesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing activity</td>
<td>Experience the totality of the organization</td>
</tr>
<tr>
<td>Growing independence</td>
<td>Increase self-responsibility and self-motivation</td>
</tr>
<tr>
<td>and interdependence</td>
<td></td>
</tr>
<tr>
<td>Limited behaviors</td>
<td>Decrease compulsive and defensive behavior</td>
</tr>
<tr>
<td>Lengthening time</td>
<td>Increase employee knowledge and concern about the broad work environment</td>
</tr>
<tr>
<td>perspective</td>
<td></td>
</tr>
<tr>
<td>Toward peer or</td>
<td>Utilize an increasing range of abilities, especially cognitive and interpersonal abilities</td>
</tr>
<tr>
<td>superordinate position</td>
<td></td>
</tr>
<tr>
<td>Growing awareness</td>
<td>Increase the portion of the flow of work and hence the time frame in which the employees attempt to be more efficient and effective</td>
</tr>
<tr>
<td>and control</td>
<td></td>
</tr>
</tbody>
</table>

a Based on Argyris, 1957.

Below, the reader can track the specifics that permit individuals to do in their organizational roles what they need to do as humans-at-their-best. And the effects below also suggest the profile of attractive consequences to which this good fit of person to organization can lead.

A NEW MODEL OF SERVICE DELIVERY

The concept emerged unevenly over time, but SQAT—the Special Quality Action Team—all along worked toward an alternative mode of delivering nursing service. It became known as Shared Care. It contrasted with the Total Care concept prevailing throughout the corporate system, as well as in much nursing everywhere. Exhibit 14.8 provides some details about Shared Care: it seeks to reinforce collaborative relationships between all care-providers for a specific patient population, building especially on the observation of a low Sense of Community in both C and E conditions.

Broadly, Shared Care may be described as nonbureaucratic, and it shares much conceptual territory with the flow-of-work model (Golembiewski, 1995). Some readers may see in Shared Care what they identify as “team nursing,” and that is correct as far as it goes. However, note that “Shared Care” as intended here should be viewed in the basic context of the various attitudes/skills about interpersonal confrontation detailed earlier as well as of the Daily Events Schedule in Exhibit 14.9 below, and of other design details elaborated above.
### Exhibit 14.8 Two Models of Delivering Nursing Services

<table>
<thead>
<tr>
<th></th>
<th>Shared Care</th>
<th>Total Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should be patient-centered</td>
<td>Needs of individual care-providers, or of their professions or specializations, often dominate</td>
<td></td>
</tr>
<tr>
<td>Provides for 3 15-minute team meetings per (12-hour) shift: near start-up, at mid-shift to &quot;touch base,&quot; and at shift’s end to debrief</td>
<td>No such provision</td>
<td></td>
</tr>
<tr>
<td>Emphasizes teamwork by care-providers with knowledge/skills required by a cohort of patients</td>
<td>Emphasizes individual knowledge/skills and especially autonomy of RN: &quot;solo nursing&quot;</td>
<td></td>
</tr>
<tr>
<td>Loyalty has multiple overlays—to patient, team, profession or specialization, employing organization</td>
<td>Basic loyalty to profession or specialization</td>
<td></td>
</tr>
<tr>
<td>Local controls relevant to each care-providing situation should dominate</td>
<td>External controls dominate—e.g., historical standards and norms</td>
<td></td>
</tr>
<tr>
<td>Responsibility and motivation are team-based to promote positive care-giving as well as to satisfy emotional needs for belonging and community</td>
<td>Responsibility and motivation are individually based to clarify assignments of error</td>
<td></td>
</tr>
<tr>
<td>Patient-as-actor in own treatment, directly or via surrogates, whenever possible</td>
<td>Patient as acted-upon</td>
<td></td>
</tr>
<tr>
<td>Decision-making encouraged at bedside level</td>
<td>Decision-making often occurs above bedside level</td>
<td></td>
</tr>
<tr>
<td>Team members with various skill levels contribute to positive care-giving, given coordination and reasonable delegation by the RN team leader (TL)</td>
<td>RN is in dominantly operational role</td>
<td></td>
</tr>
<tr>
<td>Requires managerial skills and delegation</td>
<td>Requires effective communication, conflict resolution and sharing of patient information between team members</td>
<td></td>
</tr>
<tr>
<td>Requires effective communication, conflict resolution and sharing of patient information between team members</td>
<td>Emphasizes individual nursing skills and specializations</td>
<td></td>
</tr>
<tr>
<td>Includes a mini-hierarchy among care-givers that allows advancement for an RN Team Leader, who retains contact with specific patients and with direct care-giving</td>
<td>Information about patients often will be unevenly shared among care-providers—due to lack of supportive relationships, tensions between individuals with different skills or professions, real or alleged confidentiality, and so on</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Advancement typically requires movement up the bureaucratic hierarchy into administrative positions, which implies distancing self from patients as well as from direct care-giving</td>
<td></td>
</tr>
</tbody>
</table>
Emerging Role Descriptions

Role clarification also played a major role in SQAT's planning, consistent with the emerging Shared Care model as well as with the survey and anecdotal evidence about role unclarities. Consequently, one major product of the action-planning was a detailed set of role descriptions for Shared Care—16 pages worth, in fact, in substantially final form.

Although we avoid details here, the Shared Care model facilitated already-ongoing upgrading of the LVN role while avoiding role compression in two ways. First, RN knowledge and skills were utilized more extensively by greater emphases on:

- Nursing assessments
- Certain treatments
- Critical interaction with patients and families
- Communication with physicians
- Development of the overall treatment plan and correlating its components
- Monitoring the results of care to patients
Teaching patients and families
Nursing diagnoses

Second, Shared Care involves what amounts to a second career ladder, which opens a new RN role. See also the last item in Exhibit 14.8, RNs in Shared Care, which permits nurses a sense of growth and development while continuing as a direct care-giver. In greater detail, the Team Leader (TL) is accountable for nursing care given by all team members, using a participative style of leadership, and is responsible for Shared Care applied to an assigned group of patients. The TL assesses patient needs and plans nursing care assignments based on those needs and priorities.

The Staff Registered Nurse coordinates care delivery to a specific patient group. Directing and delegating nursing staff activities, the Staff RN acts as a liaison between the patient, family, and physician, while also managing acute care for adult patients with physicians and auxiliary health care teams. The Staff RN focuses on monitoring and stabilizing patients, while coordinating, documenting, and planning diagnostic and recovery regimens with a team.

The Clinical Support Nurse assists the Shared Care team when needed, with each CSN providing the team leader with up to 50 percent of her/his time during an eight-hour weekday shift. The CSN must balance his or her time between E and C conditions, responding quickly to problems or overload in either. The position is fast-paced, involves performing a variety of tasks, and requires above-average clinical skills, the ability to move between worksites, and personality features that fit in comfortably, whenever help is needed.

In addition, a Shared Care nursing team also includes roles for Licensed Vocational Nurses (LVN) and for Certified Nursing Assistants (CNA).

Daily Events Schedule

Some additional details reinforce the senses in which the Experimentals moved toward their intended model of interactions. Paramountly, Exhibit 14.9 reflects how interactions became manifest in ward dynamics. There, considerable detail is summarized about how OD values would come to characterize the go-go of the experimental ward.

By and large, Exhibit 14.9 is self-explanatory. Usefully, the reader may trace how the schedule of events begins and ends with a marshalling of group forces—to begin to provide direction for the day, as well as to help resolve the feelings and reactions that may have been generated during the day’s work just past. The rationale is direct. Such features help mobilize the sense of groups and emotional forces, among other things, to reinforce the collective requirements of Shared Care, to improve communication as well as a meaningful sense of control over the work environment, and to increase social cohesion and combat aloneness and alienation.
These general comments can be placed in historical and substantive contexts. To begin, the find pre-experimental activities involved, a re-educative intervention whose basic features are summarized below.

**A RE-EDUCATIVE INTERVENTION**

As action-planning matured, E participants perceived a need for an enhanced sense of teamwork to implant Shared Care. A 4-hour seminar focused on positive problem-solving, and on building toward the early voicing and resolution of issues and conflicts. One action-planner noted, based on the survey data as well as her own experience, that (Rountree and Weber, 1991b) “Nursing staff everywhere fall into the mode that conflict is bad, and should be avoided at all costs. They need to communicate and delegate effectively, but aren’t taught those skills in schools” (p. 4).

Several conceptual models for dealing with conflictful situations were introduced, including the Pinch Model of John Sherwood and John Glidewell as well as Harvey’s Abilene Paradox. Reviewing their past personal successes with positive problem-solving, as informed by conceptual contexts, was expected to give E participants a shared knowledge-base from which to practice new behaviors, as well as the language to coach and reinforce each other in confronting behaviors (Golembiewski and Rountree, 1998, especially Exhibits 14.3 and 14.4).

**Experimental Structure and Schedule**

The final pre-experiment activities involved SQAT agreement about structure, patient load, and a schedule of daily events. Patient load was prearranged with management, with one main contingency. A Clinical Support Nurse could provide an additional 50 percent of an 8-hour day to the Shared Care unit, as needed. Moreover, if the patient census grew sufficiently and acuity also increased, a staff RN would be added. If the census went up but acuity decreased, a CSN would be added.

In addition to variable nursing inputs, both E and C conditions shared “fixed resources” on each of the two daily shifts. These included a nurse manager (RN), an assistant nurse manager (RN), and a unit secretary.

The schedule of daily events provided for a 12-hour day shift of the Shared Care unit, with schematic details being beyond the present scope. The night shift—during which health-care demands abate, doctors are unlikely to be on the floor, and so on—operated under Total Care on both sides of a ward, which had been split down the middle into E and C conditions. In sum, Shared Care organizes around a “whole,” with qualifications.

Note especially that three team conferences were a vital element in the holistic or integrated model of Shared Care service delivery—before the shift,
at mid-shift, and after work. Particular significance was attributed to the end-of-shift debriefings that concluded each daytime SC shift. The total E nursing staff then asked: (a) How was the quality of teamwork today? (b) Did we get it done? and (c) How do we feel about it?

The “unit model” underlying Shared Care has long been with us (e.g., Worthy, 1959; Golembiewski, 1962b), and today it receives growing attention in both business and government. The bureaucratic model still dominates, however, and it departmentalizes around the “parts” of work—separate functions or processes. Basically, the unit model departmentalizes around “wholes”—each of its basic units assembles all or many of the activities necessary for a complete flow of work. Beginning in the 1920–30s, and accelerating during the 1950s, the divisional structure form of the unit model was widely adopted at top levels of organization (e.g., Chandler, 1967). Later—sometimes as a product of natural evolution (e.g., Trist and Bamforth, 1951), and at other times as the result of conscious planning (e.g., Rice, 1958)—exemplars of the unit model could be seen at worksites. Examples include long-wall coal-mining with its autonomous teams (e.g., Herbt, 1962), as well as sociotechnical systems (e.g., Cummings, 1993). Related extensions have been made into “team nursing” and “team teaching,” but variants with such labels cover a very broad range.

MAJOR CONSEQUENCES OF INTERVENTION ON BEHAVIOR AND ATTITUDES

Shared Care sought to achieve several explicit goals—to build a stronger sense of community within the daytime shift of the Shared Care unit; to enrich satisfaction in the work itself; and to enhance nurse-physician and nurse-patient relationships. Why? For example, under Total Care, communication between care-givers tended to be “as needed” or, worse, “when time permits.” This situation reflected limited opportunities in the design of work for the development of mutual trust and understanding. Consequences included the lack of relationships that are tight-knit and dependable, and a reduced sense of sharing in common accomplishments.

Moving toward such objectives required major developments. For example, the Experimental participants had some early jitters in implementing Shared Care, but these seemed to largely disappear after the first few shifts. An in-house report observes: “The first day, staffers generally were not sure how they felt about [the change] . . . by day two they were feeling better about it . . . and by days three and four members were feeling exceptionally positive about it.”

Overall, nonetheless, the 30-day experimental period resulted in a pattern of shifts in the expected directions, with most changes being substantial on both self-reports and objective indicators. Note also that there is no simple way of describing the various subpopulations referred to below. This is due to personnel
shifts, the hospital’s policy of periodically “rolling-up” nurses from one 12-hour shift to the other, and especially the need to protect proprietary information. All appropriate nurses provided survey data—pre- and post-test, E and C. All E and C patients, and the physicians attending 85 percent of them, also provided data.

Three major emphases provide details about pre- versus post-test changes.

Patient Acuity

Patient assignments were made by an external office, with “last patient, first bed available” constituting the initial cut at assignments. Other factors also could affect final judgments—for example, the patient’s acuity, physician preferences, and the mix of nursing skills available.

Although differences in patient acuities constitute a potential threat to E versus C comparisons, that possibility does not seem strong. Thus, a retrospective review of patient acuity ratings on the Medicus System revealed no significant differences between E and C assignments. Moreover, the average acuity of admissions for the 101 days immediately preceding the experiment basically was the same as the E and C means.

Changes in Worksite Features

Broadly, Tables 14.2, 14.3, and 14.4 reveal that the interventions under the E condition had major and expected effects. The pre- versus post-test comparisons are not reported for those in the Control condition. However, they typically reflect only random differences. Note also that Cronbach $\alpha$ coefficients provide no evidence of scale unreliability.

Three emphases provide necessary detail and nuance to reflect the changes in worksite features. In turn, then, attention is directed at: five focal worksite domains; the domain Policies, Practices, and Procedures; and the two domains labeled Physical Environment and Organizational Environment.

Five Central Worksite Domains

For reasons consistent with the values underlying the interventions, the greatest impacts were expected on five of the eight domains—The Work Itself, Work Group Environment, Employee Satisfaction, Employee Motivation, and Employee Performance. The rationale for these expectations is transparent. Scores on all scales in these five domains are expected to increase, with the sole exceptions of the measures of Autonomy and Workload. That is to say, a successful intervention will result in nursing smarter and better, with no perceived increase in workload. Nurses might report a decrease in Workload or an increase in Autonomy, but “no change” is a positive outcome.

As Table 14.2 shows, these exceptions are met in all cases. Autonomy and Workload do vary randomly, as expected of and even required by noncoercive
TABLE 14.2 Summary, Tests of Effects of Five Central Domains, Pre- vs. Post-test E Condition

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Post-test shows</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Work Itself</strong></td>
<td></td>
</tr>
<tr>
<td>Autonomy</td>
<td>.79</td>
</tr>
<tr>
<td>Job content</td>
<td>.88</td>
</tr>
<tr>
<td>Meaningfulness</td>
<td>.88</td>
</tr>
<tr>
<td>Motivating potential</td>
<td>.82</td>
</tr>
<tr>
<td>Responsibility for outcomes of work</td>
<td>.88</td>
</tr>
<tr>
<td>Role clarity</td>
<td>.88</td>
</tr>
<tr>
<td>Satisfies needs that motivated one to become nurse</td>
<td>NA</td>
</tr>
<tr>
<td>Workload</td>
<td>.88</td>
</tr>
<tr>
<td><strong>Work Group Environment</strong></td>
<td></td>
</tr>
<tr>
<td>Communication within immediate workgroup</td>
<td>.81</td>
</tr>
<tr>
<td>Cooperation</td>
<td>.88</td>
</tr>
<tr>
<td>Openness</td>
<td>.79</td>
</tr>
<tr>
<td>Psychological sense of community</td>
<td>.92</td>
</tr>
<tr>
<td>Trust</td>
<td>.92</td>
</tr>
<tr>
<td><strong>Employee Satisfaction</strong></td>
<td></td>
</tr>
<tr>
<td>General satisfaction</td>
<td>.90</td>
</tr>
<tr>
<td><strong>Employee Motivation</strong></td>
<td></td>
</tr>
<tr>
<td>Attitude about quitting</td>
<td>.92</td>
</tr>
<tr>
<td>Opinion of self</td>
<td>.81</td>
</tr>
<tr>
<td><strong>Employee Performance</strong></td>
<td></td>
</tr>
<tr>
<td>Work sampling: Utilization of professional knowledge/skills</td>
<td>NA</td>
</tr>
<tr>
<td>Emotional exhaustion as burnout surrogate</td>
<td>.92</td>
</tr>
<tr>
<td>Nurse/physician relationships</td>
<td>.88</td>
</tr>
<tr>
<td>Patient satisfaction (QCI Index)</td>
<td>.91</td>
</tr>
<tr>
<td>Physician satisfaction (TMPE)</td>
<td>.93</td>
</tr>
<tr>
<td>Use of call light by patients</td>
<td>NA</td>
</tr>
<tr>
<td>Costs of nursing services</td>
<td>NA</td>
</tr>
</tbody>
</table>

Notes: *P < .05, **P < .01, ***P < .001, NA = Not Applicable.

changes. Statistically significant improvements are observed in all but three of the other measures, and these three improve substantially although they are not amenable to statistical testing.

The magnitude of these changes can be usefully illustrated by three of the scales of special concern in E—Psychological Sense of Community, Communi-
TABLE 14.3 Summary, Tests of Effects for “Policies, Practices, Procedures” Domain, Pre- vs. Post-Test, E Condition

<table>
<thead>
<tr>
<th>Policies, practices and procedures</th>
<th>Alpha</th>
<th>Post-test shows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career advancement opportunity</td>
<td>NA</td>
<td>No change</td>
</tr>
<tr>
<td>Career earnings potential</td>
<td>.88</td>
<td>No change</td>
</tr>
<tr>
<td>Compensation</td>
<td>.88</td>
<td>No change</td>
</tr>
<tr>
<td>Growth opportunities</td>
<td>.81</td>
<td>Significant improvement**</td>
</tr>
<tr>
<td>Pay-for-performance (an on-going compensation program)</td>
<td>.88</td>
<td>No change</td>
</tr>
<tr>
<td>Promotional opportunities</td>
<td>.90</td>
<td>Significant improvement*</td>
</tr>
</tbody>
</table>

Notes: *P < .05, **P < .01.

The effects of the intervention also can be judged by the record with scales that were not expected to change in the experimental period. To explain, the interventions were expected to contribute to meeting a broad range of needs for all nursing

TABLE 14.4 Summary, Tests of Effects, Two Environment Domains Pre- vs. Post-test, E Condition

<table>
<thead>
<tr>
<th>Physical Environment</th>
<th>Alpha</th>
<th>Post-test shows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical working conditions</td>
<td>.85</td>
<td>No change</td>
</tr>
<tr>
<td>Unit layout/design</td>
<td>.76</td>
<td>No change</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organization Environment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of supplies/equipment</td>
<td>.76</td>
<td>No change</td>
</tr>
<tr>
<td>Communication between departments</td>
<td>.92</td>
<td>Significant improvement**</td>
</tr>
<tr>
<td>Communication between shifts</td>
<td>.88</td>
<td>Significant improvement***</td>
</tr>
<tr>
<td>Facility image</td>
<td>.90</td>
<td>Significant improvement*</td>
</tr>
<tr>
<td>Management style</td>
<td>.88</td>
<td>No change</td>
</tr>
<tr>
<td>Nursing leadership (in understanding</td>
<td>.85</td>
<td>No change</td>
</tr>
<tr>
<td>problems/needs of care-giver)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support from other departments</td>
<td>.79</td>
<td>Significant improvement**</td>
</tr>
</tbody>
</table>

Notes: *P < .05, **P < .01, ***P < .001.
staff, and especially for RNs and LVNs, but only in the long run—for example, to improve growth opportunities, or the potential for career advancement, and so on. In the 30-day experimental period, however, no major changes were made in those policies, practices, and procedures referred to in Table 14.3. So “no change” could be expected in this domain.

Table 14.3 reflects just such a pattern, with two exceptions. Apparently, many respondents saw in Shared Care real opportunities for growth and promotion, despite the limited E period. Overall, however, “no change” dominates. This is not only expected, but also reduces the credibility of arguments that pre- versus post-test changes reflect only an indiscriminate “halo” on every scale due to a “Hawthorne effect.”

Two Environment Domains

As for the two environment domains in Table 14.4, “no change” is either expected or desirable during the E period. Since no modifications were made in the Physical Environment or the broader Organizational Environment, maintaining the status quo is expected. However, an awkward change-process might induce negative effects on both domains, and especially on scales in the Organizational Environment.

Table 14.4 implies that no such awkward effects occurred, but also possibly that the interventions propagated what might be called second-order effects in some other subsystems with which E nursing staff interfaced. Specifically, on the Organizational Environment, “no change” characterizes three of the seven scales, while the others significantly improve. These improved scales relate to communication and support involving other shifts and departments. The increase in communication between the two shifts is particularly noteworthy, since the day shift in Shared Care and the night shift retains Total Care.

Changes in Performance

The “Employee Performance” domain was reviewed in Table 14.2 but for several reasons deserves detailed attention. Its variables relate to important outcomes; they include several objective measures; and several of the variables are not commonly reported in studies of organization change.

Changes in Role Performance

Table 14.5 seems convincing that the intervention successfully shifted role definitions and performance. For RNs under Shared Care, that shift approximates a 50 percent increase in doing activities for which they were specifically educated and trained. The corresponding increase for those in the LVN role is less dramatic, but still surpasses a 15 percent shift in the intended direction.

These changes in Utilization of Knowledge/Skills are derived from a sub-
TABLE 14.5 Activities Performed for Which Knowledge/Skills of Role Are Required, in %

<table>
<thead>
<tr>
<th>Roles</th>
<th>Pretest (all nursing staff in ward)</th>
<th>Under Shared Care (E condition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>53.0%</td>
<td>78.5%</td>
</tr>
<tr>
<td>LVN</td>
<td>60.0%</td>
<td>70.0%</td>
</tr>
</tbody>
</table>

stantial system of categorization, described in the companion piece as Work Sampling. Overall, the activities in question relate to:

Direct Care—to the patient (29 activities)
Indirect Care—for the patient (18 activities)
General—for the unit/staff (20 activities)

Notably, also, this increased Utilization of Knowledge/Skills under Shared Care seems associated with satisfaction with work. For example, under Shared Care, General Satisfaction has a high correlation with Utilization ($r = .71, P = .0001$).

Reduced Emotional Exhaustion

A major concern throughout was that the pace and extent of E changes might heighten burnout. Stressors are always numerous in medical-surgical nursing, and all changes imply additional possible sources of strain (Warrick, 1981). The key question thus becomes: Are the E features attractive enough to overbalance this potential for new strain?

Three kinds of evidence suggest that the interventions avoided the dangers. First, the direct test of Emotional Exhaustion supports the view that the interventions succeeded not only in overcoming start-up strain but also in reducing the pretest level.

Second, many of the scales summarized in Table 14.2 also support the interpretation that nurses’ degree of emotional exhaustion was kept in check. For example, General Satisfaction improves significantly between the pre- and post-test; Workload does not change significantly; and Attitude about Quitting improves. The reader also can extend this reinforcing analysis, using a number of the other results summarized in Table 14.2.

Third, reasonably, the potential for strain in E probably was reduced in several reinforcing ways. Thus, relationships and jobs were restructured to be more need-satisfying to employees, consistent with Argyrian (1957) and Herzbergian (1959) notions. Moreover, the explicit emphasis throughout the action planning on regenerative interaction probably was not only more need-satisfying, but
also helped in both the planning as well as the experimental period to avoid the awkward consequences of degenerative interaction—low openness, low owning, and low trust as well as high risk. Degenerative interaction dissipates energies in seeking answers to the wrong issues, and in many other ways increases and exacerbates a broad range of stressors that contribute to advancing burnout (Golembiewski, Boudreau, Munzenrider, and Luo, 1996).

Improved Patient Satisfaction

On the six-item Quality of Caring Index, or QCI (Rountree and Deckard, 1984), Shared Care proved superior. For example, during the experimental period, E significantly had a higher QCI mean than C, with the same number of patients in each condition. Broadly, QCI in the Experimental condition was not only higher but also less variable than QCI in the Control condition. Moreover, QCI in C did not differ significantly from the pre-experiment experience in the ward.

These self-report data are confirmed overwhelmingly by patient responses to open-ended items. For example, one patient reported, “I almost never had to ring the buzzer [call light] because someone was always there.”

Improved Physician Satisfaction and Nursing Relationships

The pretest versus post-test comparisons were substantially favorable, judging from responses to The Medical Practice Experience (Rountree and Davis, 1990). This significant shift, reported by Table 14.2, occurred even though pre-test opinions collected one week prior to the 30-day experimental period already compared favorably to other nursing units. After the experiment ended, physicians reported that:

- RNs were more available in Shared Care.
- After initial reservations, they warmed significantly to both the making of shared rounds with nurses as well as to face-to-face meetings for making decisions about patients, both of which are relied upon heavily by Shared Care.
- The quality of information under Shared Care rose above the already-high pre-test levels.
- The quality of care in Shared Care was higher than in Total Care.
- The quality of charting and its completeness by Shared Care nurses improved.
- The probability increased that physicians would meet at the bedside the nurse(s) knowing the most about a patient.

Reduced Call Light Usage

Equipment also generated a record about the use of call lights in both E and C. Although technical difficulties with the recording system wiped out data during
about 23 percent of the experimental period, the surviving data showed that
Shared Care patients used the call light a full one-third fewer times per 12-hour
daytime shift than Total Care patients. The interpretation seems obvious.

Reduced Nursing Costs

Although cost data are proprietary, the overall situation seems clear enough in
E versus C comparisons. Specifically, Shared Care generated the results detailed
above, along with substantial reductions in the costs of care-giving by nurses.

Specifically, corporate conventions generate two sets of numbers relevant
for present purposes, which are reviewed here in relative terms. First, define as
100 the annualized nursing costs of Total Care in the ward for the year prior
to the experiment. Second, the 30-day experiment with Shared Care also was
annualized, to estimate what nursing costs would be if converted permanently to
the Shared Care mode.

The Shared Care annualized cost is .802. This is substantially lower than
the Total Care estimate for the year preceding the experiment—that is, 1.00.

DISCUSSION

All in all, then, multiple kinds of evidence—mostly self-reports, but including
several objective measures—indicate that a redesign of a worksite better met the
needs of all major stakeholders in a medical-surgical unit while also reducing
costs. Moreover, the factors confounding and distorting this interpretation are not
major ones, on the whole. For example, Shared Care covered only the 12-hour
daytime shift in the Experimental half of the ward. Most of the care-giving would
occur during that time so, presumably, a full-day extension of the Shared Care
mode would not have diminished the observed effects. Indeed, the shift differ-
ences in mode of service delivery perhaps could have been expected to exacerbate
normal tensions between shifts. The data indicate that did not occur.

No broad claims for generalizability will be made. Nonetheless, the E con-
dition represents many similar-type situations—those in which changes in tra-
ditional routines and structures are made necessary by competitive pressures,
and those in which individual preferences and needs can be better met by post-
bureaucratic changes.

The evidence suggests that a new and more satisfying balance was achieved
in E. On the one hand, stressors no doubt were induced by the major changes
in the Experimental condition’s roles, structures, and activities. On the other
hand, these new stressors seem to have been more than counterbalanced by a
range of improvements in individual and group relationships as well as in roles.
Thus, the increased Sense of Community credibly would have increased social as
well as instrumental support and, in turn, these would have helped keep emotional
exhaustion from escalating. Relatedly, the apparent improvements in communica-
tion and collaboration probably would have helped heighten task performance, at the same time as they heightened social cohesion and helped contain emotional exhaustion and burnout (e.g., Golembiewski and Munzenrider, 1988, pp. 203–208).

This chapter attempts to heighten the possibilities of general application, if not necessarily of generalization. Consider Exhibit 14.8, presented above, which conceptually identifies two contrary models for medical service delivery via nursing. This application also provided a mass of operating policies and procedures for activating the behaviors and attitude appropriate for “Shared Care.” Exhibit 14.9 presents one set of operating details appropriate for bringing “Shared Care” to life, and some discussion several pages earlier provides context for that exhibit.

This discussion will not dwell on the multiple ways that “Shared Care” differs from “Total Care” models. But many points of contrast are apparent on the face of it, as in building group forces that will reinforce a “shared care” effort and heighten participants’ sense of psychological community as well as decrease the emotional exhaustion of nurses.

One caveat concludes this review dealing with change in a hospital setting. While the present design and results imply hope that a ubiquitous-type situation can be constructively dealt with, we are well-advised to respect profound ethical and value challenges, if not dilemmas. What distinguishes an intervention that responds to legitimate needs from some unilateral manipulation to squeeze more work out of people? No general prescriptions apply, but this case suggests some major components of the required respect for ethical and value concerns. Thus, the focus here was determinedly from the bottom-up. Moreover, it promised no quick fix. Some four months elapsed while the experimental participants wended their way from diagnostic data, to plan, then to implementation, and to evaluating results. Finally, and perhaps most basically, participants variously escalated their confidence that they could trust one another, and that management had no hidden agendas beyond the conviction that real change was necessary, and that it would come most effectively and efficiently as a result of the local initiatives of relevant stakeholders. The design elements contributing to such effects are detailed in a companion report (Golembiewski and Rountree, 1998).

NOTES

1. The categorization was developed by Ann Van Slyck and Associates, of Phoenix, Arizona. The activities constitute an element in a broader package—data collection procedures, documentation forms, and data-coding methodology.

2. All calculations employ corporate conventions that take into account skill/professional mix; average hourly rates; patient census; and number of hours for each full-time equivalent.
REFERENCES


Rountree, B. H., and Weber, A. (1991b). Redeployment of nurse skill levels is focus of...


Illustrating Large-System Change at the Interface*
Testing Some Features of the Common Wisdom

Since the earliest appearances of Organization Development (or OD), and especially in the public sector, commentators have emphasized the relevance of the interface between what is classically labeled Politics versus Administration—roughly, the policy-determining versus technical or implementing arenas. Thus, Golembiewski (1969) warns that what with so many things subject to “going political,” at so many points in time, in response to such a broad range of stimuli, OD faces special challenges at the interface. Other observers have been absolutely dour concerning planned change anywhere in the public sector, and nowhere less optimistically so than at the interface (e.g., Burke, 1980).

The rationales build-down from the constitutional separation of powers as well as checks and balances. These institutional features are seen as distributive, if not fragmenting, whereas OD values the broadly integrative. This conceptual gap implies low success rates. With the same consequence, many observers claim few OD applications have been made in public sector. This lack of positive practice effects also implies low success rates. We shall see.

Such warnings seem alarmist, however, when the literature on OD applications is reviewed in detail. That is, public sector applications seem about as frequent as the proportion of public employment justifies; and success rates of OD applications in the public sector are about as high as in business locations, where they are formidable (e.g., Golembiewski, Proehl, and Sink, 1981; Nicholas, 1982). Indeed, in some panels of applications, public sector applications have higher success rates (e.g., Golembiewski and Sun, 1990; Golembiewski, 1998). See also Chapters 1 and 2, especially.

So what is going on? Several major ways of reconciling the several points above come to mind, of which we consider only two. First, nobody has looked very intensively at success rates of OD at the interface, and it may be possible that observers attribute low public success rates on the basis of an inept assumption about the interface. Consequently, the expectations reflected above may rest in a faulty disjoint-by-definition. Derivatively, something like the P/A interface also exists in business organizations. This chapter is not alone in attributing a public interest to all organizations of substantial scope—whether business, public, or voluntary (e.g., Dahl, 1975)—and the rationale for acknowledging various degrees of “publicness” in all organizations is growing (e.g., Bozeman, 1987; Golembiewski, 1995c). Relatedly, today’s organization forms may be too complex for such a simplicism, even if other days and minds permitted in/out classification.

Second, it might be argued that ODers have developed a substantial catalog of theoretical as well as practical ways-and-means of dealing with change, whatever the locus or level. For example, substantial if incomplete details have been published about the features of the interface, as well as about related implications for useful intervening (e.g., Golembiewski, 1985, pp. 233–365).

Both selected points receive some attention below, with emphasis on the second point above. Providing the appropriate perspective will involve three basic emphases derived from OD applications at the P/A interface, both in business and government. In introductory summary:

Several classes of designs for planned change in large systems will be described.
A panel of planned change applications impacting the interface will be described.
Several general guidelines will be assessed to determine whether or not they are supported by the panel of applications.

In sum, these ambitions have strong motivations. Thus, no such panel focusing on the interface currently exists, and such as assemblage alone will permit adding to existing knowledge. Realistically, the panel will provide only a place to start, but even that might help motivate the greater attention by others to interface effects on which more definitive work can be based.
SOME DESIGNS FOR PLANNED CHANGE AT THE INTERFACE

Basically, four classes of designs associated with Organization Development, or OD, are distinguished here. These classes of interventions, among others, have been applied at the interface:

Traditional OD designs, which often build initially on the development of values/attitudes/skills generating regenerative interaction, which then can provide the foundation for appropriate policies, procedures, and structures (e.g., Golembiewski, 1979a). Substantial evaluative research exists, and success rates are formidable (e.g., Golembiewski, 1997b, 1998).

Future search conferences, a family of related designs that bring together broad ranges of participants in joint planning exercises that are advertised as a “new paradigm at work” that emphasizes self-management by finding “common ground.” Such designs can generate enthusiasm, lead to the development of now → future pathways, and inspire committed action (e.g., Emery and Purser, 1996; Weisbord, 1992).

Large system interventions patterned on “work-out,” as pioneered at General Electric, have some features of future search conferences. Workouts seek to extend the power of interpersonal and small group dynamics into macro-system dialog. At off-site experiences, such designs “bring all stakeholders into the same room” so as to cut across functional barriers and many hierarchical levels, thus bringing all necessary resources to bear on an issue, reducing processing time, and moving toward the goal of basic cultural change that will make work-out the normal way of doing business at GE (e.g., Ashkenas and Jick, 1992; Golembiewski, 1995c, pp. 188–189).

Evaluative research is rare about “work-out” (Ashkenas and Jick, 1992), in part by design (Ashkenas, Ulrich, Jick, and Kerr, 1995, p. 329), but strong anecdotal support has led to prominent extensions like the “boundaryless organization” (e.g., Ashkenas, et al., 1995).

Appreciative Inquiry, which builds on the basic notion of the power of positive approaches and stimuli, in contrast to the problem-solving or (allegedly) deficiency orientation of the three other classes of interventions described above (e.g., Cooperrider and Srivastva, 1987).

Both evaluative research and commentary on Appreciative Inquiry (AI) are rare (e.g., Golembiewski, 1997a), but applications abound.

SOME APPLICATIONS AT THE INTERFACE

Exhibit 15.1 illustrates a rare, and probably unique, resource in providing an illustrative cohort of applications of planned change impacting the interface. Be-
### Exhibit 15.1 Selected OD Programs Involving Politics/Administration Interface, Anchoring a List of 41

<table>
<thead>
<tr>
<th>Case</th>
<th>Date</th>
<th>Locus/source</th>
<th>Character of design</th>
<th>Capsules of consequences</th>
<th>Success category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1966–71</td>
<td>U.S. Department of State, agencywide (Hamon, 1975; Golembiewski and Kiepper, 1988)</td>
<td>Basic reliance on sensitivity training for senior and middle management, as pre-work for local team-building or conflict resolution as carriers of normatively consistent OD initiatives. N is not known, but is substantial</td>
<td>Early training was often personally powerful, but many of its organization extensions were blunted—in large part by political attacks on State's major champion. Evidence suggests major &quot;sleeper effects&quot; some years later, when both the early champion and many of his opponents were gone from the agency as well as from the association of foreign service officers.</td>
<td>II</td>
</tr>
<tr>
<td>2</td>
<td>1966</td>
<td>U.S. Department of State, Administration Area (Crockett, 1970)</td>
<td>Basic interaction-centered team-building, as follow-on to sensitivity training in &quot;cousin groups,&quot; N = 10, who were forming an Executive Group.</td>
<td>Intended team-building effects are reported in anecdotal form, with the resulting empowerment leading to a basic structural change that eliminated multiple levels of review even when that change-objective seems a &quot;sacred cow.&quot;</td>
<td>I</td>
</tr>
<tr>
<td>Year</td>
<td>Event</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968-70</td>
<td>Smith, Kline &amp; French (Golembiewski and Carrigan, 1970a, 1970b; Golembiewski, 1972, pp. 408-432)</td>
<td>Basic reliance on sensitivity training for regional and divisional sales management, followed by team-building down through operating levels, to build regenerative interaction climates. Linkages with the executive level are involved. The basic OD goal is to improve feedback linkages, up as well as down the corporate chain. N approximates 75.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>August 1969</td>
<td>Ferneda Workshop on Border Dispute, Horn of Africa (Walton, 1970)</td>
<td>Design had two basic parts: a workshop experience to develop values/skills associated with regenerative interaction and country groups later seeking to agree on a resolution of the border dispute. Broadly, the workshop experience had the expected effects (e.g., on trust-building). But the country groups did not settle the dispute, in part because the later Workshop did not generate a consensus resolution.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>$1.1-billion commercial banking institution whose <em>nom de plume is First Peoples Bank, or FPB</em> (Williams, 1996)</td>
<td>Combines Appreciative Inquiry (AI) principles (Coomer, 1996) in a survey/feedback intervention in an organization experiencing financial difficulty. N = 8,000, with the prime champion being a Vice President who headed a cost-cutting campaign. A vision and a statement of values were considered crucial in guiding the FPB of the future, building on positive aspects of the initial FPB. A broad evaluation concludes that AI “can breathe life, strength and a proactive response into a crisis by emphasizing positive aspects.” A year after the survey, FPB was in part merged with another organization, and in part sold off to another firm. AI is reported to have facilitated identifying “what should be preserved” as well as in facilitating transitions (Williams, 1996, p. 50).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Exhibit 15.1 Continued

<table>
<thead>
<tr>
<th>Case</th>
<th>Date</th>
<th>Locus/source</th>
<th>Character of design</th>
<th>Capsules of consequences</th>
<th>Success category</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>1996–97</td>
<td>Comprehensive State University internal memo (Golembiewski et al., 1997a).</td>
<td>Survey/feedback design involving dean of a major college, the Senate charged with 5-year evaluation of the dean, and a Senate Committee. Action research is the guiding metaphor, but the basic short-run Senate goal was to develop an acceptable model for assessment, given major past turbulence in deans' reviews.</td>
<td>For the first time locally, a comprehensive survey was conducted, and the results broadly distributed, with the goal of supporting a program of action research. The survey report was accepted by the Senate and dean, and initial steps to act on the results were made. No strong management champion existed.</td>
<td>II</td>
</tr>
<tr>
<td>40</td>
<td>1995–97</td>
<td>U.S. Department of Labor, Employment and Training Administration (internal memo, 1997; Golembiewski and Miller, 1997; Golembiewski and Miller, 2001)</td>
<td>Basic labor/management cooperative form using participative committee and team sessions, assisted by an internal Reinvention Office, generally in an OD mode. Planning committees approximate 50, with various reviews by all Labor executives and all ETA personnel.</td>
<td>Intensive work occurred over a year-plus, and 12 restructuring options were generated shortly before the two major political champions left Labor after President Clinton's 1996 reelection. Progress was further limited by congressional contention over the replacement for Secretary Reich.</td>
<td>II</td>
</tr>
</tbody>
</table>
Unspecified, at the almost-unanimous requests of participants

Several collections of relatively permanent and regime-vetted executives in totalitarian or autocratic political systems, N = 6 to 12, typically curious about the 'Western OD' they read about and attracted by offers of 'demonstrations' by an ODer

Based on "micro-lab" experiences in day-long simulated organizational settings that permitted testing the applicability of regenerative interaction systems to internal operations of clusters of work-related executives (Golembiewski, 1979, vol. 1), all experiencing the multiple and serious costs of the closed interaction encouraged by their political systems.

All but uniformly, several iterations of a similar design led to similar reactions: personal attraction to the values underlying regenerative interaction, but general agreement that the anticipated benefits did not justify the real or fantasied costs in their political systems; and the nearly absolute conviction that accommodating changes in their political systems were unlikely, although they later occurred. Substantial interest was expressed in applying regenerative interaction to privatized contexts, often the family. Some anecdotal evidence implies success and satisfaction in such efforts.

III in executive clusters; II in other areas, usually non-work
cause the full panel of 41 applications has been published elsewhere, for convenience we take the first and last four cases from that list. The reader will get the sense of it, and trees will be saved.

Earlier attempts to both describe and guide intervening at the interface exist (e.g., Golembiewski, 1985, pp. 233–365), but practical and theoretical developments have not burgeoned. (e.g., Halley, 1994). Exhibit 15.1 illustrates applications covering all four classes of planned change described above—by date, locus, design, and a sketch of the main outcomes.

How can this author be substantially certain of a claim about the useful status of Exhibit 15.1? Four contributors to certainty will suffice here. First, literature on applications at the interface is unusual, if nothing else; and most experiences with the four approaches focal here comes from mid and low levels of organizations. So, isolating a cohort of OD interventions at the interface is helpful, even if limited.

Second, interventions at the interface often go unreported, or are reported only in sketchy form—for example, in the “Camp David accords” of President Carter. Exceptions do exist in both public and business sectors (e.g., Lippitt, Langseth, and Mossop, 1985; Moch and Bartunek, 1990), but the generalization applies broadly, for reasons that seem clear enough, in the general case. Thus, the people with appropriate technical skills and experiences will be constrained by the fact that principals have very big fish to fry, perhaps especially their own places in history. This adds massively to the normal constraints inherent in confidentiality or the exceptional demands of national security.

Third, the cases included in Exhibit 15.1 have a self-referential character in one of two basic senses. Thus, this author had a direct involvement in nearly half of the cases, as the citations suggest. Or this author had some special interest or knowledge about other cases that, while falling short of direct involvement, provide a privileged perspective. Because of these two factors, if for no other reason, Exhibit 15.1 has a limited character.

Fourth, Exhibit 15.1 reflects this author’s view that something like the Politics/Administration distinction exists in all organizations (see case 3). These interface features may appear in different guises and at different—even mercurially shifting—levels. Although this view is becoming more common, it still has not achieved conventional status.

**TESTING SOME GENERALIZATIONS ABOUT INTERVENTIONS AT THE INTERFACE**

Exhibit 15.1 will be put to direct use in helping to test several generalizations about planned change. Tentativeness is appropriate but, nonetheless, eight generalizations will be viewed in light of the applications in Exhibit 15.1, and a ninth generalization relies on a broader context.
Generalization 1

From the very outset, OD has been emphatic about the usefulness of getting support “from the top.” More recently, the prescription has called for “champions,” typically referring to well-placed executives, and especially those in “line” roles.

In both cases, the basic rationale suggests a kind of organizational street smarts useful for dealing with interface phenomena, if substantial success is a goal. Support from the top, or from the champion, has both offensive as well as defensive uses, as it were. Although OD seeks to build deliberately away from a reliance on coercive power, the next best thing to having the support of elites for OD values is to have those elites in a nonopposing posture. Looked at in another way, all change relates to issues of power or influence, and especially so at the interface during periods of implementation. That is to say, all systemic change tests the interface—it challenges and may supersede whatever contemporary conventions distinguish Administration from Politics, even if temporarily. Having elite support for this fateful relocation helps safely manage potential cross-currents, and implies the safety of participation by all organization members.

This much having been written, generalization 1 finesses serious ambiguities and limitations, to which Exhibit 15.1 helps direct useful attention. Six points are usefully made, although they fall short of comprehensiveness. Recall that, conveniently, cases 5 through 37 are omitted from the exhibit. The case numbers are reported here to facilitate cross-checking with the original source.

First, several cases in the full version of Exhibit 15.1 imply the central role of the “champion,” both by presence as well as by absence. The former effect is highlighted most dramatically in case 31, where one CEO was very impactful on the way a huge workforce now envisions its work, and perhaps especially in how they will come to do so. Case 40, concerning the Labor Department, illustrates the contrary situation—where the president, department head Reich, and a political appointee who was the latter’s direct-report initially supported an ambitious effort at change, then, suddenly, all were gone. The effects are still being felt, but the proposal was enacted into law.

Second, the common prescription was never very precise. For example, “the top” should not be taken literally, because—as in cases 3, 10, 29, and 40 in Exhibit 15.1—sufficient wriggle room for real change can exist even for smallish units of organization at lower levels. In effect, decentralized organizations both relocate and redefine “politics”—for example, responsibility for discernible chunks of work is relocated to lower levels, while the rules of the game permit all or many units “to win” in differentiated ways of increasing lower-level autonomy and discretion. In sharp contrast, traditional structures encourage zero-sum competition at supervisory levels of higher (e.g., Golembiewski, 1995c, esp. pp. 170–199).
Third, the ETA case in Labor deserves a bit more attention because it directs attention to a useful qualifier of generalization 1. As a consultant, your author usually finds it useful to encourage champions within Administration as well as Politics spheres. So, when I became a consultant in the ETA case, I did my usual thing during introductory meetings with career officials. “Who would be up-front,” the consultant asked, “if your champions were to go down?” Generally, the consultant was encouraged by careerists to think more positive thoughts, but some got the message and surfaced their support in ways that had impact with Congress even after the three levels of political support had evaporated.

Case 40 has been resolved favorably but, nonetheless, your author/consultant now more than ever believes he raised what is an apt question, and perhaps nowhere more so than at the Politics/Administration interface. The cross-pressures there are so great that, despite contrary intentions and expectations (e.g., Reich, 1997, esp. pp. 333–336), political champions can easily disappear within the time frame of a typical change of program—hence, the centrality of early attention to possible Administration champions, and not only because they will be the prime implementors of any changes.

Fourth, and a very useful qualifier of the usual OD prescription favoring “support from the top,” tactical as well as strategic considerations encourage a niche multi-approach to champions. Consider the experience in the Baltimore office of the Social Security Administration (Rainey and Rainey, 1986a, 1986b; Golembiewski, 1995b, esp. pp. 242–246). There, the champion was a career civil servant—well respected, but near retirement and, consequently, without any apparent ax to grind in supporting the successful implementation of a flow-of-work, or unit structure, for organizing work. This alternative provided sharp contrast with the long-prevailing bureaucratic approach to organizing work (e.g., Golembiewski, 1993, 1995b), and was known by political actors who were less well situated to sponsor change than the careerist.

Fifth, “hierarchy” is of diminished or nuanced relevance in many cases of applications in large systems, and perhaps nowhere more so than when “networks” play dominant roles, as in cases 12, 13, 14, 32, 36, and 37. Here, concern with “the top” takes on a plural character. The role of hierarchy also will be variously limited and conditioned by applications of work-out, future search conference, and Appreciative Inquiry applications. The focus on “the hierarchy” also will be affected, obviously, when labor/management committees play major monitoring roles in programs of change (e.g., case 27).

Sixth, to close this short catalog, many of the cases in Exhibit 15.1 suggest the value of a kind of organizational double-play: internal as well as external champions, with the latter adding useful complexities to the sense of “the hierarchy” seems as internal to an organization (Golembiewski and Kiepper, 1983a, 1983b, 1988). The latter can be technically skilled consultants; they can be disseminators-of-ideas become gurus; or, on occasion, one person may serve in both
roles. Their centrality is illustrated by several of the cases in Exhibit 15.1—cases 23, 32, and 33. Such externals can be powerfully rein-forced by internal change-agents, who provide crucial specialization in the history and culture of specific targets of change.

Generalization 2

OD approaches to planned change were, ab initio, nested in values/skills/attitudes, as well as variously built-into training and experience in small groups, and these became in effect the front-load of programs of change (e.g., French and Bell, 1993; Golembiewski, 1993). See cases 1, 3, 6, 11, 13, 17, and 28 for examples. This front-load typically was interaction-centered, and it could be approached via T-grouping in intact teams or team-building (e.g., Dyer, 1987), among other possibilities, such as the sharing of Three-Dimensional Images.

For good or ill, Exhibit 15.1 also reflects a recent trend at the interface that stands in direct opposition to this pronounced historic bias. In sum, this new trend is reflected in the following:

Large system interventions such as future search conferences or work-outs—see cases 23, 24, 30, and 31, for example—which typically provide no explicit prework in developing appropriate values, attitudes, or skills.

Groups may be utilized, as in the brainstorming employed in work-outs (e.g., case 31), but these collections are only convenient contexts, or even “nominal groups.” Such usages contrast sharply with the concern in T-groups or team-building for developing in each small collection of learners quite specific values-in-action that provide support and increase trust, as well as reinforce consensus. This is convenient, or perhaps necessary, because the learning context also is intended to become the agent of change.

This developing trend may be viewed as “getting to it, without preliminaries,” and two major points provide perspective on what may be a new trend, which in general relates to Appreciative Inquiry, future search conferences, and work-out. See Exhibit 15.1, cases 23, 24, 30, 31, 32, and 38. First, as the growth curve of its adoptions powerfully suggests, the new bias may have much to recommend it. Among other features, several of the cases in the full panel underlying Exhibit 15.1—e.g., cases 23, 24, 30, 31, 32, and 38—imply these attractions, among others:

Their underlying philosophy places great faith in the “normal person” as a generator of changes, which minimizes the need for pre-intervention training which, in turn, can reduce associated resistances to change as well as costs.
They sharply reduce reliance on “the experts,” who can be practically and philosophically troubling, as well as expensive, and especially when they are involved in diagnosis and prescription.

They reduce threats to individual participants, as compared with designs that call for greater personal openness and confrontation.

They may result in savings in time and dollars, especially in organizations that have invested in infrastructure that variously provides socialization and guidance (e.g., O’Toole, and Bennis, 1992; Golembiewski, 1995c, esp. pp. 336–348): for example, by culture statements; management institutes and other vehicles for inducing specific values and behaviors; the institutionalization of a “way of life” in various documents, including veritable organizational constitutions.

They can help avoid “paralysis by analysis” where the pre-work involves T-groups or other attractive loci for micro-analysis that may divert energies from macro-analysis of task and performance.

Second, however, these attractions have real costs. For example, OD intervenors will not only be sources of experience and theory; they also may provide protection for participants, and especially as a counterbalance to organizational authorities with questionable motives—whether they be merely overenthusiastic, uninformed, or determinedly devious and coercive. Or interventions like future search conferences or work-outs may be valued because they constitute temporary and issue-limited “parallel structures” that imply fewer challenges to traditional managerial prerogatives and power, even as they also provide useful tools to be called on by management as it wishes.

In sum, these interventions may be useful for limited conceptual and planning purposes, but they may also be less appropriate for long-run implementation, and they also have a potential for “cooling-out” uses by management intended only to “buy time” or misdirect attention.

Today’s research remains too sparse to conclude confidently on this point, but one can easily go broke betting on serendipity. Hence, the relevance of a delicate balance. It remains unclear how successful applications in large systems without pre-work can be in the long run. Nonetheless, observers must honor noble intentions—that is, the announced GE intent is that work-out replications, over a 10-year period or so, will generate a quick-reacting and radically new culture at work (Ashkenas and Jick, 1992). But such honoring must be cautious because lack of clarity exists about two crucial points: what succeeding stages may be required in getting to cultural change and what specific conditions or dynamics will facilitate those stages.

**Generalization 3**

The development of models of the change process had high practical and theoretic urgency, in early planned change, but a shift may be occurring. If so, it remains
unclear whether this reflects a well-placed confidence in more facile approaches
to change or whether the new trend is a self-defeating one.

These early models have protean uses: they not only reflected OD’s increasing
reach-and-grasp, but also described what will occur, help explain means →
ends linkages, and motivate people to move toward ideals from existing condi-
tions. In the OD literature, for example, the relevance and character of models
for individual and organizational movement always had high priority (e.g., Lip-
pitt, 1975), especially to help transfer off-site learning into systems (e.g., Gole-
biewski, 1979a, 1979b).

Exhibit 15.1 reflects a dramatic shift away from this historic OD anchoring
in the three more recent classes of interventions—future search conferences,
work-out, or Appreciative Inquiry. Anecdotal reports provide strong support for
each of these three classes of interventions, without pre-work, but little rigorous
research deals with either their developmental dynamics or consequences. This
important conclusion is supported by extended analyses (e.g., Golembiewski,
1997a). For present purposes, readers can compare the “Capsules of Conse-
quences” in Exhibit 15.1—for OD interventions such as cases 3, 6, 22, and 28,
as contrasted with exemplars of the three other classes of interventions, such as
cases 31, 32, and 39. The latter are anecdotal and impressionistic, even vaguely
so. In contrast, OD approaches often specify and test in detail their expected
consequences. In sum, one cannot now assemble for search conferences, work-
out or Affirmative Inquiry the kind of evaluation for OD and QWL reflected in
the opening chapters above. Let’s make the strongest possible case against finess-
ing pre-work, so important is the issue. Consider here only the concept of “em-
powerment,” which has a broad currency nowadays among virtually the full spec-
trum of managerial researchers and practitioners. At the level of espousal, this
encompassed spectrum includes OD, future search conferences, Appreciative In-
quiry, work-out, as well as the “Learning Organization” metaphor to be consid-
ered in the last section of this chapter.

In action, however, the four latter classes of interventions devote attention
to empowerment that is mixed, and even indirect, at best. To provide some sup-
port for this summary view, consider these four components of empowerment
(Thomas and Velthouse, 1990):

Potency, or the belief of an individual or a group that they can be effective;
Meaningfulness, a sense that the arena for empowerment is relevant to the
core beliefs, attitudes, and values of participants;
Enhanced choice or autonomy in acting in relevant arenas; and
Impact on relevant system(s) for efficiently and effectively expressing em-
powerments as well as for implementary necessary choice.

Such components also have been elaborated into broad theoretical formulations
(e.g., Kirkman and Rosen, 1997).
A few examples illustrate the typical shortfalls in action concerning “empowerment” when the focus is on the “new” classes of interventions listed on the previous page. For example, potency clearly would be reinforced by upfront training in appropriate values, attitudes, and behavior. As the full panel underlying Exhibit 15.1 implies, however, the several focal classes of interventions seldom, if ever, emphasize such pre-work. In cases, this is a direct derivative from ideology (e.g., Cooperrider, 1996); others see a convenience in the finesse of diagnosis by future conferences (e.g., Weisbord, 1997, p. 66); and the lack of public research about GE’s work-out is often explained in proprietary terms.

**Generalization 4**

Related to the previous generalization—indeed, underlying it—the historical bias in planned change puts a premium on rigorous research that seeks ways to maximize a specific set of values. But fashions in that ideal may be changing.

The historic bias can be documented easily and economically. Witnessing only the dominance in OD of “action research,” whose products are perhaps more revealingly labeled “goal-based, empirical theories” (Golembiewski, 1996). See Exhibit 15.1, in which most of the OD cases—e.g., 3, 6, 17, 26, and 28—clearly encompass empirical research in the column “Character of Design.” Comprehensive evaluations of this OD commitment to research are conveniently available (e.g., Golembiewski, 1990, 1995), and also introduce this volume.

For good or ill, and for different reasons, the three other classes of designs have a sharply moderated enthusiasm for research. For example, Appreciative Inquiry (AI) is tied firmly to social constructionist philosophy (Cooperrider, 1996), typically elaborated as independent of the discipline of classic empirical research. Why is obvious. Basically, perhaps AI postulates that the only reality is “in here,” in the person who “enacts” what exists, granted that some movement beyond this ideological position seems to be occurring (e.g., Gergen and Thatchenkery, 1996, esp. pp. 360–362). In classic empirical research, in contrast, significant realities also exist “out there”—in empirical regularities, increasingly comprehensive theoretical networks, and so on. Here “reality” surely exists as perceived, but also as what may be called interobserver reliability. Other reasons contribute to the dearth of empirical research in AI but, bottom-line, that characterization is granted by both proponents and critics (e.g., Golembiewski, 1997a). For details, see Chapter 25.

The same empirical shortfall characterizes work-out exemplars and (if less so) future search conferences. In the former case, the emphasis is deterministically on applications, with proprietary considerations reinforcing the lack of published empirical research. In search conferences, the fixation on participant inputs does not encourage empirical research, which is largely the province of “experts.”
Generalization 5A

In general, the early expectation in the planned change literature seems quite definite: interventions must overcome substantial and subtle challenges. Hence, most early observers, both proponents and opponents, expected low success rates for applications in all arenas, based on an extensive rationale (Golembiewski, 1985, pp. 234–361).

Nowadays, most observers recognize that this expectation is too pessimistic by far. Table 15.1 presents more-or-less typical estimates of success rates for OD and QWL applications, respectively, and it provides no support for Generalization 5A. Far beyond Table 15.1, an extensive summary deals with about a dozen-and-a-half separate surveys of panels of evaluations of planned change applications (Golembiewski, 1998), which details similar results. See also Chapters 1 through 3.

A few words about assigning success rate categories to individual OD and QWL applications will help interpret Table 15.1. Three scorers read each case, keeping in mind that I or II assignments required 70 percent and 50 percent or more, respectively, of effects in the expected direction that are statistically significant. IV assignments went to cases with a quarter or more of effects falling in a contrary direction and/or with more than a sprinkling of statistically significant contrary effects. After each of the three individual ratings, raters discussed each case having assignments in contention. Those differences were either reconciled, or the final rating assigned was the lowest of those remaining in contention. In effect, interrater reliability is 100 percent.

Such a massive track record does not yet exist for the three other non-OD classes of approaches to planned change of concern here.

**Table 15.1** Success Rates Estimates in OD and QWL, N = 574 and 229, Respectively

<table>
<thead>
<tr>
<th>OD Outcomes</th>
<th>Success Rate Estimates, in %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I Highly positive and intended effects</td>
</tr>
<tr>
<td>I. N = 574 cases</td>
<td>40.1</td>
</tr>
<tr>
<td>II. QWL Outcomes</td>
<td>I</td>
</tr>
<tr>
<td>N = 229 cases</td>
<td>63.3</td>
</tr>
</tbody>
</table>

*Source: Based on Golembiewski (1998).*
Generalization 5B

Almost without exception, observers—both earlier and some even today—expect substantially lower success rates in public sector attempts at planned change. Underlying rationales emphasize the purportedly lesser concern with efficiency and effectiveness, as well as the assumedly greater resistance to change.

Evaluative studies do not support this view, however. In the largest population of applications so far assembled, the success rates do slightly favor business applications, but not by much. And success rates in both sectors are substantial, even formidable (Golembiewski, Proehl, and Sink, 1981). On occasion, moreover, public-sector success rates are greater (e.g., Golembiewski and Sun, 1990). See also early chapters above.

Generalization 6A

For conventional reasons, success rates at the Politics/Administration interface are typically assumed to be very low, whether the arena of application is business, government, or the voluntary sector. The underlying rationale has been described elsewhere (e.g., Golembiewski, 1985, pp. 233–361), and has features known to even casual readers of the literature on bashing public bureaucrats and bureaucracies.

Table 15.2 summarizes the success rates in the full panel of applications underlying Exhibit 15.1 and—again so much the worse for the common view—it presents no need to reconcile differences with Table 15.1. Success rates are broadly comparable at the interface in both OD and QWL applications, which will tend to occur at middle and lower levels of management, respectively. The assignment criteria are direct. Multiple readers reviewed each case, and assigned a rating of I, II, III, or IV, with the meanings reflected in Table 15.2. The reliability of final assignments is 1.0. Any assignments in dispute were discussed and, where agreement was not possible, the case was assigned the lowest rating in contention. Overall, the cases permit only impressionistic overviews of effects. Where statistical tests of numerous variables were run, a I assignment requires

<table>
<thead>
<tr>
<th>Table 15.2</th>
<th>Overall Success Rates at Interface, 41 Cases and 59 Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highly positive and intended effects</td>
</tr>
<tr>
<td>I</td>
<td>27.1</td>
</tr>
</tbody>
</table>


that virtually all changes fall in the expected direction and that those changes achieve statistical significance in perhaps two-thirds of the $T_2$ vs. $T_1$ estimates. A II assignment admits a few more unexpected changes, with half or so of the expected changes achieving statistical significance. See also Chapter 14 for further details.

**Generalization 6B**

Observers generally expect greater subtlety and complexity, if not chaos and confusion, at the public sector interface. Hence, most assume success rates there will be very much lower at the interface than in business settings. (Golembiewski, 1985).

Given the small number of cases in the business arena in the full panel of applications, public versus business comparisons can be suggestive only, and perhaps falsely so. If “success” is defined by assignments to category I plus II, applications at the interface hardly qualify as much lower in government settings than in business settings—cases 3, 6, 26, 28, 29, 31, and 38 in the full panel. Note also that Table 15.3 provides summary data for three arenas, with “Network” including panels of participants that cannot be classified as dominantly “business” or “public.” Network exemplars are cases 5, 10, 14, 18, 27, 33, 36, and 37.

**Generalization 7**

Although opinions differ, the general consensus is that planned change interventions have a substantial decay rate, even if they initially generate intended effects, which is typically seen as unlikely.

The full panel permits no certitude here and, at best, provides only mixed and weak support for this seventh guideline. Five points summarize details. First, only a minority of the cases—e.g., 3, 6, 17, 26, and 28—give explicit attention to longitudinal designs. So any conclusions about the persistence of effects over time must be hedged. See also the fourth point below.

Second, most of the cases—for example, cases 7, 8, 23, 24, 29, 32, and 39—target short-run effects. Only a minority of cases comment on longer-run pre-test effects, and even then only in general terms.

Third, for several cases—for example, cases 12, 18, and 34—longer-run effects were intended but seem either doubtful or mitigated, or both.

Fourth, however, long-run effects can hardly be said to be absent from Exhibit 15.1. This generalization applies in three distinct senses. Most prominently, the full panel underlying several cases both target and track efforts over substantial periods; See cases 3, 6, 17, 26, and 28. They support a generalization in direct opposition to Generalizations 5 as well as 6. Not only do positive effects tend to occur in these cases, but their persistence also seems marked.

In addition, expected long-run effects tend to appear even in cases in which they are not tracked in detail, as seems to hold for cases 7, 11, 13, 17, 19, 27,
<table>
<thead>
<tr>
<th>Arenas</th>
<th>I Highly positive and intended effects</th>
<th>II Definite balance of positive and intended effects</th>
<th>II No appreciable effects</th>
<th>IV Marked contrary effects</th>
<th>In process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business (6 cases; 6 estimates)</td>
<td>66.7</td>
<td>33.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Public (27 cases; 38 estimates)</td>
<td>29.0</td>
<td>52.6</td>
<td>15.8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Network (8 cases; 15 estimates)</td>
<td>6.7</td>
<td>66.7</td>
<td>20.0</td>
<td>0.0</td>
<td>6.6</td>
</tr>
</tbody>
</table>
30, 31, and 33. In such cases, one can argue that the earlier targeted interventions built a base for later development, granting the lack of specific longitudinal research.

Perhaps most intriguingly, when targeted effects did not substantially occur in the short or intermediate run, major “sleeper effects” at times appear later. For example, this seems true in cases 1, 7, and 25. In addition, while major delayed effects cannot be confidently attributed to case 22, some evidence there also suggests sleeper effects. Initially, in these four cases, the “timing was not right,” and for related reasons. One interpretation of case 1, for example, proposes that several key retirements/separations had to take place before the targeted values could be acted on, if major resistance were to be avoided. These personnel changes in State included both proponents as well as opponents of what was known as Project ACORD, or Action for Organization Development. The implied rationale seems clear: the departure of powerful persona-in-opposition—in State as well as in the American Foreign Service Association and also in Congress—seems to have broken the stalemate concerning OD values earlier considered desired and desirable by most participants in ACORD, but whose institutionalization was opposed by various political and professional interests. Note, however, that alternative explanations for such a sleeper effect in State have been proposed (e.g., Levinson, 1973, pp. 143–175).

Fifth, even when not explicitly targeted, credible long-term effects are reported by several of the applications in the full panel underlying Exhibit 15.1. For example, this seems a fair characterization of cases 30 and 38. Something of the same may be said of cases 5, 9, 10, 14, 20, and 35, if with qualifications. Reasonably also placed in this fifth category are spin-offs from an original planned intervention. These spin-offs can be at levels different from the original intervention; they have surprising but consistent targets; and they can be serendipitous inventions not anticipated by the original intervention (see Crockett, 1970; Harmon, 1975; Golembiewski and Scott, 1989).

**Generalization 8**

The dominant opinion in planned change emphasizes a cultural relativity: that different national or local cultures will have a better fit with some classes of interventions than with others. Without a doubt, Hofstede (1980a, 1980b) is most influential here. Some even bluntly refer to “North American OD” as their way of emphasizing its limited reach-and-grasp—that is, OD reflects the culture and values of the nation-states in which it developed; these culture/values can have various misfits with other nation-states, and OD designs will poorly suit those nations not involved in OD’s development.

Propositions about culture boundedness do not get much support from either the full panel underlying Exhibit 15.1 or from the few direct tests available
in the literature. Perhaps most dramatically, the overall success rates of OD interventions in developing nations do differ from North American experience, but those rates are substantially high in both settings (e.g., Golembiewski, 1990, pp. 11–24; Golembiewski, 1998). The differences no doubt include variance attributable to organizational or broadly cultural differences between one site of application or another, but those factors clearly do not dominate, if available cross-cultural OD applications provide any representative indication. That panel now approaches 250 applications. In sum, such evidence is consistent with the view that the bulk of differences—whether in cultural or organizational condition or history—were accommodated by the OD intervenors in the numerous work settings represented in the available success rate literature. See also Chapter 2.

Exhibit 15.1 relates to only 41 applications of planned change at the P/A interface, but it provides little support for the common view. Thus, the several applications directly in the OD tradition do come from the United States and Canada, but they encompass a broad range of focal issues and organizational conditions, including start-up as well as steady state. A few exceptions do exist in Exhibit 15.1—for example, cases 4, 5, and 10.

More clearly, the search conference exemplars cover a broad range, not only with respect to specific arenas but especially in connection with difference in national and perhaps cultural sites. That is, case 39 involves a Canadian province, 24 summarizes five uses in a U.S. federal agency, 30 deals with a Pakistani setting, and 32 refers to an Australian problem setting. In any case, the generalization does not seem obscure to Weisbord (1992, p. 160). He concludes that such cases suggest “that the only limits on using [future search conference] methods are to be found in our own imaginations.”

Generalization 9

The full panel represented by Exhibit 15.1 cannot pretend to be exhaustive, if only because planned change rests on an organic model that generates new and more comprehensive classes of interventions. Exhibit 15.1 is limited, but, as Table 15.4 shows, success rates are favorable.

But we can go a little further than announce stalemate. Considering classes of intervention at earlier stages of development than the four classes reflected in Exhibit 15.1, no doubt the most attractive deals with the “learning organization,” or LO. A few signs of this status of the LO metaphor can be suggested by a small handful of features. Thus, the philosophic and anecdotal support for LO is impressive, both in its volume as well as in the quality of scholars associated with that flood of published materials (e.g., Argyris and Schön, 1978; Senge, 1990; Schein, 1993). In addition, diverse applications in many organizations can be assembled under the banner of that metaphor (e.g., Watkins and Marsick, 1993), if often loosely.
TABLE 15.4  
Summaries of Success Rates at the Interface, Public vs. Business, N = 23 (Success Rate Estimates in %)

<table>
<thead>
<tr>
<th>Arenas</th>
<th>I: Highly positive and intended effects</th>
<th>II: Definite balance of positive and intended effects</th>
<th>III: No appreciable effects</th>
<th>IV: Marked contrary effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Sector Cases</td>
<td>66.7</td>
<td>33.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Sector Cases</td>
<td>31.3</td>
<td>50.0</td>
<td>18.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (23)</td>
<td>39.1</td>
<td>43.5</td>
<td>17.4</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Nonetheless, the literature associated with the LO metaphor rests on a kind of augmented face-validity: typically, materials associated with LO seem like reasonable twists on, or extensions of, so much that many in planned change have believed, written about, had strong feelings about, applied in real-time situations, or even tested (e.g., Watkins and Golembiewski, 1995). So, what is it that is missing in the “learning organization” literature—for example, from such major sources as the special Organizational Dynamics, number 2, 1996? The shortfall there clearly relates to the paucity of anything that can be labeled “research,” whether rigorous or not.

Finally, for present purposes, the LO metaphor, for all of its attractions, does not permit confidence about what is tail and what is conceptual dog, as it were. To be specific, can OD be replaced by the sense and substance of “learning organization”? The issue cannot be settled here, but early analysis implies that the metaphor is perhaps best seen as an extension of basic OD concepts and values (e.g., Watkins and Golembiewski, 1995), rather than as a replacement for them.

REFERENCES*


* These references include all citations to the original Table 15.1 (Golembiewski, 1999), even though cases 5 through 37 are omitted from the version of the table used here.


Team-building constitutes a common family of OD designs, and for good reasons (e.g., Dyer, 1987). Those designs have substantial success rates, as Chapters 1 and 4 suggest. This record implies that interveners are generally quite clever in fitting design variants and values to specific settings, despite the general theoretical failure to explicitly provide for differences likely to be encountered in large collections of small-group settings, despite attractive beginnings (Bowers and Hausser, 1977).

Broadly, ODers need to do better than to rely on their individual sensitivities to one or a few groups, if only because it is increasingly commonplace to hold team-building experiences for twenty or even a hundred or more teams exposed to the same learning design. Individual sensitivities are not sufficient for this scale of intervention. Moreover, as the technology for differentiating groups develops, so also will individual consultants dealing even with single groups be more effective in fitting specific designs to different contexts.

This need to differentiate contexts to facilitate the fitting of tailored designs is both necessary and neglected. Not only does observation suggest a diversity of types of small groups, but both the general literature (e.g., Golembiewski, 1962) and specific tests (e.g., Bowers and Hauser, 1977) imply that a substantial number of small-group varieties need distinguishing, which in turn suggests the need for a catalog of different designs. Unfortunately, with all-too-rare exceptions (Dyer 1977, 1987), the tailoring of designs to contexts has not been an OD priority.
A CONCEPTUAL CONTEXT FOR IRONY VI

This chapter tries to do better in differentiating contexts so that appropriate designs can be fitted to them. The initial approach has three emphases: an anomaly in team-building success rates will provide a rationale for introducing two ways of distinguishing between small groups. The first emphasis deals with two classic conditions in groups, and the second directs attention to differences in psychological burnout among a group’s members.

To begin, the author acknowledges the piercing revisionist view of Taras (2000), who would have points of contention with much of the analysis below. This author remains unconvinced, however, but is nonetheless eager for readers to have knowledge of a subtle treatment that differs in important particulars from the argument below.

Toward Resolving an Anomaly

Some readers of the OD literature have recently directed attention to an apparent anomaly. On the one hand, individual applications of team-building designs seem to be increasingly popular and appear to have the expected effects, on balance (Dyer 1977, 1987; De Meuse and Liebowitz, 1981; Golembiewski, Proehl, and Sink, 1981). On the other hand, simultaneous applications of a common team-building design to large numbers of work units in the same organization have fared less well (Harris and Porras, 1978; Porras and Wilkins, 1980).

Hence, experience with team-building can be fairly characterized as Janus-faced. Most observers reflect a sunny and optimistic visage. For example, De Meuse and Liebowitz (1981) review thirty-six applications usually involving single teams and, despite concern about the casual methodology of many of the studies, they see a robust and burgeoning theory-cum-technology. Other observers (e.g., Harris and Porras, 1978; Porras and Wilkins, 1980) are dour after surveying their attempts at mass team-building. Perhaps too much was expected of a brief design, goes the overall evaluation of one of these efforts (Harris and Porras, 1978). As Porras and Wilkins conclude:

Several key measures of performance . . . [show] significant improvement. . . . Yet, measures of organizational processes indicated a deterioration of the internal dynamics of the system. This was clearly . . . unexpected. . . . Other alternatives must be sought or OD will be viewed as irrelevant to large-system change (1980, 531–532).

This chapter is not deterred by the apparent anomaly. Quite the opposite, in fact. This chapter seeks to encompass both anomalous positions in a direct way that will elaborate the theory underlying team-building and that also may enrich its practice. Specifically, the two reactions sketched above may reflect the impact of two unevenly diagnosed conditions in team-building settings:
The underlying character of the conflict or tension in the subsystem, which can be highlighted by distinguishing conditions of agreement versus disagreement in small groups.

The degree of psychological burnout experienced by members of the team

Generally, then, a design useful for some teams in a mass design may be beside the point or even counterproductive for other teams in that batch. If the standard confrontations design is utilized (Dyer 1977, 1987; Golembiewski and Kiepper, 1976; Boss, 1979), for example, there is some probability that it will be inappropriate for some or even many units in the batch, while it well serves others.

Hence, enhanced diagnostic power is the goal here. Available work provides instrumentation for such diagnosis, fortunately, and it also suggests the character of appropriate designs. That is an attractive combination on which to build.

Conditions of Disagreement versus Agreement

Attention to disagreement versus agreement as a central condition in a group has a substantial history, in various forms—from Janis (1972) to Harvey (1977) to Dyer (1977, 1987). Harvey’s is perhaps the most arresting formulation, which he began expressing as “the Abilene Paradox” and subsequently refined into a comparison of a “crisis of agreement” with a “crisis of disagreement.” Table 16.1 generates the fuller sense of the critical distinction (see also Golembiewski 1979, vol. 2, pp. 151–58).

The locus of the differences in Table 16.1 seems direct: basically, concern about membership seems to dominate in the crisis of agreement—given that prestige of membership can derive from the “best and brightest” view characteristic of the Kennedy administration (Janis, 1972), or from the desire for continued membership rooted in suspicion and fear, as seems to have been the case with Nixon appointees who reasonably seemed to believe there was no other way they could attract similar power or salary (Raven, 1974). Testing for actual agreement would risk membership. In disagreement, the primal concern is one of unsatisfactory inclusion, which typically is reflected in we/they modes. Here, there is less to lose and more to gain by dealing with difference.

Appropriate interventions in the two systems will differ profoundly. In disagreement, a legitimate process for exploring differences is needed, and that need can be easily satisfied by various confrontational and interaction-oriented designs (for an example, see Chapter 4). One key source of leverage inheres in the fact that actors risk only an unattractive relationship, which may encourage boldness and motivates progress. In agreement, valued membership must be safeguarded—first and foremost, whatever the process. This is a delicate and chancy matter, and encourages hunkering-down.
### Table 16.1 A Contrast of Underlying Behavioral Processes for Two Types of Crises

<table>
<thead>
<tr>
<th>Crisis of agreement</th>
<th>Crisis of disagreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Organization members experience pain from some specific collective problem(s), with feelings about impotence or incompetence deriving from the failure to resolve or manage the problem(s).</td>
<td>1. Organization members may or may not experience pain from some specific collective problem(s), but such problem(s) do exist although consciousness raising about them may be required.</td>
</tr>
<tr>
<td>2. Members share the same private concept of problem(s) facing their organization, and individually recognize the same or similar underlying explanations or causes of the problem(s).</td>
<td>2. Members do not share the same private concept of problem(s) facing their organization, and individually have very different underlying explanations of causes of the problem(s), if any are acknowledged.</td>
</tr>
<tr>
<td>3. As individuals, many or all organization members have similar and compatible preferences for coping with the problem(s).</td>
<td>3. As individuals, many or all organization members have different and incompatible preferences for coping with the problem(s).</td>
</tr>
<tr>
<td>4. As individuals, many or all organization members see the same or similar solution as appropriate for resolving or managing the problem(s).</td>
<td>4. As individuals, many or all organization members see different or incompatible solutions as appropriate for resolving or managing the problem(s).</td>
</tr>
<tr>
<td>5. In public settings, organization members consistently do not communicate accurately to one another—about their preferences, beliefs, knowledge of causes and consequences of organization problems—and hence mutually create a false or misleading collective reality.</td>
<td>5. In public settings, therefore, organization members may deal with one another in two basic ways: Probably in a minority of cases, members will risk open conflict, hostility, and so on, attendant to the expression of disagreements. Probably in most cases, members will avoid the risk of acrimony (as by agreeing not to disagree openly on certain issues), but at the expense of suppressing real issues and conflict.</td>
</tr>
<tr>
<td>In the longer run, the probable result is a low-energy system—careful, perhaps polite, and very conscious of roles and jurisdictions.</td>
<td>In some highly prestigious groups, the facade may emphasize potency and generate a general high-energy level.</td>
</tr>
<tr>
<td>In some highly prestigious groups, the facade may emphasize potency and generate a general high-energy level.</td>
<td></td>
</tr>
<tr>
<td>6. Given a false or misleading collective reality, on definite balance, collective decisions get made that reflect neither member preferences nor their real views of reality, with the results more than likely being counterproductive both for individual and organization goals.</td>
<td>6. The openly conflictful organizations probably will be high-energy systems, with substantial but perhaps incompatible personal commitment and involvement, great but perhaps fruitless expenditures of effort, and so on.</td>
</tr>
<tr>
<td>Greater member pain is likely, and a sense of both individual and collective incompetence and impotency probably will grow.</td>
<td>7. The suppressing organization probably will be characterized by low levels of energy, commitment, and involvement. These are probably in the majority.</td>
</tr>
<tr>
<td>8. The cycle is set to repeat itself, probably with greater speed and intensity as well as with a lessened probability of corrective action.</td>
<td>8. The probability seems high that both adaptations—if for different reasons—will lead to a false or misleading collective reality, over time. Openly conflictful organizations may develop polarizations that inhibit or preclude members from communicating accurately with one another. Suppressing organizations may create the same effect by defining certain subjects as off-limits.</td>
</tr>
</tbody>
</table>

---

*Table Entry Note: The table above provides a detailed comparison between two types of crises: crisis of agreement and crisis of disagreement. The table outlines the underlying behavioral processes associated with each type, highlighting differences in the experiences, perceptions, and outcomes of the two crises.*
Although the distinction has been in the common realm for a time, and in attractive formulations, the distinction is reflected in diagnosis and design only in rare cases (e.g., Dyer, 1977, 1987). Indeed, if anything, the distinction seems rare in the shop talk of ODers. Even more so, some change agents deny the need for any diagnosis (e.g., Cooperrider, 1996, as developed in Chapter 25). And more’s the pity, for solid guides for both diagnosis and design are available. Table 16.2, following Harvey (1977, 169–171), provides a clear guide for diagnosis. It lists a set of questions for interviews with all members of a work unit before a specific team-building design is decided on. The interview schema shown in Table 16.2 is more demanding than (for example) a paper-and-pencil diagnostic instrument that might be applied to very large populations, and can be machine-scored. But the set of questions still provides considerable aid, even as it requires interviewing prior to diagnosis.

**Table 16.2** Some Questions for Diagnostic Interviews and Guidelines for Using Them to Differentiate Agreement from Disagreement

<table>
<thead>
<tr>
<th>Diagnostic Questions</th>
<th>Diagnostic Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In general, how are things going in the organization?</td>
<td>7. If answers to question 1 are consistently positive, the organization presumably does not experience conflict of disagreement. Only reporting back to confirm the health of the organization is necessary.</td>
</tr>
<tr>
<td>2. What in particular is going well?</td>
<td>8. If answers to questions 1–4 consistently differ, the organization may be presumed to be in a conflict of disagreement. A design like the confrontation design would be appropriate.</td>
</tr>
<tr>
<td>3. What are some specific organization problems which need to be solved?</td>
<td>9. A crisis of agreement may be presumed when:</td>
</tr>
<tr>
<td>5. What problem-solving actions have you and others attempted, and what were the outcomes?</td>
<td>Much rationalization about why what should be done cannot be done gets expressed in responses to questions 5 and 6.</td>
</tr>
<tr>
<td>6. If you have not taken action, what prevents your taking action to solve them?</td>
<td>Evidence is presented about actions actually taken, especially in response to question 5, that are contrary to what respondents believe should be done.</td>
</tr>
</tbody>
</table>

Prescription of a specific team-building design for disagreement seems to pose no great problem, at least as a first approximation. The usual confrontational designs should work well, and simple interview/feedback designs also could do the job (Fordyce and Weil, 1971). My preference is for designs building around Three-Dimensional (3-D) Images, which have been widely used and reported on (e.g., Golembiewski and Blumberg, 1967; Golembiewski and Kiepper, 1976; Boss, 1979. These designs share data in response to three questions:

How do we see the Other?
How does the Other see us?
How do we see ourselves?

Two or more parties can then compare their answers to such questions, in what amounts to a design encouraging telling the less-varnished truth, as can result from acting on the regenerative model of interaction—high openness, owning, and trust, as well as low risk.

Three-dimensional designs tend to work for several reasons, with two being perhaps primary. Such designs inherently provide a procedure for gathering data, which rests on specific values and empirical theory fragments. This provides the “legitimate process” that Table 16.1 sees as central in a condition of disagreement. In addition, 3-D designs seem to quickly liberate energies, which by hypothesis are used to repress conflict in disagreement. Such energy release can be exhilarating and, in combination with the newly legitimated process for problem analysis, can induce a very tangible sense of movement or progress in what was a blocked and frustrating situation. This release of energy often comes as a tangible whoosh, in fact, once the energy is no longer needed for suppressing conflict or avoiding unsatisfactory inclusion. Crudely, one guesses that the cost to benefit ratio is very favorable in conditions of disagreement. If the design works, new data and membership become available. If the design fails, one is not very much worse off.

As for prescribing a basic design for a condition or crisis of agreement, let us begin with exclusions. For openers, it seems clear that confrontational designs have significant drawbacks in such cases. Indeed, they may be generally counter-productive. Recall that continued membership is postulated as central in conditions of agreement. Hence a confrontational design might well fail basically even as it succeeds tactically in generating the public sharing of data. In short, confrontational designs raise very troubling questions for those preoccupied with membership:

What is there about us as a group that led us to publicly suppress what we in fact largely agreed about?
How can we be confident that we will not again fall into a condition of crisis of agreement?
TABLE 16.3 Sketch of an Intervention for a Condition of Agreement

1. Sort the interview data into the basic themes of agreement.
2. Themes get reported back in a public session in ways that use respondents' own words as much as possible but that protect anonymity.
3. All those interviewed then write a collective summary of all of the data supporting each theme of agreement, decide on the action implications of the themes, and plan specific actions.
4. Summarize the theory of agreement sketched in Table 16.1, to reinforce understanding of why and how members were inhibited from sharing agreements with each other.
5. Individual organization members are coached, in private, about actions they might take in light of the agreements with their colleagues.

We can be less certain about what designs will work for a diagnosed crisis of agreement, in addition, because little research or experience has been accumulated. However, Harvey (1977) seems to provide a useful way of approach in Table 16.3. Read between the lines a little, and his basic thrusts seem clear. The design's multiple checks imply care to establish that only broadly held materials are admitted to analysis. Pretty clearly, this rechecking process no doubt seeks to reduce the probability that anyone's membership is threatened, even as it seeks to raise the likelihood that not just another crisis of agreement is being developed.

The careful character of Table 16.3 contrasts markedly with the wham-bang quality of many confrontational designs. There is too much to be lost in the ideal conflict of agreement for precipitous acting-out; and even an impactful learning design could generate very mixed results. Crudely, again, the cost to benefit ratio seems relatively balanced in a crisis of agreement. Hence, the careful character of Table 16.3.

Conclusions should not be rushed beyond the reasonable suggestions of Harvey, but the argument thus far does suggest a range of other designs that are indicated, as well as some designs that are contraindicated for the conditions of agreement and disagreement. The provisional assignments shown in Table 16.4 seem reasonable for the two diagnoses of agreement or disagreement.

In a rough sense, the designs in the Condition of Disagreement column initially emphasize differentiation and, although this will often lead to integration, such a beginning emphasis could pose definite threats to membership in the group. Obviously, this constitutes no basic problem for a condition of disagreement — where the key issue is unsatisfactory inclusion with which, obviously, even extended initial differentiation is consistent. The designs in the Condition of Agreement column in Table 16.4 focus ab initio on a common product and hence will be less upsetting where the issue is continuing membership.
In addition, some designs seem to be more or less equally applicable to the two conditions. Survey/feedback and interview/feedback designs permit such flexibility. Why? Possible reasons include a subtle relocation of responsibility for what gets said, and how it gets said, on the survey, and perhaps even the disavowal by members of any and all feedback by the external intervener if it threatens membership. In one such application, for example, members of an executive team were not willing to deal with a critical situation till they heard the results of interviews, which had this tenor: “Eleven of thirteen of you agreed about X, twelve of thirteen about Y, and all of you agreed about Z.” Preserving the myth would threaten membership more than accepting the actual agreement. Or attacks on the intervener might have the paradox effect of heightening group cohesiveness.

The distinction of conditions implies theoretic and practical leverage. Consider the work that supports the “greater effectiveness” of (for example) survey/feedback as compared to laboratory training or T-grouping (e.g., Bowers, 1973). Common interpretations conclude that some designs are better than others. The present line of thought encourages a more restricted interpretation. When you cannot differentiate targets for interventions, some designs like survey/feedback may be safer in that they will not create problems in either a condition of disagreement or one of agreement. Other designs may be more target-specific, being good fits for some settings and poor fits for others. The moral of the research is not that some designs are better than others. In contrast, the highlighted need involves diagnosing the specific conditions to which various designs are more or less applicable.

**Take Care in Boarding the Bus to Abilene**

The crisis of agreement appears in several forms (e.g., Janis, 1972), but there is no question that its most popular representation derives from Jerry Harvey. That
popularity came early in the delightful format of Jerry’s family “going to Abilene” when all (or most) would have preferred sitting on the front porch with a fan and a glass of lemonade on a blazing-hot day (Harvey, 1988a, 1977). Hence the common reference to “the Abilene paradox.” (There was no air conditioning, one gathers, either at home or in the car.) That popularity has persisted, with Organizational Dynamics having reprinted Harvey’s Abilene story (1988a), along with an extended personal commentary and update (Harvey, 1988b) as well as with two notes that highlight the importance of the Abilene paradox (Kanter, 1988; Carlisle, 1988).

So popular has Abilene become that the crisis of agreement has been extended very far—probably too far. In being associated with so many situations, the formulation stands in danger of meaning less and less about more and more. Hence this effort at better targeting.

Let me back up a bit. Few (indeed, more likely, no) empirical tests exist of the ways suggested by Jerry and others for dealing with the condition. And the few cases are not only sparse, but they also are unclear about the outcomes of specific designs for dealing constructively with a crisis of agreement.

In sum, Harvey’s seminal concept has been stretched to cover situations that it fits only poorly, or even not at all. This boosterism can devalue the concept and discourage empirical inquiry. A concept can be associated with everything only at the cost of failing to clarify anything. “One size fits all” applies even less to metaphors and situational descriptions than it does to socks, where it applies poorly enough.

Specifically, at least five types of situations normally remain undifferentiated in discussions of Abilene. They distort or camouflage useful distinctions, especially when it comes to prescribing what it will take to do the job of resolving a crisis of agreement. Consider these five types of crisis of agreement:

1. Garden variety conflicts of agreement in loving groups, as in the case of Jerry’s family that made the famous trip to Abilene
2. A crisis of agreement in various formal or work groups
3. The conflict of agreement among the “best and the brightest,” like that which seems to have affected President Kennedy’s associates in the apocalyptic decision-making surrounding the Cuban missile crisis (e.g., Halberstam, 1969)
4. The conflict of agreement like that apparently existing among President Nixon’s confidants in the Watergate days, where a “war mentality” prevailed (e.g., Raven, 1974) and where “real” self-esteem seems low among those sharing Nixon’s “bunker mentality”
5. The conflict of agreement among true believers—in an ideology, a mythology, or whatever—as they come to question their orthodoxy (e.g., Festinger, Riecken, and Schacter, 1956).

These seem very different situations, and they respond to different interven-
tions—even diametrically opposed interventions. Significantly, also, the probable success rates are sometimes at a slam-dunk level. In other cases, intervening successfully approximates a last-second, eighty-foot, behind-the-ear, game-winning fling that surprises everyone.

Let me be brief in highlighting each type, and in sketching some implications of each for theory and practice.

The garden variety case, type 1, poses no great problems when I encounter it, which is far more often than types 3 and 4. I resist even referring to it as a “crisis.” That could result in some cases, of course. Far more likely, however, this type of situation provides opportunities for poking fun, and soon after the fact, very much in the manner that Harvey attributes to his family. Typically, direct confrontations will suffice, and the probable success rate is very high. The common reaction on exposure of the agreement is spoken in a bemused tone, “Well, how about that?”

Why? Basically, powerful factors in such cases cushion the severity of the dour conditions Harvey rightly associates with a real crisis of agreement. These cushioning factors in type 1 include long-standing relationships, presumed continuation of affective bonds, the multiple opportunities for face-to-face dealing with the issues and their consequences, and the usefulness of such victories in building cohesiveness.

Type 2 can cause greater problems, lacking as it does in most cases the positive relationships between family members. Role differences also can exacerbate type 2 dynamics. It does not appear to me pure fussiness to differentiate types 1 and 2, given the general tendency to lump them together.

Type 3 conflicts seem to me to have major redemptive features that help leverage change, with the most prominent feature being the lofty self-esteem of the participants in a work or formal group. Theoretical reasons and experience alike suggest that such individuals can hear surprising or even hurtful feedback better than most, and then get on with it. Direct confrontations should be able to do the job.

There are downsides to intervening in a type 3 case, of course. The fear of exclusion exists, and that can be very real. Moreover, the feedback must come from someone most group members see as staunch as they see themselves. That someone can be an insider or an outsider, but that someone must have high standing with group members.

However, the general prognosis seems good in type 3 situations. Perhaps basically, participants pride themselves on their coping skills and have the reservoir of self-esteem to recover from situations that might herniate others. I recall one joyous post-insight whoop in such a group: “When we do it, we really do it. We set a record for going to Abilene. No doubt. Let’s set some other records now.”

Type 4 poses the most serious issues, since the potential loss of membership
can be crippling if self-esteem is low and selection processes have been sufficiently diabolic. “Where could I ever make as much money, and exercise as much power, at my age and experience?” echoes a plaintive cry common to several Watergaters as they sought to explain why they did not do what could have saved them and their country from a painful and dangerous experience (i.e., tell the truth).

Why the difference between 3 and 4? In contrast to 3, here the key descriptors are negative and even passive versus active and positive—a “connection” or serendipity versus achievement in gaining membership, dependence versus a certain rambunctiousness of spirit, a sensitivity to commands versus a questioning if not counter-dependent spirit, lower versus often towering self-esteem, and a garrison mentality versus a perception of self as a competent in a hurry to get to Camelot.

In type 4, Harvey’s conceptual elaborations seem to me to apply most directly—indeed, point by point; and his prescriptions about intervening via careful and slowly unfolding designs seem right-on. Such groups can’t win for losing—usually sooner rather than later, and often with a precipitous fall from grace rather than a setback uncovered sufficiently early to contain losses, and perhaps to motivate learning. The probability of successful intervention in type 4 seems low to minuscule to me, but no definitive study exists.

Type 5 seems most problematic to me, and the vital issue seems to me whether a high-enough proportion of members, appropriately located and wired-in, come to have similar reservations about the same leading ideas, at around the same time. If so, successful change has a real chance. Otherwise, I see the chances in the range of slim to none.

Type 5 is not strictly a conflict of agreement in Harvey’s terms, in any case. Perhaps it is best described as a movement toward a possible crisis of agreement, but as having protean forms that feature aggressive denial at many stages, and with many branches in its developmental tree. An extreme case of type 5 involves a religious group, whose members have variously prepared for the end of the world, on a specific date, as by selling or giving away their possessions. Intervening here will be extraordinarily complex, because faith will sustain the true believers at least until the specific date, when deep despair or convoluted denial suddenly set in (e.g., Festinger, Riecken, and Schachter, 1956).

Whatever one calls it, or them, I see a greater frequency of type 1 than 3 or 4, and probably of 2 than 3 or 4. An example of type 5 concerns the organizational belief about the alleged felicitous qualities of “natural fire” in the national forests, whose limits were dramatically revealed in Yellowstone Park during 1988.

I suspect Gorbachev faced a type 5 situation, for example, as did the officials of the Forest Service in the late 1960s. In both cases, their respective orthodoxies portrayed them as being in great shape, just as their respective abysses
loomed ahead of them. Growing numbers came to perceive the burgeoning distance between leading ideas and reality. The Forest Service saved the baby from the bathwater—as it added to its product line, added various new specialists, diversified the gender and race of its employees, encompassed people as well as trees as clients, and generally got in better shape to compete for budgetary support. Gorbachev barely avoided assassination.

DIFFERENCES IN PSYCHOLOGICAL BURNOUT

New evidence also suggests that team-building diagnoses and designs, especially when masses of teams are involved, need to be sensitive to one major characteristic of participants as individuals and in the aggregate—their members’ degrees of psychological burnout. Much useful detail on burnout appears in Chapter 17, and a few bare details suffice here. This emphasis on individuals complements the preceding emphasis on the character of their relationships. Broadly, burnout has a central conceptual status as “a syndrome of inappropriate attitudes towards clients and towards self, often associated with uncomfortable physical and emotional symptoms” (Kahn, 1978, p. 61). In addition to reduced productivity, burnout’s dreary inventory of effects includes “job turnover, absenteeism, and low morale [as well as] various self-reported indices of personal distress, including mental exhaustion, insomnia, increased use of alcohol or drugs, and marital and family problems” (Maslach and Jackson, 1981, p. 100).

This section receives reinforcement from other chapters, and some overlap will be risked on the general premise that repetition is preferable to overlooking any significant points. To sketch the later review of a substantial body of research (Golembiewski, Munzenrider, and Stevenson, 1985; Golembiewski and Munzenrider, 1988). Progress has been made toward the development of a paper-and-pencil measure of burnout (e.g., Maslach and Jackson, 1981, 1986), but it is not clear how best to aggregate the items of the MBI. Several alternative ways have been tested, using as a basic referent twenty-two target variables thought to tap important aspects of the worksite. The variables and predictions about them can be introduced in three categories:

1. Job Descriptive Index (JDI), which measures satisfaction with five facets of work and also provides a total satisfaction score (Smith, Kendall, and Blood, 1969). As burnout increases, one expects reduced satisfaction on all JDI measures with the possible exception of JDI-Pay. The host organization’s pay policies are considered superior, generally, and satisfaction with them consequently might not differ among those experiencing various degrees of burnout.

2. Job Diagnostic Survey (JDS), which measures satisfaction with ten facets of the job (Hackman and Oldham, 1981). As burnout increases, with the possible exception of JDS-Compensation, one expects reduced satisfaction on all JDS facets.
3. Assorted scales which, with one exception, should decrease as burnout increases. The Job Tension scale (Kahn et al., 1964) should reflect a direct relationship with burnout. These other scales should decrease as burnout increases:

- Trust in supervisors (Roberts and O’Reilly, 1974)
- Trust in fellow employees (ad hoc measure)
- Job involvement (White and Ruh, 1973)
- Willingness to disagree with supervisor (Patchen, 1965)
- Participation in decisions concerning work (White and Ruh, 1973)

Tests of four ways of aggregating MBI items may be summarized briefly. First, in Table 16.5, the MBI Total Score correlates quite regularly with a number of variables describing important aspects of the worksite. That puts it mildly, in fact. All twenty-two of the correlations are in the expected directions; twenty-one of them attain the .05 level; and the significant correlations account for a substantial 13 percent of the total variance, with several cases accounting for as much as a quarter of the total variance.

Second, in Table 16.5, the three MBI subscales have marked patterns of association with the twenty-two target variables. The level of association is lower, on average, for the three subscale scores than for the Total Score. Moreover, the subscales differ substantially in their association with the target variables. Specifically, on average, the subscales singly account for the following percentages of variance:

- Emotional Exhaustion accounts for 12 percent, with subscale items tapping the degrees to which respondents feel inadequate in coping with strain in their lives, to which they are “at the end of the rope” in psychological emotional terms.
- Personal Accomplishment accounts for 8 percent, and subscale items here relate to feelings about the effectiveness of performance on a meaningful and worthwhile job.
- Depersonalization accounts for 4 percent of the variance, with its items referring to the degree to which humans and relationships with them are considered to be objects or things.

Third, burnout can be rated from I (least) → VIII (most) on the phase model. Table 16.6 depicts that model, using High vs. Low scores on the three MBI subdomains. For details, see Chapters 19 and 20. Here, the phases isolate nonrandom variation on almost all of the target variables described above, almost all of the time, in all of the several tests run thus far. On average, eighteen or nineteen of each twenty variables on which significant variation was expected did generate F-ratios that surpass the .05 level. In addition, paired-comparisons establish the robustness as well as regularity of the differences in target variables isolated by
| Table 16.5 Correlations between Total and Subscale MBI Scores and Twenty-two Scales |
|---|---|---|---|---|
| | Alpha | Total score | Depersonalization | Emotional exhaustion |
| I JDI Scales: Satisfaction with | | | | |
| Work | .80 | -.51<sup>a</sup> | -.26<sup>a</sup> | -.34<sup>a</sup> | -.49<sup>a</sup> |
| Supervision | .83 | -.37<sup>a</sup> | -.21<sup>a</sup> | -.20<sup>a</sup> | -.37<sup>a</sup> |
| Coworkers | .86 | -.24<sup>a</sup> | -.16<sup>a</sup> | -.25<sup>a</sup> | -.29<sup>a</sup> |
| Promotion | .90 | -.28<sup>a</sup> | -.12 | -.16<sup>a</sup> | -.29<sup>a</sup> |
| Pay | .92 | -.15<sup>a</sup> | -.01 | -.07 | -.19<sup>a</sup> |
| Total JDI | .73 | -.49<sup>a</sup> | -.25<sup>a</sup> | -.33<sup>a</sup> | -.45<sup>a</sup> |
| II JDI Scales: Satisfaction with | | | | |
| Experienced meaningfulness of work | .78 | -.39<sup>a</sup> | -.20<sup>a</sup> | -.30<sup>a</sup> | -.31<sup>a</sup> |
| Experienced responsibility for work | .66 | -.19<sup>a</sup> | -.12<sup>a</sup> | -.26<sup>a</sup> | -.05 |
| Knowledge of results | .71 | -.26<sup>a</sup> | -.23<sup>a</sup> | -.12 | -.20<sup>a</sup> |
| Work in general | .75 | -.51<sup>a</sup> | -.26<sup>a</sup> | -.33<sup>a</sup> | -.50<sup>a</sup> |
| Work motivation | .58 | -.26<sup>a</sup> | -.13<sup>a</sup> | -.32<sup>a</sup> | -.08 |
| Growth | .84 | -.51<sup>a</sup> | -.17<sup>a</sup> | -.32<sup>a</sup> | -.32<sup>a</sup> |
| Job security | .84 | -.33<sup>a</sup> | -.19<sup>a</sup> | -.20<sup>a</sup> | -.26<sup>a</sup> |
| Compensation | .86 | -.13<sup>a</sup> | -.03 | -.02 | -.15<sup>a</sup> |
| Coworkers | .67 | -.36<sup>a</sup> | -.18<sup>a</sup> | -.33<sup>a</sup> | -.24<sup>a</sup> |
| Supervision | .90 | -.34<sup>a</sup> | -.18<sup>a</sup> | -.17<sup>a</sup> | -.33<sup>a</sup> |
| III Selected Scales | | | | |
| Trust in supervisors | .83 | -.43<sup>a</sup> | -.23<sup>a</sup> | -.29<sup>a</sup> | -.39<sup>a</sup> |
| Trust in fellow employees | .78 | -.36<sup>a</sup> | -.27<sup>a</sup> | -.24<sup>a</sup> | -.26<sup>a</sup> |
| Job involvement | .77 | -.45<sup>a</sup> | -.25<sup>a</sup> | -.42<sup>a</sup> | -.33<sup>a</sup> |
| Willingness to disagree with supervisor | .70 | -.12 | -.02 | -.25<sup>a</sup> | -.02 |
| Job-related tension | .81 | .40<sup>a</sup> | .23<sup>a</sup> | .13<sup>a</sup> | .42<sup>a</sup> |
| Participation | .76 | -.31<sup>a</sup> | -.08 | -.34<sup>a</sup> | -.25<sup>a</sup> |


Note: *Indicates a correlation coefficient statistically significant at .05 level.
TABLE 16.6 Phase Model of Burnout

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depersonalization</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
</tr>
<tr>
<td>Personal accomplishment (revised)</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
</tr>
<tr>
<td>Emotional exhaustion</td>
<td>Lo</td>
<td>Lo</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
<td>Hi</td>
<td>Hi</td>
</tr>
</tbody>
</table>

the phases. Depending on the specific site tested, as many as 60 percent of the
paired-comparisons of differences on the target variables attain statistical signifi-
cance. Details are available in several sources (e.g., Golembiewski, Munzenrider,
and Stevenson, 1985; Golembiewski and Munzenrider, 1988; Golembiewski,
Boudreau, Munzenrider, and Luo, 1996).

The phase-out seems to have significant implications for team-building de-
signs, particularly in mass populations. For teams with all or many members in
phases II and IV, for example, standard interaction-centered or confrontational
designs should suffice. In those phases, individuals score High on Depersonaliza-
tion, and interaction designs should be helpful in arresting or reversing such a
trend. Here new stimulation would be effective, in general: individuals receive
heightened feedback and disclosure concerning the effects of their viewing others
as objects; and, ideally, individuals also come to feel the rewards of greater accep-
tance and liking that typically follow as a consequence of responding to others
as persons.

For more advanced burnout phases, however—let us say, phases VI
through VIII—interaction-oriented designs seem awkward, even counterproduc-
tive. Individuals already have a surfeit of negative stimuli and may even possess
a response-set that encourages them to perceive neutral or even positive stimuli in
negative terms. Here, the dynamics characteristic of interaction-centered designs
might result in overstimulation, including even the potential for emotional bomb-
arding escalating to nonhelpful or even hurtful proportions. Looked at from
another perspective, cases of advanced burnout might be particularly responsive
to policy and structural amelioration—as in reductions in role-overload, the
greater clarification of roles, and so on. The evidence is far from conclusive on
the point, but some experience in day-care centers supports the present position.
There, staff with high burnout responded positively to various restructuring ef-
forts—for example, job rotation and rescheduling the children’s day so as to ease
the demands on the adult staff (Pines and Maslach, 1980).

GREATER SUCCESS BY SPECIFYING KINDS OF CRISES
AND MEMBERSHIPS

Will the specification of kinds of conditions in groups, as well as the distributions
of the burnout phases of their memberships, increase the success rates for OD
interveners? The implied support for an affirmative answer seems strong. The failure to make these two differentiations promises mischief.

In this spirit, let us bring these introductory considerations to three tentative conclusions. In turn, they sketch the basic orientation of this chapter, emphasize the possibility of the convenient instrumentation for the two kinds of differences emphasized here, and sketch some guidelines for designs sensitive to kinds of group conditions and associated burnout in groups.

Significance of Mass Team-Building

There seem to be ample reasons for interveners to broaden their conceptual perspectives to include two classes of differentia—agreement versus disagreement and degree of burnout—when considering a team-building design. This seems apt counsel even when an intervener works with a single group or team. Mass team-building efforts positively require such specification. This bottom-line conclusion alone justifies the demonstration here.

Mass team-building efforts are no longer exotic, please note, and concerns about their efficacy constitute one of the major challenges to OD’s future credibility. For a time, very few examples of mass team-building could be found in the literature (e.g., Golembiewski, 1978, vol. 2, pp. 131–138). But more recently, their incidence has increased substantially (Harris and Porras, 1978; Porras and Wilkins, 1980), as have reports about their mixed or unexpected results. By hypothesis, these surprises are provisionally attributed to the common imposition of a standard design on mass team-building populations, whose probable heterogeneity will have an uneven fit with standard designs. As the number of teams increases, roughly, so will variant conditions appear to which (for example) the usual confrontational designs adapt poorly. With single teams or small numbers of units, in contrast, alert interveners will be able to adapt standard designs to the specific texture and flow of individual teams. Here, the success rates are substantial (e.g., De Meuse and Liebowitz, 1981). For example, I have reported rates in changing group variables in individual teams via a standard design that approximate 80 percent (Golembiewski, 1978, vol. 2, pp. 338–344).

The basic notion about group heterogeneity underlying this chapter should not astound or even surprise. Even brief consideration—which can be reinforced by sophisticated research (e.g., Bowers and Hausser, 1977)—strongly suggests that work units are not essentially homogeneous. No one yet knows, but as many as a score of subtypes may be needed to encompass the diversity that exists in nature (Bowers and Hausser, 1977)—hence the lower success rates of applying standard team-building designs to large populations of formal work units.

Availability of Instrumentation

Fortunately, convenient approaches exist for testing burnout as well as for agreement versus disagreement. The paper-and-pencil instrument developed by Mas-
lach—her Burnout Inventory, or MBI—is easily applied in large populations. Boudreau and Golembiewski (2002) have developed an attractive measure—Boudreau’s Burnout Questionnaire or BBQ. Harvey’s guidelines for interviews also can be applied to large batches of work units. Clearly, this requires some effort and skill. But even brief interviews with members of teams-to-be-built seem useful, on general principles. Moreover, OD’s credibility depends on increasing effectiveness in making desired or predicted things happen. So convenience cannot be a determinative criterion, even though it remains a reductive one.

**Different Designs for Different Teams**

Provisionally, interveners ought to keep a 2 x 2 grid in mind as they diagnose team-building populations. See Figure 16.2, which supports a brief sketch of dominant properties of the four grid quadrants and also suggests implications for appropriate OD designs. Broadly, the point is that OD successes will increase, and failures become less common, as different designs are fitted to the context-sensitive quadrants of Figure 16.1.

**Quadrant I**

This is likely to be a high-energy condition, at least initially. Thus, the condition of disagreement should induce some frustration, which can be energizing for at least a time, and may indeed erupt into open conflict. In addition, low levels of burnout imply few energies bound up in symptomology and its consequences, but also some emotional slack to absorb disturbances.

Here the full battery of conventional high-stimulus designs seems applica-

![Figure 16.1 A provisional grid for team-building diagnoses.](image-url)
ble (e.g., Golembiewski, 1982), including interpersonal confrontations and sharing third-party images.

Quadrant II

Quadrant II will probably be characterized by a reduced but still substantial energy level. The positive appeal of membership should generate substantial flows of energy, discounted to a meaningful degree by the prevailing condition of agreement. The low burnout implies no great drain on the basic energy level.

Designs like Harvey’s may be appropriate here. Confrontational designs threaten membership, and thus might seriously reduce the energy available for isolating and solving problems. Harvey’s design is slow-moving and this may conserve energy, but it lacks the potential for the quick great-leaps-forward characteristic of confrontational designs.

Quadrant III

Quadrant III seems a more difficult target for change than I or II. The energy level will probably be low, but the potential for substantial mischief increases.

A two-phase design may be appropriate here. Structure- or policy-oriented designs could be used to arrest or reverse burnout, to begin, and especially low-stimulus designs such as flexible work hours, role negotiation (Harrison, 1972), and job rotation. Standard confrontational designs then might be applied. Such efforts seem likely to reinforce one another. Any increases in energy due to amelioration of burnout could be applied to the condition of disagreement, and the possible bursts of energy from such progress could in turn reduce burnout.

Quadrant IV

Quadrant IV seems the chanciest condition, especially if feelings of low self-esteem or self-worth characterize team members. In such a case, their attachment to membership might be desperate, and hence their willingness to risk its loss might be exceedingly low. A two-stage design seems indicated in IV, as in quadrant III, but the prognosis seems less favorable. The key question is whether prior amelioration of burnout can release sufficient energies to risk the threat to membership associated with any choice or change under the condition of agreement.

These general musings rest unevenly on theory and practice, especially in the case of the Abilene paradox. For example, it appears that there are two modes of response to each burnout phase. Advanced burnout, active mode, seems to permit greater leeway in the application of conventional OD designs, most of which are high stimulus and arousing. Advanced burnout, passive mode, encourages a more cautious approach, as the notes about quadrant IV outline. It also appears that most of those in advanced phases—phases VI, VII, and VIII—are passive. Active/passive distinctions have been variously made, as by above median scores on conventional measures of Job Involvement.
The distinctions above should be confirmed by the experiences of many interveners, and here offer a case-in-point. Consider one example from approximately 15 years ago, which came as something of a shock after a number of early successes with a Three-Dimensional Image design (e.g., Golembiewski and Blumberg, 1967). This design basically proposes that two or more participating work units develop three lists for sharing in response to three questions to members of two teams:

How do we see our work team?
How do we see the other work team?
How do we believe the other work team sees us?

The design’s intent is clearly confrontational, and seeks to orchestrate a mutually empathic escalation based upon (if for different reasons) both successes and failures in describing selves as others see them in the example in Table 16.7. The “successes” of shared perceptions would enhance self-esteem and in effect build bridges between members of the two work units. The “failures” also would be there for both groups to see, with the knowledge being gained and with reward and reinforcement consequently characterizing the learning environment.

Matters turned out differently—in one case, radically. The two groups were right-on in their mutual perceptions, basically, as Table 16.7 illustrates (using only one of the three cross-comparisons possible via 3-D images). Team A was the top-level Operating Committee (OC), and the larger Team B was composed of the directors who were the first to report of the OC members. After much prodding, the two groups separately developed images—two of which Table 16.7

<table>
<thead>
<tr>
<th>How team B sees team A</th>
<th>How team A believes team B sees them</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Not communicative enough</td>
<td>1. Divergent group</td>
</tr>
<tr>
<td>2. Floundering, indecisive</td>
<td>2. Lack of evidence of authority and decisiveness</td>
</tr>
<tr>
<td>3. Defensive, unreceptive</td>
<td>3. Spend much time on non-key issues</td>
</tr>
<tr>
<td>4. Too involved in day-to-day operations and</td>
<td>4. Poor catalysts</td>
</tr>
<tr>
<td>decisions</td>
<td>5. Lack of communicating a sense of direction and purpose</td>
</tr>
<tr>
<td>5. Cautious</td>
<td>6. Nice guys!!? (i.e., not competent)</td>
</tr>
<tr>
<td>6. Under tight Corporation control (i.e.,</td>
<td>7. Too much resistance rather than encouragement</td>
</tr>
<tr>
<td>restricted and at a competitive disadvantage</td>
<td></td>
</tr>
<tr>
<td>for corporate resources)</td>
<td></td>
</tr>
</tbody>
</table>
illustrates—that were more or less unanimous expressions of previously private materials. The result? General silence and apathy followed considerable effort at analysis of what happened and why. “It surprised me that pretty much everyone saw matters as I did,” reported one participant. “It was not our style to talk about such things. My reaction was despair. All of us saw things as awful, I learned, and we all did nothing. I felt communally flaccid afterward; I always had felt individually impotent. We huddled together in our common nakedness, more fearful than ever of acting.”

In retrospect, diagnosis was lacking in this case. I would now see the situation as a Type 4 in Figure 16.1—a condition of agreement, reinforced by advanced burnout, with most of the executives in the passive mode. We intervenors, in effect, failed to diagnose IV but we also disingenuously applied a type I design. The common and consensual disclosure did not release energies but instead bound them more tightly to preserving the increasingly tenuous membership. Even that strained comfort was soon forfeited when the corporation made wholesale personnel changes.

IMPLICATIONS FOR IRONY V

Just as its three predecessors, Chapter 16 contributes to Irony V. Applications are numerous and success rates are high in OD and QWL, both in the public and business sectors. All four chapters provide numerous details on how the number of applications could be increased, purely and simply. Less conclusively, but intriguingly, the four chapters suggest that success rates also might be improved, or at least maintained, and details how that might be done.

REFERENCES


A Big Contributor to Heightened Success Rates, Almost Undoubtedly
OD Designs Improve Group Properties and Reduce Burnout

This chapter preaches an old message to readers of the previous chapters and adds technology to it: OD can profit from specifying differences between the targets of interventions. People in advanced phases of burnout seem to be characterized by (for example) progressively greater physical symptoms. Given the implied differences in personal slack in responding to changes, different OD designs and expectations about their effects seem appropriate for persons in (let us say) phase I versus phase VIII. Operational definitions appear below.

Hence, OD can and should go beyond mere description, as important as that is. Initial profiling of a target population for burnout and its context should be extended to fine-tuning OD designs, as Chapter 16 proposes. This chapter takes the next compound step: it not only illustrates up-front analysis but focuses on a specific design, as well as tests for its effects, and details some conclusions about increasing OD success rates by taking into account the degrees of burnout of individuals within a specific group context. In contrast, Chapter 16 deals with the same two features, but in general terms.

The specific focal points here are burnout and group properties, with the goal of responding to an initial diagnosis so as to increase the levels of personal slack available to individuals in small systems. This contrasts sharply with merely profiling what exists. The broader purpose here is to do something about what exists, so as to ameliorate burnout and its consequences as well as to change group properties.
Increasing rather than decreasing personal slack, as represented by the phases of burnout, will not come easily. Despite the profusion of advice on how to deal with burnout, few direct interventions exist and almost none provide data sufficient to assess effects. Kilpatrick (1986) isolates 661 published items on burnout, of which 138 (or 40 percent) offer how-to prescriptions. However, only 132 of the 661 cases rest on data, and a meager four provide data from conscious interventions (Anderson, 1982; Haack, 1980; Pines and Aronson, 1983; Slutsky, 1981). Matters have improved a bit in later citations, but more changes need to be motivated by this analysis.

Of necessity, then, this chapter seeks to illustrate how to explore largely new territory. This research blends prescription and research in a Human Resources (HR) corporate group, which initially numbered thirty-one. Conventional OD designs reduce burnout, as well as improve group properties and turnover. All three effects should serve to reduce the stressors acting on people, and thus should increase their emotional and psychological margin of comfort in responding to life’s exigencies. Put in other terms, this chapter illustrates another approach to increasing OD success rates, in human and organizational terms.

Six major emphases describe this action research. Thus, crucial associations of stress and burnout with OD will be drawn, a conceptual approach to burnout phases will be sketched, methods will be detailed, a mini-history will emphasize milestones in the HR unit’s multiple transitions in response to its demanding work environment, active interventions will be reviewed, and results of the interventions over nearly three years will be highlighted.

CONCEPTUAL CONTEXT FOR IRONY VI, REVISITED

Briefly, the world is full of “stressors,” or stimuli that upset the “constancy of the [person’s] internal environment” (Selye, 1956, p. 27). Stressors can herniate or energize—with effects specific not only to different individuals but to different points in time, as well as in different contexts. In sum, stressors consequently can create strain and distress, which induce deficits or deficiencies in people and their relationships, but stressors also can generate constructive energies, or “eustress.”

We neither can nor should avoid stressors, but we can experience so many of the wrong kinds of stressors as to generate strain and then burnout. There, normal coping skills and attitudes become insufficient to handle the stressors experienced.

OD’s interest in burnout is—or should be—specific and practical. Specifically, OD’s basic orientation toward choice and change exposes individuals to potential strain. Moreover, practically, a battery of OD designs and experience with them stand ready to guide data-based efforts to reduce burnout (e.g., Golem-
biewski, 2000), and failure to tap this potential deprives OD of a major target for applications in moving with effect toward the vision of the responsibly free workplace detailed in earlier chapters.

But how to be specific and practical? Two contributors to Irony VI—Chap-
ters 16 and this one—orient the reader to the intricacies of an ongoing program of reseach with an eight-phase model. For immediate purposes, attention is di-
rected at two features of that research: the range and depth of the impact of burnout, and the apparent conceptual roadblock the phase model places in the path of OD applications.

Range and Depth of Impact
The phase model argues that people in advanced phases of burnout are character-
ized by low levels of energy, self-esteem, and efficacy. The mainline linkages seem direct, for physical symptoms as well as other human deficits or deficien-
cies. Worksite features can induce advanced phases of burnout which, in turn, are associated with such effects as a high incidence of reported physical symp-
toms. And the latter can then have feedback effects on both the worksite and on the burnout of people in it.

Advanced phases of burnout also seem to have a startling incidence and persistence. Worldwide (Golembiewski, 2000) over 40 percent of the phase as-
signments are to Phase VI–VIII, with a higher incidence characterizing respon-
dents from non–North American worksites. Moreover, advanced phase assign-
ments seem quite stable over the period of a year in about 75 percent of the cases (Golembiewski, Munzenrider, and Stevenson, 1986, pp. 135–139).

Apparent Roadblock to OD
Such effects suggest a high-leverage target for OD, but the response has been underwhelming—and for a very good reason. Obviously, the virulence, perva-
siveness, and persistence isolated by phase research constitute formidable chal-
 lenges. So why the neglect? Consider one informed speculation.

Although ODerst often recommend high-stimulus designs for advanced burnout—interpersonal encounters, confrontations, interpersonally oriented team building, and so on—the condition does not necessarily provide the coping slack that such designs require. Tersely, traditional interventions risk overstimulation. This seems to be a significant limitation on the use of OD because its theory and practice emphasize high-stimulus interventions (Golembiewski, Munzenrider, and Stevenson, 1986, pp. 191–194). The concern will be especially strong for those many ODerst who remember the single major criticism of some forms of sensitivity training or encounter: that they are high-stimulus designs capable of propelling vulnerable persons into a “stimulus overload” condition (e.g., Lieberman, Yalom, and Miles, 1973).
TABLE 17.1 Two Classes of OD Designs

<table>
<thead>
<tr>
<th>High-Stimulus Designs</th>
<th>Low-Stimulus Designs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpersonal confrontations</td>
<td>Time off from work</td>
</tr>
<tr>
<td>Interpersonally oriented team building</td>
<td>Flexible work hours</td>
</tr>
<tr>
<td>T-groups for intact work groups</td>
<td>The kinds of job rotation that require variable intensity and provide variety</td>
</tr>
<tr>
<td>Basic policy or structural change</td>
<td>Confrontive “stress management workshops”</td>
</tr>
<tr>
<td>Confrontive “stress management workshops”</td>
<td>Mild role negotiation</td>
</tr>
</tbody>
</table>


This conceptual impasse seems avoidable, in part, because two modes of individual response seem to exist to all phases of burnout—active, and passive or withdrawn modes. Directly, the distinction suggests two intervention strategies. When advanced burnout is associated with a passive mode, coping requires creating slack in an already overloaded system. Provisionally, then, interveners might well distinguish the two classes of designs shown in Table 17.1. Moreover, interveners also might consider a two-stage strategy for ameliorating advanced burnout, passive mode: to begin to reduce burnout via low-stimulus designs; and then, beyond some as-yet-undetermined level, to intervene with a broad range of designs to further reduce burnout and to anchor it at lower levels.

To illustrate, the research below uses standard high-stimulus OD designs in focusing on a headquarters HR staff group—initially numbering thirty-one—all but two of whom maintained an active orientation to their work even though they were in advanced phases of burnout. The key assumption is that “actives” retain the emotional slack to respond to high-stimulus interventions.

**METHOD FOR ASSESSING THE IMPACTS OF ACTION PLANNING**

Archival data about turnover and three kinds of self-report data assess effects of this action research. One kind of self-report permits assigning individuals to eight phases of burnout; the second kind supports a characterization of the present HR population as “active” or “passive”; and the final kind of self-report describes the immediate work setting.

**Phases of Burnout**

Other chapters provide detail about the phases, so two reminders here should suffice. The reader will remember that the phase model rests on the three subdomain items on Maslach’s Burnout Inventory (Maslach and Jackson, 1982, 1986):
Depersonalization, high scores on which indicate individuals who distance self from others and who view other humans as objects or things;

Personal Accomplishment (reversed), low scores on which characterize individuals who believe they are doing poorly on a task that is not particularly worth doing;

Emotional exhaustion, high scores on which indicate persons who are near or beyond “the ends of their ropes” in psychological and emotional terms, who are beset by stressors beyond their comfortable coping limits.

Operationally, moreover, High and Low scores on each subdomain are distinguished, using norms from a large population (Golembiewski and Munzenrider, 1984). Assuming the progressive virulence of the subdomains in the order given above, the High versus Low distinctions generate the eight-phase model shown below.

<table>
<thead>
<tr>
<th>Phases of Burnout</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depersonalization</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
</tr>
<tr>
<td>Personal accomplishment (reversed)</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
</tr>
<tr>
<td>Emotional exhaustion</td>
<td>Lo</td>
<td>Lo</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
<td>Hi</td>
<td>Hi</td>
</tr>
</tbody>
</table>

Mode of Adaptation

How to estimate active versus passive modes remains unsettled. Here, three Work Environment Scales (WES) are used to make the active versus passive judgment: Involvement, Autonomy, and Task Orientation (Insel and Moos, 1974). Operationally, mean standard scores greater than 50 on each of the three scores define “active,” based on national norms for “general work settings” (Moos, 1981, pp. 27–28). By this criterion, twenty-nine of the thirty-one initial HR members qualify as “active.” Specifically, the two lowest-scoring persons average 90 when their three WES scores are combined. These two individuals do not seem extremely “passive,” then, although they clearly differ from other HR members. The twenty-nine other HR members average over 180, and over 93 percent of their seventy-seven individual WES scores are greater than 50.

Work Environment Scales

Worksite features are measured by the sixty WES items, which track ten dimensions. Briefly (Moos, 1981, p. 2), the WES dimensions include:
Involvement: the extent to which employees are concerned about and committed to their jobs

Peer cohesion: the extent to which employees are friendly and supportive to one another

Supervisor support: the extent to which management is supportive of employees and encourages employees to be supportive of one another

Autonomy: the extent to which employees are encouraged to be self-sufficient and to make their own decisions

Work pressure: the degree to which the processes of work and time urgency dominate the job milieu

Clarity: the extent to which employees know what is expected of them in their daily routine, and the degree to which rules and policies are communicated explicitly

Control: the extent to which management uses rules and pressures to keep employees under control

Innovation: the degree of emphasis on variety, change, and new approaches

Physical comfort: the extent to which the physical surroundings contribute to a pleasant work environment

Task Orientation: the emphasis on good planning, Involvement and Peer Cohesion, perhaps most prominently. See also the concluding paragraphs in this chapter. Task Orientation provides some sense that OD is not viewed here as merely positive affect but also as a goad to performance

WES measurement properties are conveniently available (Moos, 1981) and, quite directly in most cases, high scores on all are consistent with OD values and approaches, excepting Control and Work Pressure. The point has a face-validity for some WES scores—for example, efficiency, and getting the job done.

The target here is a number of people leaving the firm divided by average HR headcount. Other measures—such as intent to leave—would provide less conservative estimates of the commitment to continued employment, but they were not authorized in this case.

Finally, for introductory purposes, the general associations of WES scales and the phases seem obvious. Thus, among others, increases in Peer Cohesion and Supervisory Support should tend to lower Depersonalization, high scores on which are gateways for advanced phases. In addition, Personal Accomplishment should be buoyed by several WES scales, including Involvement, Autonomy, Work Pressure, and Control, primarily. And heightening scores on Emotional Exhaustion would be inhibited by many favorable WES scores—perhaps especially Supervisor Support, Peer Cohesion, and Autonomy.

Notes about Uses of Data

The four kinds of data are here used to estimate the impacts of interventions, basically, and the intervention’s public focus was on “a major morale problem,”
The self-report data were gathered following this survey schedule (see also Table 17.2 for a schedule of non-survey events):

Day 45: Administration I
Day 200: Administration II (phases only)
Day 295: Administration III
Day 425: Administration IV
Day 575: Administration V

and on what the HR personnel could do collaboratively to alleviate this problem.

Table 17.2  A Schedule of Non-Survey Events

<table>
<thead>
<tr>
<th>Day</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Problem sensing</td>
</tr>
<tr>
<td>45–75</td>
<td>Baseline data gathering</td>
</tr>
<tr>
<td>115</td>
<td>Begin data flow</td>
</tr>
<tr>
<td>140–170</td>
<td>Initial action planning</td>
</tr>
<tr>
<td>200</td>
<td>Continued action planning</td>
</tr>
<tr>
<td>230</td>
<td>Presentation to corporate HR oversite committee</td>
</tr>
<tr>
<td>265</td>
<td>Major policy intervention</td>
</tr>
<tr>
<td>295</td>
<td>Review, extend, and recommit</td>
</tr>
<tr>
<td>475</td>
<td>Major reorganization</td>
</tr>
<tr>
<td>575</td>
<td>Report on all changes</td>
</tr>
</tbody>
</table>

- Human Resources (HR) managers express fears
- Solicitation by memo of “concerns” from all employees
- Share concerns with all HR staff to encourage reaction and confrontation, when necessary
- Four “interest groups” of volunteers focus on special targets for improvement
- Interest groups report recommendations to all HR staff
- HR members each report back on “five recommendations we would most like to see implemented”
- Begin planning to implement high-priority recommendations
- HR turnover rater for 1984 higher than expected
- HR Career Progression Plan proposed
- HR Career Progression Plan announced
- Ten-hour meeting of all HR staff to review progress, plan future initiatives, and celebrate achievements
- “Chunking” occurs: several “strategic operating areas” replace a basically functional structure (see especially Chapter 5)
- Share all data on administrations I–IV
- Two-hour review and planning session for adapting to the newly decentralized HR activities.
Since the major interventions occurred through day 265, administrations I and II may be considered as long and short pretests, respectively. Administrations III, IV, and V constitute post-tests, and are conveniently labeled as one of three kinds of post-test: short, long, and very long, respectively. Turnover data became available at approximately day 90 and day 455. Participants were promised a full briefing on all aggregate measurements after implementing their action plans. This occurred at day 575, just after administration V.

THUMBNAIL HISTORY OF HR TRANSITIONS

Both HR and its corporate home had grown sharply and raised their goals over the preceding years. To illustrate, mid-1981 saw a formal HR integration of two organizationally separate activities—dealing with operational matters like recruitment and employee relations versus training and development, basically—whose close collaboration was required to meet the firm’s goals of major expansion while lowering turnover via better recruitment, training, and organization development. Corporate turnover in both 1979 and 1980 approximated 43–45 percent; it fell in the next three years to 24, 18, and then 14 percent in 1983, even as total employment increased by more than 25 percent.

Still, in early 1984 the HR management team feared that such progress often came at the expense of neglecting important work within HR. Staff were strained, and the VP and his directors expected “high turnover.” Expectations concerning HR’s contributions escalated, and so did the workload.

This concern birthed an OD effort whose major events are detailed in Table 17.2. (The table does not include the five survey administrations described immediately above.) The earliest survey data confirmed the initiating concern. As rows in Table 17.3 indicate, administration I saw over 50 percent of the HR staff in burnout phases VI, VII, and VIII. The percentages are similar for that subset of fourteen who responded to the first four surveys, but the data are not reported in Table 17.3 to conserve space. In addition, 75 percent of the total staff rate as phase V or greater, and only one person among the two top levels of management scores less than phase V. This not only seems high, it was the second-least desirable profile of phases observed at that time (Golembiewski, Munzenrider, and Stevenson, 1986, pp. 127–134).

The WES scores provide no more reason for cheering in a corporation proud of its first-class status. See Table 17.4, columns 1 and 4. Recall that a standard score of 50 indicates a raw score at the mean of respondents from a large cohort in “general work settings” (Moos, 1981, p. 27). HR Peer Cohesion was especially low, for example, and Work Pressure approximated the maximum possible score.

The 1984 HR turnover rate also proved to be high—37 percent, as Ta-
<table>
<thead>
<tr>
<th>Administration</th>
<th>N</th>
<th>I 22%</th>
<th>II 30%</th>
<th>III 33%</th>
<th>IV 29%</th>
<th>V 46%</th>
<th>VI 50%</th>
<th>VII 50%</th>
<th>VIII 49%</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>31</td>
<td>19%</td>
<td>0</td>
<td>3</td>
<td>23</td>
<td>23</td>
<td>10</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>34</td>
<td>18%</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>11</td>
<td>21</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>III</td>
<td>36</td>
<td>33%</td>
<td>6</td>
<td>11</td>
<td>0</td>
<td>11</td>
<td>22</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>IV</td>
<td>35</td>
<td>29%</td>
<td>6</td>
<td>14</td>
<td>0</td>
<td>11</td>
<td>20</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>V</td>
<td>37</td>
<td>35%</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>14</td>
<td>19</td>
<td>14</td>
<td>5</td>
</tr>
</tbody>
</table>
### Table 17.4 Work Environment Scales, Standard Scores, on Three Administrations

<table>
<thead>
<tr>
<th>WES Dimensions</th>
<th>Administration I (N = 31)</th>
<th>Administration II (N = 36)</th>
<th>Administration III (N = 35)</th>
<th>MATCHED Respondents, (N = 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Involvement</td>
<td>56.3</td>
<td>59.2</td>
<td>58.8</td>
<td>55.5&lt;sup&gt;a&lt;/sup&gt; 60.4&lt;sup&gt;a&lt;/sup&gt; 59.1</td>
</tr>
<tr>
<td>Peer Cohesion</td>
<td>33.6&lt;sup&gt;a&lt;/sup&gt;</td>
<td>46.6&lt;sup&gt;a&lt;/sup&gt;</td>
<td>44.5&lt;sup&gt;f&lt;/sup&gt;</td>
<td>32.6&lt;sup&gt;fe&lt;/sup&gt; 39.6&lt;sup&gt;g&lt;/sup&gt; 42.7&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Supervisor support</td>
<td>42.4&lt;sup&gt;fg&lt;/sup&gt;</td>
<td>52.0&lt;sup&gt;g&lt;/sup&gt;</td>
<td>50.6&lt;sup&gt;fg&lt;/sup&gt;</td>
<td>43.6 48.3 49.9</td>
</tr>
<tr>
<td>Autonomy</td>
<td>54.1</td>
<td>55.6</td>
<td>57.8</td>
<td>55.6&lt;sup&gt;f&lt;/sup&gt; 60.2 63.1&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Task orientation</td>
<td>61.5</td>
<td>62.9</td>
<td>61.4</td>
<td>61.6 60.3 60.2</td>
</tr>
<tr>
<td>Work pressure</td>
<td>80.9&lt;sup&gt;mh&lt;/sup&gt;</td>
<td>75.6&lt;sup&gt;f&lt;/sup&gt;</td>
<td>73.0&lt;sup&gt;h&lt;/sup&gt;</td>
<td>82.5&lt;sup&gt;g&lt;/sup&gt; 82.0&lt;sup&gt;i&lt;/sup&gt; 77.9&lt;sup&gt;gt&lt;/sup&gt;</td>
</tr>
<tr>
<td>Clarity</td>
<td>50.1</td>
<td>48.9</td>
<td>52.5</td>
<td>40.1&lt;sup&gt;c&lt;/sup&gt; 41.5&lt;sup&gt;ch&lt;/sup&gt; 50.8&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>Control</td>
<td>55.1&lt;sup&gt;d&lt;/sup&gt;</td>
<td>49.9&lt;sup&gt;d&lt;/sup&gt;</td>
<td>51.4</td>
<td>52.1&lt;sup&gt;n&lt;/sup&gt; 44.3&lt;sup&gt;h&lt;/sup&gt; 44.9&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>Innovation</td>
<td>49.6&lt;sup&gt;d&lt;/sup&gt;</td>
<td>58.9&lt;sup&gt;e&lt;/sup&gt;</td>
<td>60.7&lt;sup&gt;l&lt;/sup&gt;</td>
<td>54.7&lt;sup&gt;d&lt;/sup&gt; 61.6&lt;sup&gt;d&lt;/sup&gt; 62.6&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Physical comfort</td>
<td>62.1</td>
<td>65.4&lt;sup&gt;f&lt;/sup&gt;</td>
<td>59.3&lt;sup&gt;f&lt;/sup&gt;</td>
<td>64.6 65.1&lt;sup&gt;m&lt;/sup&gt; 60.8&lt;sup&gt;m&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

**Notes:**

N.B. Any shared plain superscript indicates $P \leq .05$ by t-test, one-tailed. Any shared superscript indicates $P \leq .10$.

* All t-tests utilize pooled-variance estimates, except for Work Pressure, I versus III, whose F-ratio (4.52, $P < .000$) requires a separate variance estimate.
### Table 17.5 Turnover Rates, 1984 and 1985

<table>
<thead>
<tr>
<th></th>
<th>1984 Corporate Rate</th>
<th>1984 HR Rate</th>
<th>1985 HR Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management</td>
<td>7%</td>
<td>0%</td>
<td>12%</td>
</tr>
<tr>
<td>Professionals</td>
<td>9%</td>
<td>30%</td>
<td>22%</td>
</tr>
<tr>
<td>Technicians</td>
<td>8%</td>
<td>67%</td>
<td>31%</td>
</tr>
<tr>
<td>Clerical</td>
<td>15%</td>
<td>66%</td>
<td>19%</td>
</tr>
<tr>
<td>Overall</td>
<td>13%</td>
<td>37%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Table 17.5 shows. This is almost three times the 1984 corporate average, and over 80 percent of the HR separations are voluntary. HR turnover in 1982 and 1983 approximated the corporate average.

### Fitting Interventions to Transitions

The OD design sought to fit and direct transitional dynamics, as two emphases demonstrate. Consider first the variable pace of intervention, and then the consistent underlying OD strategy.

#### Variable Pace

Expeditious as well as change goals influence the design. As Table 17.2 shows, data gathering at start-up consumed the first 100 days; active interventions spanned days 115 through 295, more or less; an anticipated reorganization urged a pause between days 295 and 575; and the design concluded with a first common HR experience under the aegis of the new structure. Informed observation continued over a period of years.

The reorganization confounds the intervention but also provides a stiff test of the persistence of its effects. Briefly, a traditional line/staff structure was quickly divisionalized (e.g., Chandler, 1962, Golembiewski, 1979, 2: pp. 3–69) around separate product lines to facilitate identification with smaller and more-focused efforts. Several “strategic operating areas” (SOAs) are identified and, after day 475, nearly half of the previously centralized HR staff report directly to one of the SOA heads. For many HR staff, this requires not only a new direct supervisor, but also a strong dotted-line relationship with the VP, Human Resources (VP, HR). In addition, both “movers” and “stayers” will maintain some common ties, as by periodic meetings of all HR resources for goal setting. The reorganization was a friendly one, with the VP, HR playing a major role in its design and implementation, but its significant effects had by no means played themselves out by day 575.
An environmental model dominates in management studies, and OD represents its most focused contemporary expression. Briefly, OD provides—via values, theory, and learning designs—a coherent approach to meeting individual needs and to reducing stressors while meeting organization needs (Golembiewski, 1982). This involves organization members in diagnosis and prescription. Conventional high-stimulus interventions based on OD theory and experience usually focus early on interaction—on the character and quality of relationships between people. Later attention may shift to task, as they did in this application, and to changes in policies and procedures or in structure. The basic intent is to induce an appropriate culture and process, and then to use that socio-emotional infrastructure to generate and support changes in policies, procedures, and structures.

**Front-End Interventions: A Calculated Risk**

The intervention strategy was conventional—high stimulus and confrontational and may seem risky for advanced burnout (see Figure 17.1). HR personnel maintained an active mode in all but two cases, however, and a high-stimulus design consequently had a reasonable promise of reducing HR burnout by addressing worksite issues. The high-stimulus features include having HR staff confront superiors and peers about a range of concerns, and this imposes energetic action planning on already-busy schedules.

The high-stimulus and confrontational character of the initial design essentially seeks to reverse tendencies toward degenerative interaction, and to develop or heighten movement toward regenerative interaction. In capsule, this involves increasing openness, owning, and trust, while decreasing risk (Golembiewski, 1979, vol. 1).

To illustrate, all HR staff were asked to list the three “best things” about their department, as well as three “concerns” they would most like to change. In 50 percent of the responses, the best things involve the professionalism, skill, and dedication of their coworkers; and an additional 30 percent emphasize the proactive nature of their work and its rewarding character. In effect, these attract-

---

![Figure 17.1](image)

**Figure 17.1** Front-end interventions. Main effects are indicated by solid lines; feedback or reinforcing linkages indicated with broken lines.
tions constitute a solid base from which to deal with concerns, in effect. Nearly half of the respondents focus on their “worst things,” the isolation and lack of cohesion within and between HR units, and 40 percent cite the pace of work—too many projects and too many hours of work. Sharing these data in a general session (day 115) involved developing examples and voicing reactions to them. In the process, employees and managers often had to confront the VP, HR. He was widely seen as demanding and unrelenting, and HR staff had to satisfy themselves that trust was sufficiently high, and risk sufficiently low, to warrant their expenditures of energy required to diagnose and remedy the unattractive aspects. As appropriate, the data also required confronting unit heads and their superior—the VP, HR.

After extended feedback and disclosure focusing on the HR state of affairs, which sought to model and even propose the ideal of regenerative interaction, four “interest groups” of volunteers were formed to gather additional information as well as to make recommendations for improvement. In effect, these recommendations (days 140–170) deal with the five special targets among WES scores highlighted in Table 17.4. Specifically, in the initial condition, Peer Cohesion, Supervisor Support, and Clarity rated too low, and scores on Control and Work Pressure were too high.

In sum, these front-end design features propose doing within HR what was more or less routinely done with and for clients. The goal involves building appropriate “processes”—a culture, values, and relationships—that will support changes in task features and foster regenerative interaction.

**Major Downstream Response: Policy Intervention**

Around day 200, the downstream search came to highlight the lack of HR promotion opportunities. A Career Progression Plan was developed and accepted by consensus, after being blessed by a corporate committee. The plan has these major objectives:

- To clarify career progression paths and opportunities
- To communicate information about such paths and opportunities
- To increase internal promotions, especially for the clerical and technical employees
- To hire individuals into jobs with a greater congruence between the person’s abilities and aspirations and the job’s realistic promotion opportunities

The Career Progression Plan tangibly tries to motivate and reinforce the improved HR infrastructure of processes and relationships, and also provides a clearer psychological contract with HR employees.
RESULTS OF ACTION PLANNING THROUGH DAY 425

Three effects signal an effective intervention in this case. To review the sense of earlier discussion, success can be inferred when:

The profile of phases of burnout will tend toward less-advanced phases. VIII → I;

WES scores will vary in three specific ways: Involvement, Autonomy, Task & Orientation, Innovation, and Physical Comfort will remain the same or increase; Control and, especially, Work Pressure will decrease; and Peer Cohesion, Supervisor Support, as well as Clarity will increase; Turnover rates will fall in 1985 compared to 1984, and remain lower.

In capsule, the initial emphasis on interaction relates primarily to two WES targets—Peer Cohesion and Supervisor—but increases in Involvement, Autonomy, and even Innovation might well occur as barriers to regenerative interaction are removed. Changes in policies and procedures more directly impact the three other WES targets—Work Pressure, Control, and Clarity—and all together provide both a motivation as well as rationale for regenerative interaction.

Administration III constitutes the first assessment of effects following major interventions. Since no formal interventions were made in the 130 days prior to administration IV, even maintenance of administration III levels implies a potent intervention. The intended effects seemed to occur and were maintained.

Profile of Burnout Phases

Table 17.3 indicates that the proportion of those in phases I–III approximately doubles for All Subjects, and nearly quadruples for Matched Subjects (not shown), comparing administrations I and II with III. Chi-square analysis ($3 \times 2$) shows that I versus III changes closely approach significance for All as well as Matched respondents ($P = .055$ and .06, respectively).

Note also that, even without explicit reinforcement, the shift is not only maintained for All Subjects but is extended for Matched Subjects. Chi-square analysis appropriately indicates only random differences in III versus IV profiles for both HR populations. In addition, the I versus IV differences not only fall in the expected direction but also achieve statistical significance for matched respondents and approach it for All ($P = .088$).

Changes in WES Scores

The shifts in WES scores also trend in expected ways, with All respondents providing the sharper pattern of differences. The data on Matched subjects are not reported here, to conserve space.
All Subjects

Considering administration III versus I in Table 17.4, nine of ten cases are in the expected direction. Moreover, four of the eight cases on which major increases are expected in fact do achieve the .05 level of statistical significance, and an additional case approaches that level (.05 > P < .10). Moreover, for the five target variables, four trend as expected. Three achieve significance, and the fourth approaches significance. Clarity trends lower.

For administration IV, all comparisons show random changes only. Physical comfort drops, although not significantly, reasonably due to the ongoing restructuring.

Turnover Record

Table 17.5 shows that turnover falls about 45 percent in HR, in 1985 compared to 1984. No change occurs in the corporate rates, in addition, which closely approximate 13 percent in both years.

No substantial reasons encourage attributing the sharp drop in HR turnover to factors other than the intervention. If anything, 1985 was a better time than the previous year for HR personnel to seek new jobs. Moreover, the years following 1985 suggest that the change had substantial staying power. In sum:

<table>
<thead>
<tr>
<th></th>
<th>1986</th>
<th>1987</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate turnover</td>
<td>15.0%</td>
<td>15.7%</td>
</tr>
<tr>
<td>HR turnover</td>
<td>13.0%</td>
<td>14.3%</td>
</tr>
</tbody>
</table>

RESULTS OF ACTION PLANNING THROUGH DAY 575

Only very robust effects would survive in the face of the reorganization as well as the nine-month interval between administration V of the survey and the last planned intervention: Three emphases provide summary perspective on what seems to be a substantial persistence of changes. Additionally, although formal research did not extend beyond 1990, longer-term observation also supports the sense of an organization fueled by positive affect. For example, the host organizations had a high and growing rank in its market sectors beyond 2000.

Changes in Phases of Burnout

In Table 17.3, Administration V shows two patterns of decay in the extreme phases of burnout. For All respondents, the decay is modest and not significant by chi-square test. For Matched Subjects who were with HR during the full obser-
vational period, the decay is substantial after administration IV. In both cases, 25 percent fewer individuals are classified in phases VI–VIII, comparing administration V to I. Moreover, the employees in phases I–III increase by 104 and 65 percent for All and Matched respondents, respectively.

**Changes in WES Scores**

Major and expected changes on the WES variables remain through administration V, and even increase, especially for All Respondents. A detailed review of administration V appears elsewhere (Golembiewski, Hilles, and Kim, 1986), but a sense of the data trends can be suggested economically. For all variables, I versus IV and I versus V, thirty-six of forty comparisons show differences that fall in the expected direction. Moreover, 50 percent of these cases are statistically significant as well as in the expected direction, and an additional 17 percent of the cases closely approach statistical significance ($P < .10$). In sum, the very long post-test basically retains the post-intervention impact.

**Trends in Turnover**

The trend for HR turnover in the first four months of 1986 continues downward from the very high levels of 1984. In sum, the rates are: 37 percent in 1984, 20.3 percent in 1985, and 16.8 percent in 1986, annualized. In all three years, corporate turnover closely approximates 13 percent per annum.

**SEVEN FINDINGS FOR GUIDING ACTION**

This theory-based effort to ameliorate burnout had substantial and persisting effects, in sum, and five points provide highlights. First, overall, the results imply the usefulness of high-stimulus OD designs in ameliorating burnout, active mode, as well as in affecting the character of the worksite. This constitutes new news and confirmation of old news. Again, the relevance of regenerative interaction is supported, as in the ability of standard OD designs and approaches to moving away from the degenerative. Consistently, also, this chapter reinforces the salience and attractiveness of OD values, as they are reflected in various ways throughout this volume. For example, WES variables seem tied quite closely to OD values: see Peer Cohesion, Involvement, and so on.

The finding concerning burnout and its association with group properties is unique and constitutes very good news for conventional OD designs. This complements the strong case, in both theory and research, that can be made for the association of OD designs and low burnout.

Second, the results recommend distinguishing modes of adaptation to advanced burnout and, on the present showing, the combination of high burnout
and active mode constitutes a reasonable target for conventional OD interventions. Passive cases of advanced burnout constitute a challenge for future work.

But consider three points of theoretical and practical significance. Thus, no one knows the overall proportion of actives and passives in the universe of those in advanced phases of burnout, but it seems 25 percent or fewer are actives (Golembiewski and Munzenrider, 1984), overall. However, countries like Japan seem to have higher proportions in active status in all burnout phases, even though the decreasing incidence of active phases still exists (Golembiewski, Boudreau, Munzenrider, and Luo, 1996, esp. pp. 215–218).

In addition, this study does not tell us whether active status is a prelude to the passive mode. If so, not only would this signal the criticality of timely identification, but it provides a conceptual explanation of what seems an anomaly—the existence of two passive employees in an active HR culture. If an active mode derives from personality features—for example, from high self-esteem or a low propensity to helplessness—promptness is less critical.

Finally, we know a bit about the relative significance of the major conceptual actors in such planned change, but we need to know much more. For example, do the phases play the central role in explaining effects, or do they constitute more of a reflection of the basic distinction between active and passive modes? In addition, acute observers might propose that the interaction-centered components of the learning design sketched above encourage the exchange of “social support,” which can be defined briefly as the significant sense that HR members are valued and sustained by their colleagues. In general, early explorations of such critical assignments of theoretical and practical potency reinforce the usefulness of the conceptual network sketched in this and other chapters.

Third, experimental mortality was high, but personnel changes do not explain the changes in phases. Most important, the Matched respondents improve their phases distribution substantially more than All respondents, through administration IV. Moreover, of those leaving, about 48 percent are in phases VI–VIII and 28 percent are in phases I–III, on the last recorded observation. These are roughly the same as the weighted proportions of phases in the Total population at the approximate times of departure—43.2 and 30.6 percent, respectively.

Fourth, the research design probably generates a conservative estimate of effects. Deliberately, administration III was held before a ten-hour design concluding the active interventions in the action research. Moreover, administration IV followed that intervention by four months. So administration III picked up only expectation effects about day 295’s major integrating experience; and fade-out had ample time to occur before administration IV. Administration V came long after the prior planned intervention, and before a two-hour general session.

Fifth, theoretical reasons suggest that the two passives are not likely to respond positively to high-stimulus designs, but they do not seem to have been
adversely affected. Those two persons improved their WES scores substantially and had these sets of five phase assignments:

\[
\begin{array}{cccccc}
\text{WES Outlier 1:} & \text{VII} & \text{VIII} & \text{VIII} & \text{VIII} & \text{IV} \\
\text{WES Outlier 2:} & \text{VI} & \text{VII} & \text{VI} & \text{III} & \text{VI}
\end{array}
\]

The Employee Assistance Program officer was alerted concerning possible difficulties, but none surfaced.

GREATER SUCCESS BY SPECIFYING DIFFERENCES BETWEEN CONTEXTS, REVISITED

Despite some major open ends and blank spots, then, this analysis and the prior chapters should establish the attractiveness of seeking to improve success rates of OD applications, this time by taking differential burnout into account, and by taking group differences into account. This chapter outlines a major way to do something about acting on this knowledge of burnout’s strategic character—specifically, by influencing the group context.

Consider this sketch of the very real potential for improving the success rates of OD applications. Attention goes to one direct and two reinforcing contributors to this conclusion.

Directly, the less-advanced phases of burnout are associated with group properties congenial to OD, and especially to the regenerative model of interaction. Put the other way, the most-advanced phases of burnout are uncongenial to OD processes.

Theoretical support comes from regenerative interaction, which involves low levels of risk as well as high levels of openness, owning, and trust. For example, high Clarity—that is, the extent to which employees know what is expected of them—suggests high openness and owning, or at least an unwillingness by management to consider human responses to having the heat kept on, or to generate alternative ways of mobilizing effort—as by higher Involvement or Task Orientation, which imply a willingness to risk and trust, as well as high openness and owning. From another perspective, high levels of Peer Cohesion, Supervisor Support, and Autonomy all suggest substantial degrees of trust as well as a moderated degree of personal risk. Individuals are encouraged to make their own decisions, but in the context of peer and supervisory support. Individuals are not left hanging, to twist in the wind, whenever their reach exceeds their grasp.

Similarly, high Involvement is a prime feature of OD efforts and group processes—and so are Peer Cohesion and Supervisory Support. Recall also that
OD can be defined as increasing responsible freedom in organizations. Translated quite literally, greater freedom implies Autonomy as well as a moderated Control—or the use of rules and pressures to keep employees under control—that are ideally reinforced by diminished Work Pressure. Relatedly, responsible freedom requires high degrees of Involvement, Task Orientation, and Peer Cohesion.

Broadly, burnout seems fundamentally rooted in specific group contexts. Illustratively, if you know the burnout phase of any single member of an immediate work unit, you have a very high probability of knowing the dominant pattern of phases (Golembiewski, Boudreau, Munzenrider, and Luo, 1996, pp. 215–217).

REFERENCES


Acknowledging Some Limitations of “One Person, One Vote”*
Survey/Feedback Realities and Classes of Respondents

OD has some definite populist stripes, and practitioners can wear them proudly, and often do so. That is, OD interveners often give voice to the underdogs in helping them speak truth to the top dogs—that is, to those in power. And this is a critical activity and clearly one that is central to the movement toward regenerative interaction and away from degenerative interaction. Early chapters provide detailed rationales for why this movement is strategic in both theory and experience, and how this movement is indispensable in inducing greater responsible freedom in organizations.

At the same time, however, OD can easily have too much of even the best of things. At least, that is what the present position argues directly. For some purposes, equating “one person, one vote” is not only disingenuous but possibly misleading, if not actually dangerous.

Let not the reader conclude that this chapter intends more than the demonstration that converting data into information often will require differentiating classes of individuals. At earlier times this could only be done by elites with computers and especially skills in programming. ODers often saw this as consequential. That is, reliance on those elites could threaten the commitment and involvement of survey respondents, whom ODers saw as both the owners and interpreters of the responses they supplied.

* The substance of this chapter appears in Ironies 1.
Nowadays, this critical view has lost most of its power, even for the most populist ODers. Computing machines are now quite inexpensive and widely available and, in addition, now many canned programs are available for reshuffling data in multiple ways.

**A CONCEPTUAL CONTEXT FOR IRONY VII**

Put the essential point in bold terms. OD has substantial success rates—and most ODers do not regularly seek to classify differences between people; and some ODers have made a fetish out of not distinguishing individuals or small groups of them in the service of confidentiality. Since individuals often differ in important regards, it seems almost necessary that some portion of OD’s failures derive from applications to a panel of respondents differentiated enough to make one design a good fit for some people, but a poorer fit for others. Ergo, relevant differentiations of people will increase the probability of success rates beyond usual levels, if ODers’ attitudes change.

Other details add useful perspective to this basic view, as illuminated by a broadly historical summary. The survey/feedback design may not yet be the general choice in efforts to develop large systems that are more efficient and effective, more productive, and more human. But it is close. Moreover, the design has had substantial success rates. And a spate of how-to-books a decade or two ago clearly signaled the arrival of survey/feedback, and the end is nowhere in sight (i.e., Bowers and Franklin, 1977; Golembiewski and Hilles, 1979; Nadler, 1977).

This burgeoning status disguises but does not eliminate Irony VIII. In fact, the success of survey/feedback designs has been achieved without specifying important differences between people. Generally, little attention has been directed to the conditions under which specific versions of that design are indicated or contraindicated. Probably then, an attractive margin for improvement is close at hand.

This burgeoning reliance on survey/feedback in OD (French and Bell, 1973, pp. 25–29) has not generated a corresponding concern with when and how that design is most useful. Bowers’s (1973a) mammoth comparison of several standard OD designs in many organizations exemplifies the state of the literature. His results suggest the special potency of survey/feedback interventions, but caution seems appropriate. Bowers’s research has generated critical comment (Torbett, 1973; Pasmore, 1976), and serious methodological questions can be raised as to whether the results have any currently interpretable meaning (Golembiewski, Billingsley, and Yeager, 1976); Bowers himself (1973a, 41–42) guesses that important intervening variables determine when (for example) laboratory training is more appropriate than survey/feedback, despite the huge advantage his data accord the latter design.
The common assumption that a respondent is a respondent constitutes perhaps the greatest point of unspecificity in survey/feedback designs. To be sure, survey respondents often will be differentiated in terms of various demographics—age, race, sex, organization unit, and so on. These categoric variables are often not very revealing, however. Whether the issue is assessing an organization's climate or evaluating employee acceptance of some practice or policy, far more rather than less, a survey respondent has been considered a survey respondent.

Ideology reinforces the analytic convenience of the assumption that respondent A is equal to respondent B, and this easy pairing helps explain why the assumption has been so little challenged as well as why it will be difficult to supplant. Like the U.S. Supreme Court's prevailing political philosophy—one person, one vote—the OD version rests on democratic, even populist, values. Albeit little tested, the one person, one vote assumption seems to rest on shaky empirical foundations. Admittedly sparse empirical evidence as well as common sense both imply that such unspecificity regarding survey/feedback interventions cannot be tolerated over the long run. For example, one study demonstrates that people who differ in basic personality predispositions will report seeing different aspects of reality (Golembiewski and Munzenrider, 1974). Survey results, in short, tell us something about the respondent as well as about the stimulus reality. Interpretation of survey results consequently will be limited in the absence of knowledge about differences between the specific persons who provide the data.

This chapter tests this conclusion by assessing differences across a broad range of indicators between survey respondents who differ in their performance appraisals. Despite their problems (McGregor, 1960, esp. pp. 77–89), performance appraisals do stand as a major tangible expression of the organizational value placed on specific individuals. This significant fact motivates the test here of differences in appraisals—of how they may influence interpretations of responses to surveys, and how the interpretations of results may vary with differences between people.

This effort implies numerous questions—empirical and normative, ethical and practical. Empirically, this chapter tests whether employees with high versus low appraisals differ significantly on a wide range of attitudinal scales and items. Where such empirical differences exist, important ethical issues with serious practical implications must be addressed. For example, what should be done if survey research reveals a major problem, but “major” only to those with low appraisals? Most survey/feedback designs finesse such questions because the designs do not distinguish between those high versus low distinctions on performance appraisals.

And what of the temptation (if not the danger) implicit in this chapter? Normative and ethical considerations help explain the common acceptance of the
one person, one vote assumption in survey/feedback designs. These considerations urge caution in whether and how to differentiate survey respondents. In short, such knowledge can hurt the careless or the uninformed, or be used to hurt them. At an extreme, knowledge that some aggregate have anti-management attitudes might be used to finger them for firing or other pretexts.

**RESEARCH DESIGN**

This study focuses on differences in the performance-appraisal ratings of 1,474 respondents to a broad QWL survey, whose date come from all hierarchical levels and major functions in a soft-goods firm. In all cases, the appraisal closest in time to the date of the QWL survey was coded for each individual. Matches of specific appraisals to QWL respondents proceeded in absolute isolation from company officials. To scotch even the suspicion that the linking of survey and appraisal data might be used punitively, in addition, appraisal data were not made available to the researcher until eighteen months after the QWL survey, long after the action research based on the data was completed.

Officials of the host organization much preferred rigidly safeguarded anonymity for individual respondents. They agreed only with great reluctance to a complicated provision whereby *willing* respondents could identify themselves to the consultant/researcher while their identity was withheld from the employing organization, within which the coding and keypunching were done. This convention permitted basic research without contaminating the survey.

**Performance Appraisal Classes**

The firm used five appraisal rankings, but these are collapsed for present purposes into three categories. See Table 18.1 for details. Basically, a Good appraisal or worse in this firm seriously affects employees, as in the length of the period they must wait for a normal salary review as well as in the maximum percentage salary increase available to them. The details would be burdensome, but they

<table>
<thead>
<tr>
<th>Table 18.1 Data about Appraisals in Study Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm's appraisal category</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>Exceptional</td>
</tr>
<tr>
<td>Excellent</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Acceptable</td>
</tr>
<tr>
<td>Unacceptable</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
massively reinforce the decision here to classify a Good appraisal or worse as Sub-Modal. Exceptional and Excellent appraisal ratings were labeled as Superior and Model, respectively.

The distribution of appraisals in Table 18.1 closely corresponds to organizational practice in recent years. Approximately 80 percent of all ratings awarded have been Exceptional or Excellent, with about 70 percent in the latter category. This compares with 77.95 and 71 percent in the present survey, respectively, in the present subpopulation of 1,474. More specific comparisons—such as those distinguishing major operating centers—also support this conclusion.

**Four Kinds of Comparisons**

The three appraisal categories are used to distinguish subgroups of respondents on four clusters of measures, and represent a loss of variance in categories that is costly but unavoidable here. These clusters include fifty-three variables, distributed in the following ways:

- Three bottom-line measures of employee perceptions of the firm
- Seventeen multi-item scales dealing with a range of managerially relevant variables
- Twelve single items referring to various management practices
- Single-item evaluations of twenty-one benefits, programs, and practices in the firm

Each cluster of measures is treated similarly. F-ratios test for the significance of overall variance between the three appraisal categories. Where significant overall variance exists, statistically significant differences between all possible pairs of appraisal categories are isolated using the modified Least Significant Differences (LSD) procedure. This procedure has the virtue of providing an exact value when sizes of cells are unequal (Nie et al., 1975, pp. 427–28), the condition that obviously exists here. The overall F-ratios might detect nonrandom differences, of course, while the LSD test fails to isolate any statistically significant pairs of differences. Hence, the LSD procedure provides the definitive test of the magnitude and regularity of covariation between differences in performing appraisal and the fifty-three target variables.

**Differences on Three Bottom-Line Measures**

The three items shown in Table 18.2 are considered to tap important employee perceptions of their firm and imply that differences in performance appraisals can enrich analysis. Table 18.2 shows that the average employee in each appraisal subgroup similarly perceives the firm as “among the best” places to work ($\bar{X}$ approximates 2.0 on a 7-point scale); and employees in each appraisal subgroup
### Table 18.2 Differences in Appraisals and Three Bottom-Line Measures of Employee Attitudes toward Firm

<table>
<thead>
<tr>
<th>Item means, classified by appraisal category</th>
<th>Overall F-Probability</th>
<th>Test for significance of differences between pairs of means (where √ designates a difference achieving .05 level by modified LSD procedure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior</td>
<td>1 vs. 2</td>
<td>1 vs. 3</td>
</tr>
<tr>
<td>How firm rates as a place to work</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Firm management better now than three years ago?</td>
<td>2.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Years employee expects to work for firm</td>
<td>1.5</td>
<td>1.6</td>
</tr>
</tbody>
</table>
generally expect long tenure in the firm ($X = 1.5$ on a 5-point scale, where $1 = \text{“until retirement”}$ and $2 = \text{“more than 5 years”}$). Significant differences do show up on the item providing an overall evaluation of the quality of management now versus three years ago, at which time a significant management change occurred and an OD effort was begun. Sub-Modals provide the lowest evaluation (3.1), but even that average score implies that management is “about the same.” The mean (2.4) for employees with Superior appraisals, in comparison, approaches the next verbal anchor for respondents, which is that management got “better” over the three-year interval.

Single-item measures imply interpretive problems but, with that caution in mind, does specifying differences in performance appraisals enrich the analysis of the three bottom-line measures? The answer seems affirmative. Without that specification, the data in Table 18.2 imply only that the firm is a good place to work, pretty much as all employees see it. This may be comforting news, but it seems misleading in at least one major particular: when appraisal differences are specified, the responses of Sub-Modals suggest that even their poor appraisal experiences have not proved alienating, at least at a gross level.

This may be welcome news for some managements, but this good news suggests one major potential point of unreality. Even employees with the poorest performance appraisals expect to remain with the firm for long periods, in almost all cases until retirement. This suggests a number of potentially serious issues, centering around the probability that any tightening of performance appraisals in the future will be seen as violating widely (but perhaps not wisely) held expectations. As the business grows more complicated, profit margins shrink and competition increases; as a result, satisfactory performance today might be seriously inadequate tomorrow.

The storm clouds were clearly on the horizon at the time of the survey, and major personnel actions in the firm were taken somewhat later. Generally, the firm moved its basic emphasis from loyalty toward competence. Sub-Modals may have been poorly prepared for this outcome, the enriched survey data suggest, despite having received some clues that boded no particular good for their careers in the firm.

In sum, specifying differences in appraisals raises a warning not apparent in the overall results. In fact, the overall results encourage a too-sanguine interpretation of the responses on this first cluster of three items. The warning signal takes on added significance because of management’s special interest in the three bottom-line measures in Table 18.2.

**Differences on Seventeen Managerially Relevant Scales**

Differences in appraisals become even more salient in analyses of the seventeen managerially relevant scales shown in Table 18.3. That is, despite suggestions...
### Table 18.3 Differences in Appraisals and Seventeen Managerially Relevant Scales

<table>
<thead>
<tr>
<th>Alpha Coefficient</th>
<th>Superior</th>
<th>Modal</th>
<th>Sub-Modal</th>
<th>F-Probability</th>
<th>1 vs. 2</th>
<th>1 vs. 3</th>
<th>2 vs. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>JDI-work</td>
<td>.8056</td>
<td>38.5</td>
<td>36.7</td>
<td>34.0</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JDI-supervision</td>
<td>.8705</td>
<td>45.8</td>
<td>42.8</td>
<td>40.7</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JDI-coworkers</td>
<td>.8889</td>
<td>42.5</td>
<td>41.8</td>
<td>40.2</td>
<td>.046</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JDI-promotions</td>
<td>.9000</td>
<td>13.2</td>
<td>12.4</td>
<td>12.4</td>
<td>.488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JDI-pay</td>
<td>.7786</td>
<td>19.5</td>
<td>18.1</td>
<td>16.8</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>JDI-total</td>
<td>.9239</td>
<td>190.3</td>
<td>180.3</td>
<td>170.6</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Efficacy of upward communication</td>
<td>.5723</td>
<td>12.8</td>
<td>13.3</td>
<td>13.7</td>
<td>.269</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upward mobility</td>
<td>.8660</td>
<td>9.3</td>
<td>10.3</td>
<td>9.5</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Supervisory influence</td>
<td>.6866</td>
<td>6.6</td>
<td>7.2</td>
<td>7.7</td>
<td>.038</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Trust I (R)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.8551</td>
<td>4.6</td>
<td>6.2</td>
<td>7.4</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Trust II (R)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.7459</td>
<td>6.7</td>
<td>7.7</td>
<td>8.2</td>
<td>.001</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Participation in decision making</td>
<td>.7932</td>
<td>14.7</td>
<td>13.4</td>
<td>11.7</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Job involvement</td>
<td>.8237</td>
<td>26.7</td>
<td>25.9</td>
<td>24.2</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Motivation</td>
<td>.5897</td>
<td>18.8</td>
<td>18.2</td>
<td>17.7</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Identification with firm</td>
<td>.5985</td>
<td>6.7</td>
<td>6.5</td>
<td>6.5</td>
<td>.351</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingness to disagree with supervisor</td>
<td>.6685</td>
<td>12.2</td>
<td>11.9</td>
<td>10.2</td>
<td>&lt;.000</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Job-related tension</td>
<td>.8153</td>
<td>16.5</td>
<td>18.3</td>
<td>17.9</td>
<td>.044</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: <sup>a</sup>R designates a reversed score; i.e., a high Trust I or Trust II score indicates low trust.
in Table 18.2 that Sub-Modals have not been grossly alienated by their appraisal experience, impressive evidence suggests that they clearly differ in major particulars from those with Superior and Modal appraisals. Specifically, this study focused on seventeen scales:

The eighteen-item scale Job Descriptive Index (JDI)-Work (Smith, Kendall, and Hulin, 1969)
The eighteen-item scale JDI-Supervision (Smith, Kendall, and Hulin, 1969)
The eighteen-item scale JDI-Coworkers (Smith, Kendall, and Hulin, 1969)
The nine-item scale JDI-Promotions (Smith, Kendall, and Hulin, 1969)
The nine-item scale JDI-Pay (Smith, Kendall, and Hulin, 1969)
The seventy-two-item JDI-Total (Smith, Kendall, and Hulin, 1969)
A set of fourteen items Efficacy of Upward Communication (Roberts and O’Reilly, 1974)
A two-item measure of Upward Mobility (Roberts and O’Reilly, 1974)
A measure of Supervisory Influence (Roberts and O’Reilly, 1974)
A three-item measure Trust and Confidence in Immediate Supervisor (Trust I on Table 18.3) (Roberts and O’Reilly, 1974)
A five-item measure of Composite Trust and Confidence in All Levels of Supervision (Trust II on Table 18.3)
A five-item scale tapping Participation in Decision Making (White and Ruh, 1973)
A nine-item scale measuring Job Involvement (White and Ruh, 1973)
A two-item scale measuring Identification with Firm (White and Ruh, 1973)
A four-item measure of Willingness to Disagree with Supervisor (Patchen, 1965, pp. 48–51)

Patterns in this panel of measures will be consequential, for the measures seem both sound and significant. Confidence in the instruments seems well placed, to begin. Alpha coefficients for each scale are given in Table 18.3. Except in three cases, those coefficients reach levels that permit optimism that the scales deal with significant dimensions of reality. Moreover, sometimes-voluminous research testifies to the validity and reliability of most of the scales. Hence, any consistent pattern of results on these seventeen scales must be taken seriously.

Several technical notes usefully precede discussion of the data. The scales are so coded that, in all but two cases, high scores imply more of the attitude/behavior in question than low scores. The two Trust scales have reversed scores. Note that the direction of questionnaire items was often varied to inhibit response set.
The pattern of variation between the seventeen scales and appraisal differences is almost monolithic. Table 18.3 reveals statistically significant overall variation between appraisal categories on fourteen of the seventeen scales, with random differences existing only for JDI-Promotions, Upward Communication, and Identification with Firm. Promotion was a major problem throughout the firm, due to its growth history, which may explain why those in all performance appraisal categories see it in approximately the same light. The other two deviant scales have low alpha coefficients, so Table 18.3 may reflect only technical problems in those cases.

The differences are regular as well as robust, moreover, and send a clear and constant message. Those with Superior appraisals have the most positive profile, with the Sub-Modals reflecting the least positive profile. Specifically, nonrandom differences exist between twenty-eight of the forty-two possible pairs, or nearly 67 percent of the cases. This suggests formidable differences.

Table 18.3 poses no interpretive puzzle, then. Sub-Modals see their environment as being significantly less attractive. For example, those with Sub-Modal appraisals report less satisfaction on five of the six JDI dimensions; they are lowest on Trust, Motivation, Involvement, Participation, and so on. Sub-Modals also rate their immediate supervisors as highest on Influence, while being least inclined to Disagree with Supervisor. These suggest a reasonable reaction to the low appraisals of the Sub-Modals, in which their immediate supervisors played the most prominent role.

Table 18.3 is mute on some important questions, patently. That table obviously does not reveal whether the low appraisals leads to the negative attitudes, for example, or whether the low appraisal is more of a dependent variable. Perhaps the relationship is interactive, but causality definitely cannot be inferred.

No doubt exists, however, about whether the relationship is marked and regular. Patently, it makes a major difference which category of employees provide self-reports on the present seventeen scales. Differences in performance appraisals co-vary significantly and regularly, on definite balance, with how employees see their worksite.

Knowledge of this pattern of variance could be important, even critical, in influencing possible managerial responses to the survey data. For example, efforts to do something about promotion procedures and policies could be motivated by the fact that they pose problems for all employees, regardless of their appraisal, judging from JDI-Promotions as well as from much confirming survey evidence not reviewed here.

Conversely, a proposal to “increase participation” would get an even longer look, and especially so if quantitative as well as qualitative enhancements were desired. Those with Superior and Modal appraisals already score high on participation, that is, relative to both Sub-Modals as well as with respect to the highest possible score on the scale. Increased participation by Sub-Modals, moreover,
Limitations of “One Person, One Vote”

might be constrained by their lower Job Involvement, Motivation, and Willingness to Disagree with Supervisor, as well as by higher Job-Related Tension. Efforts to increase participation, that is to say, might be reasonable only after some success in increasing (for example) one’s willingness to disagree with supervisors. There may even be a darker reality. Prior improvements in appraisal ratings might be required.

Differences on Twelve Management Practices

Differences in performance appraisals also seem to co-vary significantly with the twelve single-item measures of various management practices shown in Table 18.4. Except for the first variable, for which the table provides the response stems, all other items utilize a 5-point scale: (1) Strongly Agree, (2) Agree, (3) Not Sure, (4) Disagree, and (5) Strongly Disagree.

Table 18.4 reflects a dominant pattern with significant covariation existing on ten of the twelve items. Moreover, for the thirty pairs of comparisons between appraisal categories involving the ten items whose overall variation is almost certainly not random, eighteen pairs (or 60 percent) achieve statistical significance.

To be sure, the reliance on single-item measures does not permit certainty about the multi-dimensionality of the phenomena tapped by Table 18.4. Whether those single items measure twelve dimensions, or one dimension twelve times, however, appraisal differences clearly must be accorded major potency. Note that the discussion of Table 18.5 employs factor analysis to illustrate how the dimensionality issue can be dealt with in the present data. A similar analysis was performed on items in Table 18.4, but is not reprinted here to conserve space and because Table 18.5 presents more interesting problems.

Specifying appraisal categories for such management practices as those in Table 18.4 can serve at least three purposes in tailoring responses to the data from a survey/feedback design. First, whatever their performance appraisal, employees see things similarly on two items, at least at first impressions. Crisis Atmosphere constitutes a high-leverage target on which ameliorative action would be widely welcome, and High Standards taps a practice pretty much seen as already in good shape.

Second, employees with different appraisals see some items in significantly different ways, and these differences merit managerial attention about whether, and when, to take corrective action. Consider the greater (and reasonable) enthusiasm of Sub-Modals for a program by which subordinates could evaluate supervisors.

What should management do? The data help in moving toward an answer. Subordinate evaluations do not seem to have a high priority. Those with Superior appraisals approach an average response of Not Sure; and not even the Sub-
<table>
<thead>
<tr>
<th>Test for significance of differences between pairs of means (where ( / ) designates a difference achieving .05 level by modified LSD procedure)</th>
<th>1 vs. 2</th>
<th>1 vs. 3</th>
<th>2 vs. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant supervisor mode:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Primarily reward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Mostly reward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Reward and punishment</td>
<td>1.7</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>4. Mostly punishment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Primarily punishment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managers insist on high standards</td>
<td>1.7</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Management creates crisis atmosphere</td>
<td>2.9</td>
<td>2.8</td>
<td>2.9</td>
</tr>
<tr>
<td>Last merit pay increase left employee dissatisfied</td>
<td>3.9</td>
<td>3.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Satisfaction with information available</td>
<td>2.2</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Generally understand reasons for management policies</td>
<td>2.3</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Management criticizes more quickly than praises</td>
<td>3.5</td>
<td>2.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Employees treated like children</td>
<td>3.6</td>
<td>3.2</td>
<td>2.9</td>
</tr>
<tr>
<td>Work involves learning new things</td>
<td>2.3</td>
<td>2.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Employee feels overqualified for present job</td>
<td>3.9</td>
<td>3.8</td>
<td>3.5</td>
</tr>
<tr>
<td>Employee should evaluate supervisors</td>
<td>2.7</td>
<td>2.4</td>
<td>2.2</td>
</tr>
</tbody>
</table>
### Table 18.5: Differences in Appraisals and Twenty-One Benefits, Programs, and Practices, Classified by Appraisal Category Item Means

<table>
<thead>
<tr>
<th>Category Item</th>
<th>1 Superior</th>
<th>2 Modal</th>
<th>3 Sub-Modal</th>
<th>Overall F-Probability</th>
<th>Test for significance of differences between pairs of means (where √ designates a difference achieving .05 level by modified LSD procedure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retirement plan</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>.853</td>
<td>√</td>
</tr>
<tr>
<td>Savings plan</td>
<td>2.7</td>
<td>2.5</td>
<td>2.4</td>
<td>.018</td>
<td>√</td>
</tr>
<tr>
<td>Vacation policy</td>
<td>1.7</td>
<td>1.8</td>
<td>1.9</td>
<td>.112</td>
<td>√ √ √</td>
</tr>
<tr>
<td>Sick leave policy</td>
<td>1.7</td>
<td>1.8</td>
<td>2.0</td>
<td>&lt;.000</td>
<td>√ √ √</td>
</tr>
<tr>
<td>Dispensary and first aid</td>
<td>2.4</td>
<td>1.3</td>
<td>2.3</td>
<td>.578</td>
<td>√</td>
</tr>
<tr>
<td>Medical insurance</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td>.969</td>
<td>√</td>
</tr>
<tr>
<td>Parking lots</td>
<td>2.5</td>
<td>2.8</td>
<td>2.9</td>
<td>.018</td>
<td>√</td>
</tr>
<tr>
<td>Absence policy</td>
<td>1.9</td>
<td>2.1</td>
<td>2.5</td>
<td>&lt;.001</td>
<td>√ √ √</td>
</tr>
<tr>
<td>Career development</td>
<td>2.6</td>
<td>2.8</td>
<td>2.6</td>
<td>.007</td>
<td>√ √ √</td>
</tr>
<tr>
<td>Employee promotions</td>
<td>2.6</td>
<td>3.0</td>
<td>3.0</td>
<td>.007</td>
<td>√ √ √</td>
</tr>
<tr>
<td>Tuition payments</td>
<td>1.9</td>
<td>2.0</td>
<td>2.0</td>
<td>.361</td>
<td>√</td>
</tr>
<tr>
<td>Local phone policy</td>
<td>3.2</td>
<td>3.2</td>
<td>3.1</td>
<td>.853</td>
<td>√</td>
</tr>
<tr>
<td>Affirmative action programs</td>
<td>2.3</td>
<td>2.4</td>
<td>2.6</td>
<td>.119</td>
<td>√</td>
</tr>
<tr>
<td>Severance pay policy</td>
<td>2.5</td>
<td>2.4</td>
<td>2.5</td>
<td>.386</td>
<td>√</td>
</tr>
<tr>
<td>Life insurance program</td>
<td>1.9</td>
<td>1.9</td>
<td>2.0</td>
<td>.146</td>
<td>√</td>
</tr>
<tr>
<td>Disability insurance program</td>
<td>2.1</td>
<td>2.2</td>
<td>2.1</td>
<td>.801</td>
<td>√</td>
</tr>
<tr>
<td>Security practices</td>
<td>2.5</td>
<td>2.4</td>
<td>2.3</td>
<td>.087</td>
<td>√</td>
</tr>
<tr>
<td>Maternity leave policy</td>
<td>2.0</td>
<td>2.1</td>
<td>2.1</td>
<td>.875</td>
<td>√</td>
</tr>
<tr>
<td>Lounges</td>
<td>3.0</td>
<td>2.9</td>
<td>2.6</td>
<td>.003</td>
<td>√</td>
</tr>
<tr>
<td>Elevators</td>
<td>2.7</td>
<td>2.6</td>
<td>2.4</td>
<td>.017</td>
<td>√</td>
</tr>
<tr>
<td>Restrooms</td>
<td>2.5</td>
<td>2.5</td>
<td>2.3</td>
<td>.022</td>
<td>√</td>
</tr>
</tbody>
</table>
Modals average response is Agree. Moreover, some observers might be concerned that some Sub-Modals only reflect a desire to get even.

Third, Table 18.4 also contains some surprises. For example, it implies that Sub-Modals seem satisfied with their last salary increase. Given the appropriate policies put them at a significant advantage in regard to the timing and amount of such increases, the mean response implies a substantial satisfaction with processes about which Sub-Modals could reasonably have negative feelings. From one perspective, the data thus provide some welcome evidence about the perceived fairness if not beneficence of management practices and policies with regard to performance appraisal. Specifying appraisal differences, patently, highlights the perhaps surprising generality of this perceived fairness and beneficence.

From another perspective, more somber interpretations of the data should be considered. For example, management might wish to consider whether the salary increments and the appraisal differences to which they are attached convey a consistent message to employees. And management might also consider whether the received message is the one intended.

**Differences on Twenty-One Benefits, Programs, and Practices**

Survey/feedback designs often seek to elicit data about how a range of benefits, programs, and practices are perceived by organizational members. Such data can motivate ameliorative action, or can confirm suspicions that well enough should be left alone. Table 18.5 summarizes the results of one such overview of twenty-one benefits, programs, and practices, each of which respondents scored on a 5-point scale including these responses: (1) Excellent, (2) Good, (3) Fair, (4) Not Very Good, and (5) Poor.

Overall, Table 18.5 suggests that employees see the firm in question quite positively. Beyond that, specifying performance appraisals enlightens analysis in several ways.

Paradoxically, to begin, performance appraisals do not seem to co-vary significantly with differences on specific items, and this provides useful information that would otherwise be unavailable. Specifically, differences in appraisals do not seem relevant in twelve of the twenty-one cases. Some of these cases seem to be high-leverage opportunities for possible ameliorative action—for example, Local Phone Policy. Moreover, other items—for example, Vacation Policy—suggest that no urgent need for change is seen by any employees.

Nonetheless, differences in performance appraisals seem associated with nonrandom variation in evaluations of nine of the twenty-one benefits, programs, and practices. On these nine items, thirteen of the twenty-seven possible paired-comparisons of appraisal categories show noteworthy differences.
None of these statistically significant items can be interpreted definitely, but some suggest major grist for the mill. Consider a few interpretive puzzles. For example, Sub-Modals score lowest on Lounges, Elevators, and Parking Lots. Does that merely reflect a general malaise—an emphasis on hygiene factors in the absence of motivators like recognition and achievement? Or what? At least for parking lots, speculation about the more negative attitudes of Sub-Modals seems safe enough. Company parking spaces are scarce in the host firm and are awarded on two grounds: hierarchical status and tenure. Patently, employees with Sub-Modal appraisals are less likely to gain either.

Intriguing patterns also can be teased out of other items on which differences in appraisals seem important. Note that Sub-Modals are least positive about Sick Leave Policy and Absence Policy. Does this mean that some supervisors make too much in their appraisals of two of the most concrete measures available to them? Or does the pattern suggest an opportunity that might be responded to by a low-cost but high-leverage policy like flexible workhours? See also Chapter 21.

Alternatively, that association also might mean that employees perceive that productivity gets too little attention, while too much attention goes to a few days more or less of sick leave or absence. Or, far worse still, do some truly sick employees come to work, unaware of or uncertain about firm policy, afraid to report sick for fear that they could never again do much right in the eyes of some myopic supervisors? When other pieces of evidence suggest such dynamics are operating, the following program of corrective action might be appropriate:

- Supervisory training about appropriate policies
- Emphasis on nonthreatening management styles
- More carefully informing employees about the sickness and absenteeism policies applicable to them
- More effective monitoring of supervisory behavior
- Policy changes, such as flexible work hours, that can reduce hassling about short-term indisposition or via flexibilities about arrival times.

Factor analysis permits an even more searching interpretation of the data summarized in Table 18.5. Basically, that interpretation deals with the troublesome issue of how many dimensions of reality are tapped by the twenty-one variables represented there. In general, for many important particulars—for example, choosing targets for ameliorative action—it is necessary to establish the specific multi-dimensionality underlying a data set.

Table 18.6 reports the results of a factor analysis of the twenty-one variables using principal components with iterations, followed by Varimax rotation to a terminal solution. The choice of factors is dictated by the number of eigenvalues greater than 1.0. The structure in Table 18.6 is quite clean, with only a few variables having even modest loadings on more than one factor. Table 18.6 reports all loadings of 0.3 or higher.
### Table 18.6 Summary of a Factor Analysis of Twenty-One Personal Benefits, Programs, and Practices

<table>
<thead>
<tr>
<th>Factors, or dimensions of reality, isolated by factor analysis</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worksite amenities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximate policies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Career development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-run financial security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retirement plan</td>
<td>.6151</td>
<td></td>
<td>.3536</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Savings plan</td>
<td>.4880(S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacation policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sick leave policy</td>
<td>.6517(S)</td>
<td></td>
<td>.4549</td>
<td>.3361</td>
<td></td>
</tr>
<tr>
<td>Dispensary and first aid</td>
<td>.4204</td>
<td></td>
<td>.6458(S)</td>
<td>.3664</td>
<td></td>
</tr>
<tr>
<td>Medical insurance</td>
<td>.4594</td>
<td></td>
<td>.5653(S)</td>
<td>.7546(S)</td>
<td></td>
</tr>
<tr>
<td>Parking lots</td>
<td>.4594</td>
<td></td>
<td>.5653(S)</td>
<td>.7546(S)</td>
<td></td>
</tr>
<tr>
<td>Absence policy</td>
<td>.6683(S)</td>
<td></td>
<td>.3664</td>
<td>.3664</td>
<td></td>
</tr>
<tr>
<td>Career development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employee promotions</td>
<td>.5653(S)</td>
<td></td>
<td>.7546(S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuition payments</td>
<td>.3696</td>
<td></td>
<td>.3504</td>
<td>.3063</td>
<td></td>
</tr>
<tr>
<td>Local phone policy</td>
<td>.4164</td>
<td></td>
<td>.3664</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affirmative action programs</td>
<td>.3696</td>
<td></td>
<td>.3504</td>
<td>.3063</td>
<td></td>
</tr>
<tr>
<td>Severance pay policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life insurance program</td>
<td>.6575</td>
<td></td>
<td>.5298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability insurance program</td>
<td>.6575</td>
<td></td>
<td>.5298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security practices</td>
<td>.4749</td>
<td></td>
<td>.3104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternity leave policy</td>
<td>.3182</td>
<td></td>
<td>.3104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lounges</td>
<td>.5563(S)</td>
<td></td>
<td>.6016(S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevators</td>
<td>.5563(S)</td>
<td></td>
<td>.6016(S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrooms</td>
<td>.5628(S)</td>
<td></td>
<td>.6016(S)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: An "S" designates a variable which when classified by appraisal category achieved a statistically significant overall difference in Table 18.5.
TABLE 18.7 Percentages of Variance Accounted for by Five Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>% of Variance</th>
<th>Cumulative % of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Worksite amenities</td>
<td>26.1</td>
<td>26.1</td>
</tr>
<tr>
<td>II Proximate financial supports</td>
<td>8.3</td>
<td>34.4</td>
</tr>
<tr>
<td>III Policies re time off</td>
<td>6.1</td>
<td>40.5</td>
</tr>
<tr>
<td>IV Career development</td>
<td>5.7</td>
<td>46.2</td>
</tr>
<tr>
<td>V Long-run financial security</td>
<td>4.8</td>
<td>51.0</td>
</tr>
</tbody>
</table>

With some confidence, then, we conclude that the twenty-one variables tap at least five dimensions of reality, which may be thought of as potential targets for managerial or employee action in four significant ways. First, the percentages of total variance accounted for by each of the five factors may be thought of as estimates of their relative significance to respondents in the host organization at the time of the survey. Table 18.7 shows specific percentages of variance for each of the five factors.

Second, the individual loadings provide some guidance about the relative impact of specific variables contributing to a factor. In sum, these loadings help identify variables that provide special leverage for inducing change on a specific factor. For example, the loadings of Worksite Amenities suggest that the character and quality of elevator service, lounges, and restrooms dominate in the factor which accounts for most variance in the data. Ameliorative action in their cases might be especially cost-effective.

Third, information about the interaction of performance appraisals and responses on the twenty-one items also provides further guidance to management. Recall that Table 18.5 indicates nine cases in which differences in appraisals co-vary significantly with differences in responses to survey items—those nine cases are indicated by an “S” in Table 18.6. In the case of Factor I, for example, those respondents with the highest appraisals had the most negative scores on elevator services and restrooms. This implies a relatively unconflicted choice for management. Not only does Factor I account for the largest portion of the variance in the twenty-one relevant variables, but the two highest loadings on that factor relate to variables on which those with the highest appraisals also have the greatest concern. The situation is more complicated for Factor III, where those with the lowest appraisals are significantly less favorable on the two key target variables—sick leave and absence policy.

The data in Table 18.6 cannot decide for employees or management, patently. Just as clearly, however, the analysis does highlight central issues for choice.
Fourth, available knowledge does not permit definite prediction of the con-
sequences of ameliorative action on any of the factors. Will production increase
if resources are expanded to improve Factor I? The best guess is, probably not,
because Factor I seems to be a hygiene rather than a motivating factor, in the
vocabulary popularized by Herzberg, Mausner, and Snyderman (1959). In their
schema, motivators tend to be built into the work itself, and derive from elements
such as potential in work for fulfillment, achievement, recognition, and so on. In
Table 18.6, Factor IV probably comes closest to being a motivator in Herzberg’s
language.

GREATER SUCCESS BY SPECIFYING IMPORTANT
DIFFERENCES BETWEEN PEOPLE

This one approach to Irony VII reinforces the virtues of specifying differences
between people, with the likelihood that OD will become more successful in
making intended things happen. Both the data, as well as illustrations of how
they might be used, strongly suggest the usefulness in survey/feedback designs
of specifying differences in performance appraisals. Pretty clearly, major practi-
cal and normative issues get neglected in equally weighting each survey response
by aggregating individual responses which are differentiated, if not at all, only
in terms of common demographic variables.

Three questions take stage center, and attention will be given to each in
turn. The questions are:

What are the attractions of giving survey responses equal weight?
Are there any general developmental trends that seem likely to raise the
attractiveness of differentiating individuals?
Are there any trends in OD itself that will encourage differentiating individ-
uals? For example, various market and other forces might place a special
premium on executive development, as happened in the years sur-
rounding 1970, and that also seems in the wind nowadays.

In general, the answers to the second and third questions imply that the attractions
elicited in response to the first question will have a reduced salience in the near
future. Hence, improving OD success rates via differentiating respondents will
face reduced challenges.

Why Weight Responses Equally?

Numerous attractions encourage weighting each survey response equally, despite
the evidence above of the usefulness of differentiating respondents. Consider five
varieties of such attractions—existential, motivational, philosophical, mensural,
and stylistic.
Existentially, the implicit basic rationale in survey/feedback resembles that in political democracy: those who wear the shoes know best if they pinch. Hence the reasonableness of a guarantee to each shoe-wearer of an equal voice via a vote or a survey, so that no major pinching goes unremarked for very long. The orthodox rationale in political philosophy supporting this guarantee emphasizes the danger if such pinching continues without redress, and especially so in the case of the masses of system members. For all systems ultimately depend on mass support but, for many pressing reasons, mass opinion may be slow to form and often ineffectual even long after it does coalesce. Since mass opinion nonetheless will be ultimately determinative, the alternative to incremental amelioration is periodic but massive outbursts of pent-up frustration that may indiscriminately damage many innocents, and perhaps long after the possibility of timely action has passed.

Motivationally, in addition, the promise of equal weighting might encourage participation in survey/feedback designs, especially by those who normally are disadvantaged in other modes of influencing their system. Of course, few systems are democratic in the full “one person, one vote” sense.

Philosophically, many good things in our social and political systems rest on a belief in some kind of ultimate equality, such as equality before the law. Reality often falls far short of the ideal, of course, and booster shots via “one person, one vote” surveys thus may be tactical or even strategic.

Mensurally, in addition, “one person, one vote” variants avoid some ticklish and even dangerous issues. We can sample only. What are the criteria for unequal treatment? And who determines and enforces them, with what effects on the freedom of which participants in a system?

Take the present case, which selects performance appraisal as the criterion. The rationale is uncomplicated. Appraisals exist; they provide one important measure of the perceived value of the individual to the organization; and appraisals influence who gets what in organizations. However, appraisals are not carved in stone; individuals can progress or regress; and appraisals may be invalid as well as unreliable, to reflect more prejudice, lock-step thinking, and perhaps malefic than to reflect tough, dispassionate analysis.

The linkage of appraisal and survey data also could result in an unfortunate concentration of power, as in the risk of a management purge of the Sub-Modals because of their lower level of support for the organizational regime. The danger will be especially great where informed consent is lacking, as when organization members innocently provide questionnaire responses that are linked to other data that might motivate action against specific individuals, without sufficient appreciation by those providing the data of the awkward possibilities of their action. In effect, insidious managements might entrap respondents into providing unwitting evidence against themselves or others.

Here, the ODer will face a challenge in protecting sources of data, and especially so in those cases when management’s interest is somehow enjoyed.
Style attractions, finally, also can encourage equal weighting of opinions in survey/feedback designs. Some OD professionals prefer a style that is unreliedly helpful versus even potentially hurtful or punitive. Differentiations of the kind illustrated above may not be their cup of tea. For differentiation raises tough questions, whose answers may be helpful to the system while punishing to some members. These questions include: Who really counts in System X? Who do we want to keep, and who is more dispensable?

Of course such questions get raised and answered in all systems, in one way or another. But facing such questions explicitly and directly is no easy matter, especially since objective answers seldom exist. And just because someone will raise and answer such questions does not mean that OD specialists will do so, or should.

These constitute powerful reasons for weighting all survey respondents equally, but counterforces also are growing. The next two subsections illustrate this tendency, and also reframe the differentiation argument in terms of various developmental features. Broadly, differentiation is not a yes/no issue, as this subsection suggests. Rather, progressive differentiation can enhance OD success rates as two classes of developments mature.

**Trends in Three Developmental Areas**

Some major components in the antidifferentiation rationale have lost substantial force in recent years. Consider the revolution in access to personal computers. Not long ago, data-processing problems militated against routinely differentiating respondents, even in large organizations. Now, the required technology is within the convenient reach of all interested observers, even though beliefs about analytical problems still persist.

Patience and care definitely are required, but elaborate libraries of computational programs and hardware now serve all but the most casual users. Developments in the computational arena, in short, reduce the practical force of the antidifferentiation rationale. There is no reason in principle why everyone cannot do their own interpretative analyses of the same data sets.

In addition, the extension and maturation of OD activities tend to reduce the credibility of a second major reason for not differentiating individuals. Consider the objection that employees who know that appraisals or similar information will be integrated with survey data might provide bogus responses. Such linkages typically require that survey respondents be identified and—even given a trusted and discrete third party—the potentials for, and temptation toward, mischief inhere in anyone’s ability to match names and responses.

Since there seems no convenient ways out of the dilemmas of identifying respondents, some observers conclude that this precludes differentiation. Devious
or underhanded approaches to linking survey and other data are ruled out by OD values, as well as by ethical standards and law.

Alternatively, asking respondents to supply their own appraisals as one of their own appraisals as one of their responses to the survey questionnaire probably will not do. This approach does not require identifying individuals, but experience with asking individuals for some recall data—like family income—reflects major reliability and validity problems.

This second class of components of a rationale against differentiation, however, can be looked at in developmental terms. In effect, such objections resolve into this truism: Where OD values do not exist, one can expect behaviors and attitudes to reflect anti-OD values. Stated positively, differentiation will be more accepted as an organization develops regenerative systems—high openness, owning, and trust, along with low risk. And differentiation can both contribute to regenerative interaction and be legitimated by it. As OD is extended, in short, so will differentiation be more probable. In this sense, OD success rates enhance its credibility which, in turn, will legitimate differentiation. And that will further raise success rates.

As a final point concerning developments bearing on the issue of differentiating respondents in survey/feedback designs, many observers have emphasized the probable resistance from employee associations or labor unions. Adversarial relations have tended to dominate in the labor/management arena; OD has not given generous attention to unions, in general; and the labor movement has not been very positive toward management-initiated surveys to begin with. Especially in industries where costs are easily passed on to consumers, management may be motivated to avoid negative reactions from professional or union officials, or any other source.

The question remains: So who needs data? Nowadays, the arguments against withholding data have less currency, because of market turbulence and shorter product half-lives.

Developmental trends favoring differentiation appear even in the arena of historical “hard ball” between labor and management. Most signal, cooperation between these historical adversaries has become increasingly common, as in the QWL movement (e.g., Sun, 1988). Competition from abroad and loss of jobs have proved powerful reinforcing of those who believed in the gospel of joint labor/management initiatives as a matter of principle.

**Trends in OD Tensions**

The bias against differentiating respondents gets additional support from certain historical tensions within OD as an area of professional activity. The basic question: Is OD practice elitist (Ross, 1971) or populist (Bass, 1967, esp. 221–225)?
The fact of the matter seems to be that OD practice not only does, but must, encompass aspects of both orientations. OD practice will have strong elitist features, if only because clients often will be figures in authority. At the same time, OD practice needs a populist thrust in order to generate credible data for analysis and to develop motivation for implementing any action plans. In short, lack of balance has costs, and either/or will not do.

Guaranteeing respondents anonymity is a convenient way to achieve the required working balance of elitist and populist orientations, at least early in the game. That is obvious.

Developmental trends in this critical professional particular are mixed, but the trend seems unmistakable. ODers are increasingly sensitive to the complex character of their profession, and balance seems more the order of the day than procrustean in/out distinctions (e.g., Browne, Cotton, and Golembiewski, 1977). Moreover, managements seem to understand more and more that OD’s basic value inheres in its multiple in-betweennesses, and hence resist ODers forcing choices as to whether “you’re with us or against us” (e.g., Golembiewski and Kiepper, 1998 esp. pp. 216–226).

In such a developmental context, reasonably bounded differentiation can develop. Sharp elitist/populist orientations favor nondifferentiation.

CONCLUSION

Three points summarize this analysis, then. There seem to be multiple advantages in specifying differences between respondents to survey/feedback designs, even though this approach contains potential for mischief. This potential helps explain why survey/feedback designs have seldom differentiated their respondents, and encourages the prediction (and the prescription) that such differences will (and should) come only after careful deliberation, reinforced by effective guarantees of individual anonymity and/or high levels of mutual trust and confidence. Finally, several developmental tendencies encourage the view that reasonable progress toward differentiation can and will be made.

Research with burnout illustrates this latter class, and often can be used in combination with surveys.

A Personal View

President Truman once noted that he needed more “one-handed” advisors. Like the present analysis, many of Truman’s advisors were good at generating multiple alternatives: “on the one hand . . . , but on the other hand. . . .” Truman wanted fewer hedged opinions and more specific recommendations for action.

So what is my one-handed view, once the alternatives are stated for the
record? My goal in this chapter is to suggest the virtue of a principle or operating guide. Whenever you can safely do as I propose—differentiate respondents—do so. Also try to do more of that as time goes by, and recognize that the present approaches to specificity about personal differences are illustrative only. The discussion of the several developmental tendencies should encourage the leading toward the conclusion that differentiation does not merely mean swimming up a torrential stream.

When circumstances seem to encourage unspecificity or anonymity, I suggest two actions: reevaluate to discount the possibility that you are simply taking the easy way out; and then if your initial sense of the situation remains, be mindful of what may be lost, or even surrendered, because of the lack of specificity.

In short, this chapter has a consciousness-raising quality—a sense of what can be, even though the conditions at any specific time might tether OD short, in a specific place at a certain time. In the long run, progress gets defined as the tension between what can be and what has to be. This chapter seeks to contribute to that tension, but without prescribing indiscriminately for the diverse settings that are encountered in the world of work.

Moreover, this chapter reinforces the rationale for seeking to develop and heighten regenerative interaction in organizations: to increase trust, openness, and owning while reducing risk. Such settings raise the probability of constructive specification of individual differences. Indeed, in the long run, it is only as more and more differences are specifiable that advanced regenerative interaction is possible.

NOTES

1. Total respondents to the QWL questionnaire numbered 2,671, which represents a voluntary response rate of some 70 percent of the total employment. All respondents were located at headquarters or several nearby sites, representing the full range of functions from Research and Development through Manufacturing to Marketing. Of the respondents, 770 chose not to identify themselves, and 376 were excluded because they were in the field sales force and posed different analytic problems. This left 1,525 cases for the present study. Performance appraisals were not available for 51 "identifiers." Hence, the present N of 1,474.

   Elaborate comparisons on a large number of demographics establish that the 2,671 voluntary respondents constitute a close replica of the total employment.

   The 1,525 cases considered here are merely an interesting subpopulation, however, and no implied or express claim about their representativeness will be made. Convincing evidence (Golembiewski and Billingsley, 1976) does establish that identifiers provide survey responses that are clearly more favorable or optimistic about their organizational experiences and attitudes. Here, however, it will not be possible to test whether appraisal differences among the non-identifiers generate patterns similar to those described in the text for identifiers.
2. Specifically, all respondents were asked to enter a self-chosen, nine-digit code on their response sheets, which were to be deposited in large boxes scattered throughout the organization. If they wished, respondents also could enter that code number plus their name on a separate sheet, which they sent directly to the consultant/researcher in a stamped, self-addressed envelope. A letter to respondents explained that this procedure would permit analyses like the present one, and also might later facilitate efficient resurveys to test for trends.

3. In at least one case, only minor differences existed between performance appraisals gathered by self-reports and those from archival data. See Golembiewski, Munzenrider, and Stevenson, 1986, 87–101.

REFERENCES


Golembiewski, R. T., Yeager, S., Hilles, R., and Carrigan, S. B. (1977). Toward increasing the specificity of survey/feedback designs: Costs/benefits of differentiating respon-
Limitations of “One Person, One Vote”


Common sense tells us that individuals will differ in the degree of their personal slack—the physical and emotional reserves that different individuals can mobilize to deal with the different phenomena they encounter—and the research reported in this chapter not only confirms common sense but also draws attention to associated theoretical and practical opportunities. The focus is on the relationship of an individual’s experienced strain with normal coping limits. That strain can variously challenge those coping capacities or even overwhelm them.

The specific focus is on burnout, with the goal of testing several critical points. The clinical and anecdotal literatures on burnout suggest that noxious effects accompany and exacerbate high levels of strain. Building on these general observations, this research uses a paper-and-pencil instrument to become specific—to assess degrees of burnout in a large population—and relates those differences to a panel of nineteen self-reports of physical symptoms. Burnout proves to have direct and marked associations with a broad range of physical symptoms, as expected, and with much managerial relevance.

These findings add urgency to dealing with stress at work, as in corporate wellness programs, and the results also underscore the conceptual myopia of much of the OD literature. That literature tends to neglect individual differences between the targets of OD interventions, in general as well as in the specific case of the degree to which individuals have physical and emotional slack that can impact their basic abilities to choose or to change or even just get along.
Despite the face-valid case to the contrary, the neglect of individual differences has a long tradition in OD. For example, selection of participants became an issue in the use of T-groups, following reports of psychiatric casualties (e.g., Yalom and Lieberman, 1971), but the filters have been for gross in/out decisions, such as cautions that individuals not consider the experience a substitute for therapy, as contrasted with determining which individuals are likely to profit from specific designs. Some exceptions exist, but they are few and far between (e.g., Harrison, 1965). Similarly, team-building designs seldom take explicit account of personal differences, despite some suggestions of the myopia of this general practice.

The results presented in this chapter encourage seeing this neglect of individual differences as a significant shortfall. Again, this chapter provides another perspective on the theme that the substantial estimates of OD success rates can be enhanced in direct ways by enlarging OD’s theoretical base to encompass individual differences, such as those in burnout.

Overall, burnout is significant, and pervasive. Hence, this chapter may be conveniently seen as part of the rationale for guaranteeing emotional slack in organizations to facilitate choice and change by careful attention to burnout. The health costs implied by advanced burnout also provide a powerful motivation for doing better-than-average in connection with burnout.

A CONCEPTUAL CONTEXT FOR IRONY VII

In sum, then, there is no problem in isolating a significant context for burnout and Irony VII. Virtually all observers expect that greater incidences of a long catalog of psycho-physical disturbances will be associated with increasing burnout—high rates of headaches, lingering colds, backaches, gastrointestinal disturbances, and other indicators of impaired physical functioning, among a huge list.

Most of these reports are anecdotal or involve small populations, in contrast to this effort to identify physical covariants of differential burnout in large populations using a paper-and-pencil instrument. Details follow about a host organization and conventions for measurement.

The purposes are profoundly simple. If we can define a valid and reliable measure of burnout, and if we can relate it consistently to a growing range of phenomena, then we will be progressively in the business of attaining higher success rates—that is, getting more of what we want from OD interventions, more of the time, because of our knowledge of the empirical world.

Host Population

The present data come from one division of a federal agency, which has about fifty similar offices throughout the country performing the same set of people-
helping activities. Details are available elsewhere (Golembiewski, Munzenrider, and Stevenson, 1986), and only a general sense of the population will be provided here. The division has a total employment of approximately 2,600, with the effective response rate for different items approximately 55–60 percent. Division missions-and-roles center around direct contact with clients, under conditions that can be emotionally arousing and even hostile. The population might best be described as a “middle slice” of a federal bureaucracy, with the bulk of the present respondents falling in position classes GS-5 through GS-12. The full federal employment schedule includes eighteen grades, plus a senior executive group.

The Basic Measures

This study utilized three basic sets of measures, and the project received support at three levels. Headquarters provided the initial encouragement, and general design and questionnaire format were later approved by the division’s union. Field managers proved to be a more difficult target, and a presentation was made at a national managers’ meeting that emphasized the organizational consequences of burnout. Three of the fifty-plus field managers refused to have their employees participate, and several others took lukewarm stances toward the research project.

This study utilized three basic sets of measures—for example, burnout, physical symptoms, and over a score of managerially relevant measures: participation, job, involvement, and trust. The focus here will be on the first two sets of measures, a major convenience to illustrate the pattern of associations that also holds in very much the same pattern for the managerial measures. The two sets of focal measures here are introduced briefly.

Burnout Measures

Psychological strain is estimated by the Maslach Burnout Inventory, or MBI (Maslach and Jackson, 1982, 1986), which contains twenty-five items and is modified for present purposes (Golembiewski, Munzenrider, and Carter, 1983). The items tap three subscales, to review:

- **Depersonalization.** High scores indicate that respondents tend to distance themselves from others—to reify human contacts and to view them as categories of things.
- **Personal Accomplishment (reversed).** Low scores imply that individuals are doing well on jobs they consider worthwhile.
- **Emotional Exhaustion.** High scores describe individuals who are strained beyond their comfortable coping limits—that they are nearing or beyond, the “end of the rope,” emotionally and psychologically speaking.

The MBI subscales are used to generate a phase model, which basically proposes that increases in Depersonalization characterize the earliest and least-virulent stages of burnout. Up to a point, some depersonalizing may not adversely
affect performance—indeed, may even enhance it. Beyond that currently indeterminate point, however, increases in Depersonalization negatively impact Personal Achievement. Sufficient increases in the two subscales, in turn, can trigger high levels of Emotional Exhaustion.

This simple decision rule generates an eight-phase model, when respondents are each distinguished as High versus Low on the three MBI subscales. High versus Low assignments are made in terms of median cuts in the present population, which constitutes a substantial (N = 1555) population from a single organization with a relatively homogeneous set of missions and roles. Subscale norms have been reported elsewhere (Golembiewski and Munzenrider, 1984). Basic features of the phases are defined in Table 19.1.

**Physical Symptoms**

Self-reports concerning physical well-being derive from responses to a conventional list of possible complaints (Quinn and Shepard, 1979, esp. p. 30). Respondents used four possible response stems for describing their experiences with each symptom: Often, Sometimes, Rarely, and Never. The response stems relate to the following nineteen indicators of well-being, with the scoring being uniform so that Often is coded “4” and Never is “1.”

1. Cramps in my legs
2. Pains in my heart
3. Tightness or heaviness in my chest
4. Trouble breathing or shortness of breath
5. Swollen ankles
6. Pains in my back or spine
7. Pains in my stomach
8. Headaches
9. Coughing or having heavy chest colds
10. Stiffness, swelling, or aching in my leg muscles and joints
11. Becoming very tired in a short period of time (fatigued)
12. Having trouble getting to sleep
13. Having trouble staying asleep

---

**Table 19.1** Phases of Psychological Burnout

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depersonalization</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
<td>Lo</td>
<td>Hi</td>
</tr>
<tr>
<td>Personal accomplishment</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
</tr>
<tr>
<td>(reversed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional exhaustion</td>
<td>Lo</td>
<td>Lo</td>
<td>Lo</td>
<td>Lo</td>
<td>Hi</td>
<td>Hi</td>
<td>Hi</td>
<td>Hi</td>
</tr>
</tbody>
</table>
14. Finding it difficult to get up in the morning
15. Feeling my heart pounding or racing
16. Hands sweating or feeling damp and clammy
17. Feeling nervous or fidgety and tense
18. Being completely worn out at the end of the day
19. Having a poor appetite

Note that Cronbach’s alpha for Total Symptoms is 0.89.

FINDINGS

In general, the measures of burnout co-vary in regular and predictable ways with self-reports about physical symptoms. Four sections detail the findings which, in turn, focus on the phases of burnout, a factor analysis of the 19 Michigan physical symptoms, emotional health, and one estimate of medical costs, and a comparison with two national samples. Straightforwardly, direct associations are expected between burnout and all of the measures of physical symptoms.

Phases and ANOVA

The usefulness of the burnout phases gets major support. Total Symptoms varies regularly as well as robustly by phases of burnout, as two points establish. See Table 19.2. Thus, over 96 percent of all paired-comparisons vary progressively: that is, Total Symptoms for Phase I is less than for Phase II, which is less than for Phase III, and so on. In addition, over 60 percent of the twenty-eight possible paired-comparisons are statistically different at the .05 level, as judged by the LSD test, modified for unequal sample sizes (for details, consult Table 19.6, first row, and Table 19.7).

<table>
<thead>
<tr>
<th>19 Michigan Symptoms</th>
<th>In expected direction and statistically significant</th>
<th>In contrary direction and statistically significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.4%</td>
<td>60.4%</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

Source: From Golembiewski, Boudreau, Munzenrider, and Luo, 1996, p. 86.
Factor Analysis of Nineteen Symptoms

Do specific clusters of the nineteen symptoms differentially co-vary with phases of burnout? Yes. Factor analysis implies that the nineteen symptoms do not simply elicit generalized feel-good/bad responses from respondents. As Table 19.3 suggests, at least four separate dimensions seem to exist, and they are labeled tentatively in Table 19.3. Factor I accounts for over 33 percent of the variance, and the other three add about 20 percent in total.

To facilitate analysis, each physical symptom is assigned to a single cluster of symptoms, as Table 19.4 details. The assignments took into account the original factorial loadings.

This analysis of the differential affinity of burnout for specific clusters of symptoms can be extended in two ways. First, Table 19.5 displays the simple correlations of five aggregate symptoms scores with four MBI scores. Factor I and Total Symptoms share an average variance with the four aggregate burnout measures of nearly 16 and 13 percent, respectively. For the other three burnout measures, that average drops to 4.72 percent.

Second, this pattern of differential associations of burnout and specific clusters of symptoms is both reinforced and extended by a focus on the phases. The point is developed in three stages, beginning with Table 19.5, which reports on the variation of the five aggregates of symptoms and the eight phases of burnout. Note that in all five cases, statistically significant overall variation exists, and especially so in the cases of two clusters of symptoms—Factor I and Total Symptoms, in that order. This establishes the differential but substantial robustness of this association: self-reports of physical symptoms increase as burnout phases progress from I to VIII.

We can establish that these differences are generally regular as well as robust, overall, with the most marked regularity again being associated with Factor I and Total Symptoms. The central point is established by the Least Significant Difference test. LSD determines the statistical significance of the differences between each of the 140 possible paired comparisons for all aggregates of symptoms of phases; that is, for any measure of symptoms, the question is: is the score for Phase I less than that for II; is the score for Phase II less than that for III; and so on through Phase VIII?

The data are not reproduced here but, overall, 43.6 percent of all 140 paired-comparisons attain the .05 level of statistical significance. There are no absolutes in such matters, but anything greater than 20 percent statistically significant cases could be taken as indicating a noteworthy regularity of increases in scores, phase by phase.

However, this aggregate record obscures even as it supports the conclusion that the phases of burnout and physical symptoms co-vary both regularly and robustly. Table 19.6 provides useful specificity. Two clusters of symptoms again stand out as having the most marked associations with the phases of burnout:
TABLE 19.3 Factor Analysis of Nineteen Physical Symptoms, by Varimax Rotation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factors</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I General</td>
<td>II Cardiovascular</td>
<td>III Noncardiac pain</td>
<td>IV Sleeplessness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>enervation and agitation complaints</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leg cramps</td>
<td>.56</td>
<td></td>
<td>.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart pain</td>
<td>.68</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tightness in chest</td>
<td>.73</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trouble breathing</td>
<td>.32</td>
<td>.56</td>
<td>.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swollen ankles</td>
<td>.41</td>
<td></td>
<td>.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back pain</td>
<td>.51</td>
<td></td>
<td>.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stomach pain</td>
<td>.36</td>
<td>.36</td>
<td>.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headaches</td>
<td>.42</td>
<td>.42</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coughs and colds</td>
<td>.30</td>
<td></td>
<td>.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stiffness</td>
<td>.64</td>
<td></td>
<td>.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatigued</td>
<td>.52</td>
<td>.36</td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting to sleep</td>
<td>.31</td>
<td>.43</td>
<td>.76</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Staying asleep</td>
<td>.82</td>
<td></td>
<td>.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting up</td>
<td>.56</td>
<td></td>
<td>.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart pounding</td>
<td>.40</td>
<td>.53</td>
<td>.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweating hands</td>
<td>.36</td>
<td>.32</td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nervous</td>
<td>.60</td>
<td></td>
<td>.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worn out</td>
<td>.67</td>
<td></td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor appetite</td>
<td>.35</td>
<td></td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>6.4</td>
<td>1.5</td>
<td>1.1</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Percent variance</td>
<td>33.4</td>
<td>7.9</td>
<td>6.7</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>Percent cumulative variance</td>
<td>33.4</td>
<td>41.4</td>
<td>48.1</td>
<td>53.5</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *N = 1,512, with alpha for all items = 0.89.
* Only loadings > .30 are listed.
**Table 19.4** Modified Factors, Nineteen Physical Symptoms

<table>
<thead>
<tr>
<th>Factor I: General enervation and agitation (alpha = 0.81)</th>
<th>Corrected item-to-total correlation</th>
<th>Alpha, Item deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach pains</td>
<td>.49</td>
<td>.79</td>
</tr>
<tr>
<td>Headaches</td>
<td>.47</td>
<td>.79</td>
</tr>
<tr>
<td>Coughs and colds</td>
<td>.35</td>
<td>.81</td>
</tr>
<tr>
<td>Fatigued</td>
<td>.61</td>
<td>.78</td>
</tr>
<tr>
<td>Trouble getting up</td>
<td>.52</td>
<td>.79</td>
</tr>
<tr>
<td>Sweating hands</td>
<td>.45</td>
<td>.80</td>
</tr>
<tr>
<td>Nervous</td>
<td>.62</td>
<td>.78</td>
</tr>
<tr>
<td>Worn out</td>
<td>.61</td>
<td>.78</td>
</tr>
<tr>
<td>Poor appetite</td>
<td>.42</td>
<td>.80</td>
</tr>
</tbody>
</table>

Factor II: Cardiovascular complaints (alpha = 0.81)

| Heart pain                                               | .62                                 | .76                 |
| Tightness in chest                                       | .69                                 | .72                 |
| Trouble breathing                                         | .61                                 | .76                 |
| Heart pounding                                            | .57                                 | .78                 |

Factor III: Noncardiac pains (alpha = 0.67)

| Leg cramps                                               | .48                                 | .58                 |
| Swollen ankles                                           | .36                                 | .66                 |
| Back pain                                                 | .44                                 | .61                 |
| Stiffness                                                 | .53                                 | .54                 |

Factor IV: Sleeplessness (alpha = 0.85)

| Getting to sleep                                         | .73                                 |
| Staying asleep                                            | .73                                 |

Factor I and Total Symptoms. In sum, over 96 percent of the comparisons of pairs increase, phase by phase, for those two clusters of symptoms, over 62 percent of those expected differences in paired-comparisons achieve statistical significance, and none of the 3.6 percent of contrary cases attain the .05 level. See Table 19.6.

The other three aggregates of symptoms also support the regular increase in reported symptoms, by progressive phases of burnout, but with variable if-always-substantial force. Factors III and IV patently reflect the least-marked pattern of association of phases of burnout and reported symptoms, albeit still a noteworthy pattern.
<table>
<thead>
<tr>
<th>Aggregates of symptoms</th>
<th>Depersonalization</th>
<th>Personal accomplishment</th>
<th>Emotional exhaustion</th>
<th>Total score</th>
<th>Average shared variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total symptoms</td>
<td>.29</td>
<td>.20</td>
<td>.46</td>
<td>.43</td>
<td>12.99</td>
</tr>
<tr>
<td>Factor I</td>
<td>.32</td>
<td>.22</td>
<td>.51</td>
<td>.47</td>
<td>15.79</td>
</tr>
<tr>
<td>Factor II</td>
<td>.22</td>
<td>.14</td>
<td>.31</td>
<td>.29</td>
<td>6.18</td>
</tr>
<tr>
<td>Factor III</td>
<td>.12</td>
<td>.10</td>
<td>.22</td>
<td>.20</td>
<td>2.82</td>
</tr>
<tr>
<td>Factor IV</td>
<td>.21</td>
<td>.13</td>
<td>.27</td>
<td>.27</td>
<td>5.17</td>
</tr>
<tr>
<td>Average shared variance, %</td>
<td>5.87</td>
<td>2.70</td>
<td>13.78</td>
<td>12.04</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 19.6 Summary Associations of Five Aggregates of Symptoms with Burnout Phases, All Paired Comparisons

<table>
<thead>
<tr>
<th>Aggregates of symptoms</th>
<th>In expected direction and statistically significant(^a)</th>
<th>In contrary direction and statistically significant(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Symptoms</td>
<td>96.4</td>
<td>60.7</td>
</tr>
<tr>
<td>Factor I General enervation and agitation</td>
<td>96.4</td>
<td>64.3</td>
</tr>
<tr>
<td>Factor II Cardiovascular complaints</td>
<td>92.9</td>
<td>42.9</td>
</tr>
<tr>
<td>Factor III Noncardiac pains</td>
<td>92.9</td>
<td>17.9</td>
</tr>
<tr>
<td>Factor IV Sleeplessness</td>
<td>71.5</td>
<td>32.1</td>
</tr>
</tbody>
</table>

Note: \(^a\)Refers to 0.5 level on Least Significant Difference test, as modified for unequal sample sizes.

### Phases and Emotional Conditions

Several independent investigations show the same pattern, almost without exception, no matter how burnout is defined (e.g., Goldberg, 1972). Here, let us stick to the phase model, where the research is absolutely uniform (e.g., Golembiewski, Boudreau, Munzenrider, and Luo, 1996, esp. pp. 95–135), both in North America as well as throughout the world.

Consider here only one operational measure of emotional health—General Health Questionnaire, or GHQ (Goldberg, 1972), a paper-and-pencil instrument on which high scores indicate nonpsychotic psychiatric symptoms at a level which, if isolated by psychiatric interviews, would justify clinical intervention.

No surprises are found in the research. Increases in GHQ scores are associated, regularly and robustly, with advancing phases of burnout. Note that there are several ways of estimating GHQ scores, which all lead to the same general conclusion. One useful way of estimating GHQ scores, for example, involves distinguishing whether respondents are “normals” or “cases,” with the latter deemed appropriate candidates for clinical intervention. In most cases, the proportion of “cases” increases phase-by-phase, which association varies in magnitude but is the same in pattern for all of the numerous populations so far studied.
Preliminary Profile for Burnout

TABLE 19.7 Incidence of “Cases” by Phases of Burnout, in percent

<table>
<thead>
<tr>
<th>Phases</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16.1</td>
<td>30.7</td>
<td>21.0</td>
<td>57.7</td>
<td>42.6</td>
<td>45.5</td>
<td>54.0</td>
<td>66.9</td>
</tr>
</tbody>
</table>

Chi-square = 75.43, with P < .0000


See Table 19.7 for one kind of support for this central generalization. The data in this table come from one North American setting and are close analogs of the usual finding: the growing proportions of “cases” as the phases progress I → VIII in all other North American settings (Golembiewski et al., 1996, pp. 100–103). The same pattern also holds for all global settings so far studied (Golembiewski et al., 1996, pp. 103–105).

The direction of the proportion of “cases” X phases is not absolutely uniform, but almost so. In Table 19.8, for example, 23 of the 28 paired-comparisons—or about 82 percent—fall in the expected direction.

Phases and Medical Costs

Given the two points above, plus the fact that perhaps 40-plus percent of the participants in relevant studies are assigned to advanced studies VI–VIII, it requires no sophisticated math to conclude that the costs of burnout will be high in the present regard. On the face of the summaries in this chapter, the magnitude of the costs of the phases will include decrements in physical and emotional costs related to a broad range of variables, and will be suggested by comparisons of (let us say) the proportions of those in Phases I–III versus those in Phases VI–VIII.

Evidence is hard to assemble, if for no other reason than that confidentiality often sets limits on the uses that can be made of appropriate data. In what may still be a unique study (Jackson and Manning, 1995), direct health costs in an organization’s insurance program varied quite directly with the phase assignments of individual employees. The distribution of costs X phases in Table 19.8

TABLE 19.8 Paired-Comparisons, Insurance Costs X Phases

<table>
<thead>
<tr>
<th>In expected direction</th>
<th>In expected direction and statistically significant</th>
<th>In contrary direction</th>
<th>In expected direction and statistically significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.9%</td>
<td>21.4%</td>
<td>32.1%</td>
<td>5.6%</td>
</tr>
</tbody>
</table>

Source: From Golembiewski, et al., 1996.
achieves $P < .05$, and this summary of paired-comparisons indicates a substantially regular pattern. The pattern clearly is not a perfect one, but its general message seems quite clear for employees, unions, and managements.

**SO HOW BAD IS BAD?**

These detailed data manipulations provide strong evidence that physical symptoms worsen as measures of burnout increase, with the phases seeming to be particularly useful indicators of an individual’s state.

We can provide no complete sense of the magnitudes involved, but that is remedial here in part, and briefly. With a focus only on physical symptoms and the phases, we can provide a ballpark estimate of this central question: How bad is bad? Comparisons with two national samples of respondents provide a real sense of the great costs of advanced burnout.

Details are available elsewhere (Golembiewski, Boudreau, Munzenrider, and Luo, 1996), but some gross comparisons suggest the virulence of burnout’s covariants. In sum, those in Phase VIII experience some multiple of the symptoms reported by two national samples—two, five, or even ten times more symptoms (e.g., Quinn and Shepard, 1979). In contrast, those in Phase I experience a small fraction of the symptoms reported by all survey respondents.

More useful details put two sources of data to combined use. Thus, for many worksites, we know the distributions of phases as well as the reported incidences of the 19 Michigan symptoms. And for two national surveys, we also know the distributions of physical symptoms reported by survey respondents (Quinn and Shepard, 1974; Quinn and Stains, 1979).

Details would overwhelm this text (e.g., Golembiewski et al., 1996, pp. 85–91), but the pattern is so clear that generalizations cannot go far wrong. Roughly, those in Phases VI–VIII regularly tend to report far more symptoms than those in Phases I–III; and the latter, in turn, report fewer symptoms than the two national populations. Consider only the symptom “feel fatigued,” and those who respond “never” and “often,” as summarized in Table 19.9.

<table>
<thead>
<tr>
<th>Table 19.9</th>
<th>Selected Comparative Data on Physical Symptom “Feel Fatigued”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category of respondent</td>
<td>Number</td>
</tr>
<tr>
<td>Phases I–III at Site B</td>
<td>653</td>
</tr>
<tr>
<td>Phases VI–VIII in Site B</td>
<td>652</td>
</tr>
<tr>
<td>1972 National Sample</td>
<td>2145</td>
</tr>
<tr>
<td>1977 National sample</td>
<td>1081</td>
</tr>
</tbody>
</table>
Clearly those in Phases VI–VIII do more poorly on “feel fatigued,” and even worse on some of the other Michigan symptoms.

**GREATER SUCCESS BY SPECIFYING PHASES OF BURNOUT**

At a global level, these findings suggest that people differ in their levels of burnout, and these differences are associated directly with a person’s willingness and ability to successfully undertake choice or change. That establishes a direct connection with the probable success of OD applications. In this case, the physical symptoms associated with progressive burnout suggest a reduced willingness and ability to deal with daily living, let alone accelerated change. In preview, the results above suggest three foci: some guides for action; a potential flood of OD applications related to burnout and structure; and, in conclusion, several qualifications on assuming too much.

**Clear Guides for Action**

It seems reasonable to conclude that OD interventions will be more successful if they initially specify differences between targets, including (but not restricted to) their degree of burnout. Such specification is rare, although increasing, but even at this early point it seems obvious that individuals at advanced phases may be poor choices for some kinds of planned change—for example, job enrichment. Such designs require, or at least profit from, a sense of personal efficacy, a quantum of start-up energy, and self-esteem sufficient to buoy a person beyond toleration of unattractive conditions. Those in advanced phases seem to lack precisely such qualities (e.g., Golembiewski and Kim, 1989). To simplify while extending this conceptual network, the results of a complex research design, self-esteem levels of one population covaried regularly with the phases. A few details provide context for this summary. In about 90 percent of the cases, Phase I assignees had lower self-esteem scores than those in Phase II, those in Phase II surpassed the levels associated with Phase III, and so on. In addition, over 30 percent of the paired-comparisons were statistically significant as well as in the expected direction. No cases falling in a contrary direction achieve significance, in contrast.

One need not be tentative in generalizing beyond such studies. The results of tests of over 300 variables similarly imply that individuals in advanced phases will be preoccupied with personal issues. That is to say, advanced burnout is associated with such factors, regularly and robustly (Golembiewski, Boudreau, Munzenrider, and Luo, 1996).

Poorer profiles on self-reports that relate to the worksite: for example, lower satisfaction, higher job tension, lower participation and involvement, and
so on (Burke, Shearer, and Deszca, 1984; Burke and Deszca, 1995; Deckard, Rountree, and Golembiewski, 1986)
General physical symptoms (Burke, Shearer, and Deszca, 1984)
Lower performance appraisals
Generally lower productivity
Higher levels of affective states associated with poor mental health (Deckard, 1985)
Lower cohesiveness
Reduced sense of autonomy and control over work
Greater need for social support, but less of it (Burke and Deszca, 1985)

Due diligence requires attention to the phases, then. Not only do they regulate regularly and robustly with many central covariants, but the phases also suggest useful interventions in ways beyond the other four measures of burnout used here.

Broadly, people in the several phases of burnout might be differentially responsive to the same intervention. Specifically, interaction-centered designs seem more appropriate for those in Phase II, who are having problems with depersonalization. These designs include short-cycle designs such as confrontations or the sharing of three-dimensional images (e.g., Golembiewski 1979, pp. 318–323), as well as more time-extended interventions such as interpersonally oriented team-building (e.g., Dyer, 1987). In contrast, interaction-centered designs seem less appropriate for those in Phases VI, VII, or VIII, at least initially. For them, low-stimulus designs to reduce burnout may have a priority. These designs include flexible work hours (e.g., Golembiewski, Hilles, and Kapiro, 1974), role negotiation (e.g., Harrison, 1972), and job rotation.

It also seems both possible and useful to be more elemental even here. Indeed, the focus on physical symptoms was deliberately chosen to suggest various practicalities—the likely association of burnout with sick days, insurance usage, and perhaps workmen’s compensation. On the present showing, it seems advisable to use the phases of burnout to profile an organization. Such monitoring might well isolate trouble spots—both individual and organizational. In addition, periodic surveys might help anticipate evolving trouble spots by observing changes in the profile of the phases over time.

Some opportunities might lead to success in broader terms, by allowing an organization to react to emergent situations as well as by allowing the organization to proactively anticipate them, for personal and organizational benefit. The required database might be misused, of course, but appropriate models seem to be evolving that would provide safeguards against abuses. The Occupational Safety and Health (OSH) group of Shell Canada, for example, represents one effort to provide sophisticated intelligence about system health while retaining both executive support and employee trust that guarantee good-quality data. This
Preliminary Profile for Burnout

is no place to detail the Shell OSH network but, illustratively, it has its own computer capabilities independent of those available to general management.

A Potential Flood of Applications Related to Structure

More expansively than the profiling discussed above, burnout research clearly implies that it may be possible to deal with syndromes of related variables that derive from different forms of structure about which much is known (e.g., Golembiewski, 1987, 1995). In several ways, that is to say, the discussion above variously implies how OD’s reach-and-grasp can be extended, as in a better fit of designs to burnout status. However, these extensions are only now beginning to be made (e.g., Myers, 2000).

A final point indicates both more challenge and far greater potential in the role of structure in relation to burnout. Virtually the same point applies, in all management arenas.

Structure is an issue in several earlier chapters, and nowhere more explicitly than in Chapter 4, which deals with the reorganization of a medical ward in a hospital. Basically, the intended movement there is from the inherently fragmented bureaucratic model to an integrated or holistic model. The general sense of those arenas is anchored by Figure 19.1, which simplifies somewhat to greater facilitate analysis. That is, the two structural variants there organize activities A, B, and C, whose combination yields some product or service. Figure 19.1A departmentalizes around the several activities, and may be labeled “bureaucratic”; Figure 19.1B organizes around the total flow of work A + B + C.

FIGURE 19.1 (A) Bureaucratic model; (B) flow of work model.
Detailed comparisons of the two variants are conveniently available (e.g., Golembiewski, 1987, 1995), but a useful sense of those consensual products can be suggested by this list of particulars of the probable reactions of individuals in the two structures:

<table>
<thead>
<tr>
<th>Bureaucratic Model</th>
<th>Flow of Work Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less participation in decisions at work</td>
<td>Greater participation in decisions at work</td>
</tr>
<tr>
<td>Less involvement</td>
<td>Greater involvement</td>
</tr>
<tr>
<td>Less autonomy</td>
<td>Greater autonomy</td>
</tr>
<tr>
<td>Less cohesiveness</td>
<td>Greater cohesiveness</td>
</tr>
<tr>
<td>Less satisfaction, etc.</td>
<td>More satisfaction, etc.</td>
</tr>
<tr>
<td>Sub-interest developments that fragment work and identification with it</td>
<td>More holistic developments that integrate work and identification with it</td>
</tr>
<tr>
<td>Less satisfaction</td>
<td>More satisfaction</td>
</tr>
</tbody>
</table>

Illustratively, employee participation is more constrained in Figure 19.1A, where decision-making will either be centralized or characterized by chaotic localism. That is, only \( M_{ABC} \) in Figure 19.1A can make reasonable decisions about the flow of work \( A + B + C \). In Figure 19.1B, in contrast, those in the five roles are in a position to make reasonable decisions about the flow of work.

The evidence for such comparative analysis seems to me quite strong, especially in integration of related research (e.g., Golembiewski, 1987, 1995), which rests on thousands of studies by others. Specifically targeted research on burnout and structure, however, remains both rare and inconclusive in major regards (e.g., Golembiewski and Myers, 2002; Myers, 2000).

To be direct, only beginnings have been made in establishing the direct linkages of structural features with burnout. For example, Myers (2000) has launched a great leap forward, but the two lists above could with equal credibility be labeled “High Phases of Burnout” just as well as “Bureaucratic Model,” and “Low Phases of Burnout” just as well as “Flow of Work Model.” One such summary demonstration is available (Golembiewski, et al., 1996, esp. pp. 217–220). The two lists above of comparisons could be extended, but even as is they support a case of consequentiality. For example, to continue the lists:

**Bureaucratic Model**

- More centralized
- Favor management by direct observation
- Less amenable to “new management” techniques: i.e., job rotation, job enrichment, cross-training, etc.
Longer communication linkages between observation ↔ action in “taller” organizations, etc.

Flow of Work Model

Can be more decentralized
Favors management by objectives
More amenable to “new management” techniques: i.e., job rotation, job enrichment, cross-training, etc.
More direct communication linkages between observation ↔ action in organizations with fewer levels, etc.

This second list also could be lengthened, in turn, but we already risk overkill in support of this basic generalization: under many conditions, bureaucratic structures imply major costs that many organizations would prefer to avoid under many conditions.

So where does this leave us? Analysis has not yet arrived at the definitive point, to be sure, but the conjectural case suggests the strong potential for a flood of OD and QWL planned-change applications with high success rates. To review the argument:

The common consequences of the bureaucratic model are very much the same as the apparent covariants of advanced burnout (e.g., Phases VI–VIII);

The common consequences of the flow-of-work model are very much the same as the apparent covariants of low burnout (e.g., Phases I–III); As several chapters of this volume establish, we know much about the covariants identifying progressive burnout phases;

We also know some useful things about structural changes from bureaucratic → flow of work; and as several chapters above also show, we are beginning to learn how to trigger VIII → I movement on the phase model of burnout. See especially Chapters 4 and 17.

All in all, today’s developments leave us in a pretty good position to launch a great leap forward. And the need is clearly great and growing. Both provide a formidable base of what should be done, and how it can be done, to ameliorate burnout in organizational contexts.

Some Tethers on Enthusiasm

One should be careful not to be overly exuberant, beyond these general expectations. Four points provide more detailed perspective on the present findings, and some of them urge “Whoa” as well as “Go.”

First, although the results imply a substantial covariation of physical symptoms and psychological burnout, as variously measured, present research seldom
specifies the direction of the arrow of causality, so no firm conclusions are appropriate about what causes what. Nonetheless, the associations are marked and consistent, and results to date suggest this pattern of major linkage:

```
"Poor" worksi[497:678]e descriptors -> Increased physical symptoms
Heightened burnout
```

As in this chapter, other studies show a large majority of tests of various measures of this prototypic linkage attain statistical significance. No doubt feedback loops exist, but the case for the mainline effects sketched above seems quite strong (Golembiewski, Munzenrider, and Stevenson, 1986). Thus, sixteen worksite descriptors and burnout phases share approximately 17 percent common variance; various measures of burnout and of physical symptoms share a range of 5–20 percent of their variance, with the highest proportions involving Total Score and Factor I derived from the sixteen symptoms. However, worksite descriptors account for only about 5 percent of the variance in physical symptoms, on average. However, research on the direction of causal arrows has only begun in burnout research (e.g., Deckard, Rountree, and Golembiewski, 1989). Hence this first tether on enthusiasm.

Second, burnout (as variously measured) is not uniformly associated with all symptoms or clusters of symptoms, although all association seem to be statistically significant. Medical authorities report no great surprise about the apparently substantial salience of Factor I, or General Enervation and Agitation, for example. Factor I also substantially contributes to total Symptoms, which has the next most marked pattern of association with burnout as variously measured.

Third, the results provide variable support for using the MBI items to calculate three subscale scores. Tables 19.2 and 19.5 imply that—when separate MBI subscales are used, as in the conventional practice—Emotional Exhaustion reflects the most marked degree of association with both individual symptoms and aggregates of symptoms. For example, Emotional Exhaustion shares nearly 14 percent common variance with the five aggregates of symptoms, on average, or 2.5 to 5 times the magnitudes of the other two subscales.

The MBI Total Score also reflects marked associations with self-reports of symptoms. For example, Total Score shares somewhat over 12 percent variance with the five aggregates of symptoms.

For multiple practical and theoretical reasons, however, the associations of the phases of burnout with self-reported symptoms seem most compelling. Most prominently here, perhaps, Factor I and Total symptoms share approximately 18 percent of their variance in common with the burnout phases. Reinforcing this working conclusion, the present analysis supports other work (e.g., Golembiewski and Munzenrider, 1988) that establishes the regularity and robustness of the co-
variation of burnout with numerous worksite descriptors. In sum, as burnout phases progress, numerous indicators of the perceived quality of working life deteriorate.

The phase model has other basic attractions. Of great potential significance, the phase model implies a variety of practical and theoretical opportunities. For example, targeted interventions—at least at first blush—should differ as Total Scores are differentially loaded by Depersonalization, Personal Accomplishment, or Emotional Exhaustion. The phase model provides one approach to including such differences in subscale components. Inferences drawn from the phase model have been developed elsewhere (Golembiewski et al., 1996), and these deviations at times challenge or even contraindicate major features of the conventional wisdom concerning prominent managerial theories and practice.

A final point concerns the validity of self-reports as data. I will resist the conventional tendency to write: “Of course, the focus on only self-reports requires that results be confirmed with the use of objective and unobtrusive data.” The self-reports here should be cross-checked by a variety of other kinds of data, but evidence indicates that self-reports can lay substantial claims to usefulness. Indeed, for many purposes, self-reports provide the data of choice (e.g., Howard et al., 1980). Much confirming evidence of the usefulness of self-reports exists (e.g., French, Caplan, and Harrison, 1980, pp. 102–103). Hence, the expression “only self-reports” is avoided here.

REFERENCES


One of the common dicta in OD, a kind of working rule-of-thumb, used to be heard frequently: “The O in OD stands for organization.” That point needed emphasis and is still relevant for those pursuing individual interests, as many do via one individual or personality measure or another (such as Myers-Briggs or whatever). Early ODers often had transmuted their interest in T-groups or in sensitivity training into an emphasis on individual emotional experiences, and a general neglect of macro-system properties resulted.

This finesse of “organization” never was as common as many critics alleged, and large-system interventions also were always at the heart of OD, both early (e.g., Marrow, Bowers, and Seashore, 1967) and late, as several chapters in this book show. Readers who have persevered to this point will already have coped with Chapters 12 through 15, which illustrate the reach-and-grasp of large-system interventions in OD.

Whatever this case, it still remains to demonstrate how failure to highlight some crucial theoretical and practical differences between people can be costly in the OD where O = organization. Hence, the focus below on the phase model of burnout, globally. Thus, this chapter augments its predecessor in a signal way. Their combined point is momentous: burnout is worth paying attention to, virtually everywhere. Why? Increases in burnout are associated with more of virtually all of the many bad things in life that have been looked at. In short, burnout is a ubiquitous and costly disease.
A CONCEPTUAL CONCEPT FOR IRONY VII

While respecting the dictum, it still remains the case that in OD one size does not fit all when it comes to individual properties. That point was acknowledged from the earliest days (e.g., Harrison and Oshry, 1967), and the present goal involves demonstrating that at least one individual difference can be important in a broad range of OD situations, globally.

The relevance of burnout to OD is broad, and only a few perspectives will be emphasized here. Thus, individuals with high versus low burnout will profit from different kinds of designs, if not require them. Viewed in another way, levels of burnout may be so high as to require reduction as an early step in a broad OD program, and with delays or necessary cancellations.

In such ways and many others (e.g., Golembiewski and Myers, 2002), differences in burnout can heighten OD effects or contribute to the nuanced evolution of functions such as Human Resource Development.

GLOBAL REPLICATIONS OF PHASE MODEL OF BURNOUT

Do we know enough to justify expansive ambition for burnout as a perspective in OD? This summary chapter will allow readers to judge for themselves, given that by the time of this reading much more will be known.

To provide some sense of this accumulated practical and theoretical knowledge, substantial additions to the panel of burnout studies in global work settings have been accumulated recently, but analysis here shows that they require no modification in principle of earlier interpretations of phase model results. In sum, the phase model seems a valid and reliable operational definition, the consequences of typical covariants of the phases are serious and costly to individuals and organizations, and substantial distributions of the advanced phases of burnout are common in North America as well as globally. Specifically, this analysis adds 30 percent more worksites to the panel underlying earlier interpretations (Golembiewski, 1997). Throughout, major findings and interpretations have remained both stable in essentials and expanding in their details.

EVIDENCE MAINTAINS CONFIDENCE ABOUT VALIDITY*

Although major conceptual issues still remain open (e.g., Leiter, 1999), there seems little doubt that two approaches to burnout are the most capable of accounting for the dynamics seen in nature. Thus, one approach utilizes the three subdomains of the Maslach Burnout Inventory, or MBI (Maslach and Jackson, 1982,

1986), and focuses separately on their covariants as well as on their sequential ordering. This will be referred to as the Maslach-Leiter approach, using the names of its two most prominent North American proponents. The other approach, the phase model of burnout (especially Golembiewski, Boudreau, Munzenrider, and Luo, 1996), proposes a specific ordering of the three domains that differs from the Maslach-Leiter approach, and also provides a specific estimate of what proportion of those tested have which progressive degrees of burnout in an eight-phase model. The phase model also uses items from the Maslach Burnout Inventory, but with revisions of some items from an early form of the MBI.

Here, the focus will be on the second of the phase model approaches, basically to determine whether new information requires a change in the earlier favorable interpretations of the model and its properties. Some time ago, substantial confidence seemed appropriate about the consistent concurrent validity of the phase model, but only fifteen worldwide worksites were available (Golembiewski, 1997, p. 58). Here, the panels contain full data from twenty-seven worksites and as many as thirty-two worksites provide less complete data sets. Both add to the confidence of interpretations and justify the attention here. See Appendix 20. A for details about the full panel. Recent replications come from Belize; Canada, two data sets; Ireland; Israel, two locations; Malaysia; New Zealand, four data aggregates; and Taiwan.

Four emphases dominate this chapter. First, the conceptual and operational definitions of the phase model receive attention. Second, the record of the concurrent validity of the phase model in North American replications is reviewed. Third, the augmented global panel of replications is compared with several other panels of studies. Fourth, a brief but unsettling discussion of the several and serious practical as well as theoretical implications of basically consistent research from North America and from worldwide settings concludes this chapter.

Circumscribing the Phase Model

The phase model differs from all those measures for estimating stress or strain at work that focus on the number and severity of the bundles of stressors to which individuals are exposed. In sharp contrast, the items from the MBI on which that model rests inquire about how individuals experience whatever stressors they encounter, and differences in those experiences can be consequential for even similar bundles of stressors as they impact different individuals.

This phase approach builds on a common observation: the acceptable number and intensity of stressors will differ between individuals at any one point in time, as well as within each individual at different points in time. Hence, the phase model focuses on this question: Are the stressors one now experiences too much, whatever their kind, number, and severity?
The Phase Model

The phase model attempts to build a holistic measure from the items of the Maslach Burnout Inventory, or MBI (Maslach and Jackson, 1982, 1986). To begin, to again risk repetition, a simple paper-and-pencil test permits characterizing each individual by scores on three subdimensions:

- **Depersonalization**, with high scores indicating a marked tendency to think of others as things or objects, to distance self from others;
- **Personal Accomplishment [Reversed]**, with low scores indicating persons who report doing well on a worthwhile job; and
- **Emotional Exhaustion**, with high scores indicating people who are at or beyond their comfortable coping limits, as from inquiring how close each individual is to the “end of the rope” in emotional terms.

In turn, high or low assignments for each individual on each subdomain are made in terms of “universal norms” derived from the median scores of a large population of federal employees operating under a heavy load of stressors (Golembiewski and Munzenrider, 1988, pp. 27–28). A few users of the phase model have relied on different cutting points, usually the medians on the three MBI subdomains for each research population (e.g., Burke, Shearer, and Deszca, 1984). No further attention here goes to such “local” norms. Many details are available elsewhere (Golembiewski et al., 1996).

The phase model assumes for testing this sequencing of the subdomains: depersonalization, then personal accomplishment, and emotional exhaustion. Directly, high emotional exhaustion contributes more to burnout than high personal accomplishment (reversed), while both are larger contributors than high depersonalization.

Simple operations permit developing an eight-phase model of progressively virulent burnout. Table 20.1 provides detail on these simple decision rules. Beyond Table 20.1, the phase model does not propose that any individual will proceed in order through each phase to highest burnout—that is, from Phase I → VIII. Rather, the model proposes that those in Phase I will be characterized

<table>
<thead>
<tr>
<th>Table 20.1 Phase Model of Burnout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phases of burnout</td>
</tr>
<tr>
<td>I  II  III  IV  V  VI  VII  VIII</td>
</tr>
<tr>
<td>Depersonalization</td>
</tr>
<tr>
<td>Personal Accomplishment (reversed)</td>
</tr>
<tr>
<td>Emotional Exhaustion</td>
</tr>
</tbody>
</table>
by fewer deficits or deficiencies in a broad range of marker variables than those in Phase II who, in turn, should be better off than those in Phase III. And so on. In total, there are 28 such paired-comparisons for each variable arrayed by the phases.

Other features of the phase model—for example, that it encompasses both chronic and acute onset—also are useful. But here, a few powerful details get more attention.

CONCURRENT VALIDITY OF PHASE MODEL

Here, the usefulness of the phases will be estimated by their concurrent validity—that is, by how well the phases permit us to predict expected changes in important features of worksites. To illustrate, as the phases progress I → VIII, a useful phase model should show the following (Golembiewski, et al., 1996):

- Job involvement and all facets of job satisfaction decrease.
- Turnover increases, both in intent and in actual departures.
- Symptoms of physical and emotional ill-health increase.
- Many features of family life deteriorate.
- The cost of medical insurance increases.

Illustratively, physical and mental symptoms could at once derive from advanced burnout as well as contribute to it.

The summary data for the newly augmented global panel are presented in Table 20.2, and the relevant worksites are briefly described in Appendix 20.A. Clearly, Table 20.2 indicates that the phase model of burnout applies with substantial validity at all twenty-seven global worksites that replicated tests of the concurrent validity of the phases.

This conclusion does not have to search for support in Table 20.2. Thus, overall, very nearly nine of every ten marker variables contain nonrandom variance when arrayed by the phases. Note also that the actions do not require any qualifications about interpretations. Just the opposite is the case, in fact, perhaps in large part due to the large $N$s in these additions, which range from 281 to 2771.

This overall pattern gets strong and detailed support from the paired-comparisons of markers $\times$ phases, as Table 20.2 also shows. In sum, a very large majority of paired-comparisons fall in the intended direction—for example, Tension at Work is greater for those paired-comparisons than chance allows to attain statistical significance as well as fall in the expected direction. This is very powerful evidence. Of course, 5 percent could be expected to attain significance by chance alone, and about five times that proportion does so in Table 20.2, on average. Finally, Table 20.2 reflects a minority of paired-comparisons falling in a contrary direction, and almost none of those deviant cases achieve significance.
TABLE 20.2 Summary, Marker Variables × Phases, in 27 Global Work Settings, N = 9714*

<table>
<thead>
<tr>
<th>Marker variables × phases, statistically significant by ANOVA</th>
<th>Paired comparisons, in %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In expected direction and statistically significant</td>
</tr>
<tr>
<td>A. Original Cohort (N = 22) 111 of 130 runs, or 85.4 percent</td>
<td>78.7 23.9 21.5 1.3</td>
</tr>
<tr>
<td>B. Additions to Cohort (N = 5) 24 of 24 runs, or 100 percent</td>
<td>84.9 40.3 15.3 0.0</td>
</tr>
</tbody>
</table>

* Sites are briefly described in Appendix 20.A. The five most recent “additions” are Canada B; New Zealand B, C, D; and Taiwan C.

Of course, perfect support for the phase model would isolate an even-sharper pattern: 100 percent of all marker variables would achieve $P < .05$; all paired-comparisons would attain significance; and no paired-comparisons would fall in a contrary direction. Table 20.2 does not achieve this perfect status but clearly approaches it. This clear pattern for the global panel in Table 20.2 does not differ appreciably from the record for thirteen U.S. work settings investigating the covariation of similar marker variables with the phases. Table 20.3 provides summary details.

When Table 20.2 is compared with Table 20.3, only modest support is found for the common view that cultural relativity dominates over generalizable findings. No doubt, it may be a bit much to conclude that the phases isolate a universal pattern of covariation. At least, Tables 20.2 and 20.3 do not load the dice. To explain, two of the global populations—Israel B and Malaysian A—were tested only to see whether quite small populations would isolate the expected pattern of covariation. Table 20.2 in effect provides a conservative estimate of concurrent validity.

INCIDENCES OF THE PHASES

Since the phases isolate much the same pattern of covariants in 40 relevant populations in global and U.S. panels, the phases reflect a pattern that is not basically
TABLE 20.3  Summary, Marker Variables \(\times\) Phases, in 13 U.S. Work Settings, \(N = 11,215\)

| Marker variables \(\times\) phases, statistically significant by ANOVA | Paired comparisons, in % |
|---|---|---|---|
| In expected direction and statistically significant | In contrary direction and statistically significant |
| 102 of 111 runs, or 91.9 percent | 82.4 | 41.1 | 17.6 | 0.49 |

Source: From Golembiewski, 1997, p. 49. The U.S. panel is described in that source, p. 58, and includes only loci where marker variables were tested.

distorted by organizational or broad cultures. Several comparisons below provide quite precise magnitudes of such generalizations.

To begin, Table 20.4 indicates a worrisome distribution of the phases, with one-half of all respondents in global work settings being classified in the three most advanced phases, both subtypes A and B. Row B reflects a somewhat health-

TABLE 20.4  Summary, Incidence of Phases of Burnout, 31 Global Work Settings, \(N = 15,863\)^{a}

A. Original Cohort (22 cases, \(N = 10,190\))

| Phases of burnout, in % |
|---|---|---|---|---|---|---|---|
| I | II | III | IV | V | VI | VII | VIII |
| 11.9 | 6.0 | 10.1 | 12.3 | 5.7 | 12.5 | 4.7 | 36.9 |

| 28.0 | 54.1 |

B. Additions to Cohort (9 cases, \(N = 5673\))

| Phases of burnout, in % |
|---|---|---|---|---|---|---|---|
| I | II | III | IV | V | VI | VII | VIII |
| 20.9 | 7.6 | 8.2 | 7.4 | 8.0 | 14.9 | 5.5 | 27.7 |

| 36.7 | 48.1 |

^{a} These additions include the five mentioned near Table 20.1 and are described in Appendix 20.A, for which marker variables were tested. The additions also include four studies that provide data only about the phases of burnout: Italy A, Korea A, Poland A, and Yugoslavia A. See Golembiewski, Boudreau, Munzenrider, and Luo, 1996, pp. 169–170.
ier pattern, but that is due to the inclusion of three New Zealand worksites. Here, as everywhere (e.g., Cooper and Hensman, 1985), New Zealand seems an outlier—for example, counts 44 percent in Phases VI–VIII, while Taiwan approximates 59 percent.

Given the deficits and deficiencies of market variables associated with the phases as they progress $I \rightarrow VIII$, the distributions of incidence suggest many troubled worksites. For example, individuals in advanced phases consistently report symptoms of declining physical and emotional health (e.g., Golembiewski, Boudreau, Munzenrider, and Luo, 1996, pp. 83–128). As a benchmark, those in Phase VIII report 3–10 times more of nineteen physical symptoms than persons in Phase I, who, in turn, report only a small fraction of the symptoms reported by all respondents in two national surveys.

In contrast, data from a large panel of U.S. respondents, summarized in Table 20.5, reflect a bit more favorable distribution of phases than Table 20.4; 43.3 versus 54.1 percent of the assignments in U.S. worksettings belong to phases VI through VIII. If discounted some, however, the costs of burnout in U.S. worksettings still seem substantial. For example, consistent with expectations, people assigned to phase VIII are the significantly largest users of medical insurance, with other pairwise-comparisons falling in line (e.g., Golembiewski, Boudreau, Munzenrider, and Luo, 1996, pp. 145–151). The specific cost/benefit calculations have not yet been made, but the case for phase effects as serious and costly seems a reasonable premise of management and research decision-making.

Conveniently, a panel of Canadian worksettings permits a check on the comparisons implicit in Tables 20.4 and 20.5. Too fine a point should not be put on Table 20.6, but it does not require any modification in principle of the comparisons detailed above. Obviously, the Canadian panel is smaller than the other two panels, but the distribution of phases consistently suggests a bimodal versus normal distribution. A reasonable theoretical rationale supports this effect (Golembiewski, Boudreau, Munzenrider, and Luo, 1996, pp. 55–57), and virtually all observers would anticipate the similarity of the U.S. and Canadian panels.

### Table 20.5
Summary, Incidence of Phases of Burnout, 62 U.S. Work Settings, $N = 24,721$

<table>
<thead>
<tr>
<th>Phases of burnout, in %</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.4</td>
<td>5.7</td>
<td>11.4</td>
<td>9.0</td>
<td>6.2</td>
<td>14.2</td>
<td>8.5</td>
<td>20.6</td>
</tr>
<tr>
<td></td>
<td>41.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>43.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 20.6 Summary, Incidence of Phases of Burnout, 11 Canadian Work Settings, N = 3,240

<table>
<thead>
<tr>
<th>Phases of burnout, in %</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.0</td>
<td>6.0</td>
<td>12.8</td>
<td>8.4</td>
<td>6.8</td>
<td>12.8</td>
<td>7.9</td>
<td>20.2</td>
</tr>
<tr>
<td></td>
<td>43.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40.9</td>
</tr>
</tbody>
</table>


PERSPECTIVES ON EMPIRICAL RESEARCH IN PROGRESS

It appears to be steady as it goes, then, when it comes to interpreting the phase model of burnout. Substantial additions of the global panel do not require any modifications in principle of earlier interpretations of phase model results. If anything, indeed, the augmented evidence implies a heightened practical and theoretical urgency to act on what the phase model tells us about the world of work, or pay the costs.

The implications of the phase model—for general management (Golembiewski, 1997, esp. pp. 53–57) as well as for Human Resources (Golembiewski and Myers, 2002)—have been developed elsewhere in some detail, so only a few summary points will be emphasized here.

1. The phase model seems substantially reliable and valid, as judged by its consistent record in replications in isolating a clear pattern of covariants. The consistent patterns of covariants in the United States and in a substantial number of overseas work settings does not encourage a contrarian view based on large social, cultural, or organizational differences. Such differences certainly exist, but they do not seem to account for large proportions of the variance in phase findings.

2. Factor analysis implies the consistency in terms of both pattern and magnitude of the structures generated by respondents to the items topping the three MBI subdomains. Such evidence provides no major support for a culturally relative approach.

3. The phase model suggests that the costs of burnout are high, both socially and organizationally. The catalog of covariants of the phases is both regular and robust, and substantially so, as Tables 20.2 and 20.3 reflect.

4. The levels of burnout seem “high,” even under the “good” conditions that seem to underlie research on the phase model in New Zealand or in specific North American organizations. For example, no investigator has yet isolated in a substantial population less than 22 percent of respondents in Phases VI through VIII (Golembiewski et al., 1996, pp. 159–170). Uncomfortably, indeed, one can reasonably argue that mostly “better” organizations will allow the testing on which the phase model rests.
5. “Local” norms generate results and incidences similar to those associated with “cosmopolitan” norms, but with “universal” norms isolating stronger and more consistent patterns.

6. The complexities get no attention here, but there seem convenient designs for ameliorating burnout. Several designs have been replicated with consistent and expected effects (e.g., Golembiewski et al., 1996, pp. 222–228).

7. The concurrent variables tested in most research with the phase model are still dominantly self-reports, but “objective” or “hard” variables appear with growing frequency. Phase associations seem similar for all variables.

8. Perhaps the major gap in the research supporting the phase model is the general lack of specific cost/benefit analyses of advanced phases of burnout. Available work suggests very high costs (Jackson and Manning, 1995) but does not establish it. For example, those in advanced phases report symptoms of physical and emotional distress 5–10 times more frequently than those in Phase I; and, correspondingly, those in Phase VIII report the greatest reliance on health insurance, by far.

APPENDIX 20.A DESCRIPTIONS OF GLOBAL WORKSITES TESTING FOR COVARIANTS OF THE PHASES*

1–6. Belarus A–F Six collections of employees, many from industrial settings. \( N = 312, 792, 200, 474, 93, \) and 197, respectively. Response rates vary broadly (Boudreau and Levin, 1996; Boudreau, Zhilina, and Faiferman, 1996).

7. Belize A A convenience population from three organizations, \( N = 261, \) in white collar and professional settings (Aldinger and Golembiewski, 1999).

8. Canada A \( N = 130, \) a composite of three convenience populations (Duplcea and Hubert, 1996).

9. Canada B \( N = 678, \) all registered nurses from a western Canadian province (Barsky, 1999).

10. China A \( N = 196 \) middle managers and above assembled for a conference (Rowney and Cahoon, 1987).

11. China B \( N = 259 \) from 10 departments of Beijing municipal government, with about 60 percent in operating roles (Golembiewski and Luo, 1996).

12. Ghana A \( N = 287, \) from three Accra hospitals, all employees (Fiadzo, Golembiewski, Luo, Bradbury, and Rivera, 1995).

* All global worksites dated 1996 or earlier also are further detailed in Golembiewski (1997), p. 58.
<table>
<thead>
<tr>
<th></th>
<th>Country/Region</th>
<th>Sample Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.</td>
<td>Ireland A</td>
<td>$N = 185$, an aggregate from four diverse educational settings (Coghan and Golembiewski, 1999).</td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Israel A</td>
<td>$N = 100$, all from two departments in large public metropolitan hospital (Pines-Malach, 1999), all females.</td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Japan A</td>
<td>$N = 981$ respondents from six health-care locations. All jobs represented, except doctors (Boudreau and Golembiewski, 1989).</td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Japan B</td>
<td>$N = 498$ from several comprehensive health-care centers. All jobs represented, except doctors (Boudreau and Golembiewski, 1990).</td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Malaysia A</td>
<td>$N = 57$, a mixed population from public and business loci (Sinoway, Baxter, and Santa Maria, 1999).</td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>New Zealand A</td>
<td>$N = 194$, composite of employees from three organizations, 70 percent female (Boudreau, Everett, and Golembiewski, 1999).</td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>New Zealand B</td>
<td>$N = 281$ police, 78 percent male, from Otago and Southland Districts, with 13.5 mean years of service (Boudreau, 1999).</td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>New Zealand C</td>
<td>$N = 509$ police, 78 percent male, Canterbury and Tasman Districts, with average age of 39 with 13.5 mean years of service (Boudreau, 1999).</td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>New Zealand D</td>
<td>$N = 1082$, composite population but with police dominating (Boudreau, 1999); involves double counting with worksites 20–22.</td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>Saudi Arabia A</td>
<td>$N = 264$ professional trainers and support personnel at Institute of Public Administration, Riyadh (Al-Ebedah, 1995).</td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>Taiwan A</td>
<td>$N = 623$ from 7 types of R.O.C. police, with about 16 percent in “inside” or office jobs. All ranks of police are represented (Golembiewski, Sun, Lin, and Boudreau, 1995).</td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>Taiwan B</td>
<td>$N = 553$ street-level urban bureaucrats from all but the three highest of twelve hierarchical grades (Lin, Sun, and Golembiewski, 1996).</td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>Taiwan C</td>
<td>$N = 2771$, over 62 percent male, from twenty departments in Taipei municipal government (Huang and Sun, 1999).</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


Golembiewski, R. T., and Myers, T. G. (2002). Phase model of burnout as leverage for meeting “future forces challenging HRD”. In A. Farazmand (Ed.), *Human resources development and planning*. New York: Praeger, in press.


Enlarging the Empowering Potential of the Workweek
Flexible Work Hours as Exemplar

The concepts “workweek” or “workday” are not defined by a divine act: they are circumscribed by social and cultural practices, the cumulative acts of men and women. For a long time for most observers, and perhaps still so for some, the notion of a workweek seemed quite definite: Work started at 8:00 AM or so, permitted 30–60 minutes for lunch, and then ended something like 480 minutes after it had begun daily.

This seems a quaint notion nowadays. For many “knowledge workers,” for example, much less certainty is possible about when “work” begins and when it ends. And there certainly is nothing divinely inspired about 8 AM to 5 PM, even if we neglect the incidence of shift work.

The focus here is on a single and simple exemplar of what can be done to make managerial concepts more responsive to human rhythms and needs, and with attractive cost/benefit ratios. The focus is on flexible work hours—henceforth, FWH—which presents a kind of mixed record, if not a schizoid one. FWH is at once an “easy piece” to use and has a broad range of positive consequences. Prominent among these effects are facilitated childcare for one-parent families (e.g., Ezra and Deckman, 1996). But the reports about FWH applications cover a very broad range: thus one report (Government Finance Review, 1998, p. 68) characterizes FWH as “a rapidly growing trend” and another reports that “few employees are taking advantage of it” (Federal Times, 1997b, p. 1), despite its easy attractions.
This chapter seeks to reconcile the bouquets with the thorns, to put the mixed state of FWH affairs in flowery language. And two purposes dominate. Some useful contributions to greater responsible freedom are achievable by enriching the concept of “workweek” via flexible work hours; and almost certainly far more can be done at most worksites, with only a little extra wit and will.

LIMITED-PURPOSE DESIGNS AS GENERIC

So which prescription is correct? Is it the prescription that counsels the lofty objective, but really implies pie-in-the sky, by-and-by? It lets our aspirations soar but risks being impractical and sets objectives so high as to assure disappointment. Or do we accord priority to the “next reasonable bite” prescription? It proposes that we start from the client’s present state—dismal and primitive though that may be, and even though that risks a fixation on Band-Aids, if one is careless.

Typically, of course, OD literati answer “Yes” to the prescription for basic cultural change, dangers and pie-in-the sky notwithstanding. And that is a reasonable response, even generally commendable, and much of Ironies II reflects it.

My heart is with those who counsel, “Yes, go for it,” but my experience urges stout tethers on exuberance.

The position here is that both prescriptions are appropriate, but in a complex way. That is, both prescriptions probably will be appropriate in the life of any organization, but not at the same time. Moreover, neither one of the prescriptions is likely to be useful at the same time for all parts of any organization. Further, and especially for openers, I suspect that most organizations are relatively poor targets for the comprehensive social contract implied by pervasive cultural change. For most organizations, that is, limited-purpose designs will have a greater applicability, ab initio. In sum, limited-purpose designs, in many cases, will facilitate incrementally expansive movement toward OD ideals, and this progress at some future date may provide a solid foundation for more pervasive interventions.

Hence the centrality of this chapter and the pair following it. Essentially, each illustrates a limited-purpose design but at the same time can be convenient ways to begin the bolder journeys to organization operating under comprehensive OD contacts.

A CONCEPTUAL CONTEXT FOR IRONY VIII

Irony VIII deals with three easy pieces that illustrate limited-purpose contracts. The focus here is on flexible work hours; Chapter 22 adds emphasis to expanding the concept of the worksite by considering Flexi-Place; and Chapter 23 seeks to expand the concept of development.
Flexible Work Hours as Empowering Potential

The common intent? These easy pieces have success rates that seem very high, even formidable, and they will help lead organizations incrementally down the road to the hard piece—basic cultural change.

The chapter will deal with an easy piece—“easy” in the sense of convenient, simple, requiring little training, and with very high success rates—but it deliberately chooses to focus on the worst case of the flexible work hours genre. The focus will be on the public sector, which is widely said to be more recalcitrant in the extreme to efforts at planned change via OD, although readers of Chapter 1 will know better. In addition, the focus will be on productivity effects concerning which, again, the public sector is widely viewed as being generally deficient, if not broadly unconcerned.

Programs of flexible work hours, or FWH, have proliferated around the world, and most observers generally associate them with a broad range of positive effects for organizations as well as for their members, in personal life as well as at work (e.g., Nollen, 1982). FWH may rightly be said to be a major managerial innovation, simple though it seems.

The innovation initially permeated business organizations, and later got more public-sector notice. Paradoxically, the diffusion of this innovation had early proceeded apace, in both arenas, without two crucial kinds of knowledge—a comprehensive review of the available FWH literature and a sophisticated understanding of why, when, and how FWH works.

More recent years have seen a decided improvement in the first particular, especially around the turn of the new millennium. Specifically, several comprehensive summary studies (Golembiewski and Proehl, 1978, 1980; Nollen 1979; Ronen, 1981; Ronen, Primps, and Cloonan, 1978) indicate that FWH applications have been associated with a variety of favorable effects in public as well as in business organizations. These summary studies deal with hundreds of separate applications, and the FWH applications cover some 30-plus years. So any conclusions about FWH can lay substantial claim to relevance and generality.

The second crucial gap in our knowledge—a sophisticated understanding of why, when, and how FWH works—has proved far more resistant, and reducing that gap a bit preoccupies this chapter. It considers sixteen public-sector FWH applications, with the specific goal of illustrating how general concerns about public-sector OD applications are substantially met by one intervention. Specifically, OD applications in the public sector often have been seen as constrained by three factors:

Confusion has been expressed concerning how the behavioral theories underlying OD apply in government, given that their basic development has been rooted largely in business or industrial contexts.

Unique, or at least especially intense, institutional and historical constraints
are said by many to make OD interventions more difficult in the public sector, if not impossible.

The effects of public-sector interventions are more difficult to judge, since “hard data” are seldom available and because “soft data” relating to changes in attitudes and opinions are said by some to pose crippling problems of reliability and validity.

The discussion below will show how these three factors apply only in diminished senses, if at all, to the OD intervention called FWH.

Specifically, the FWH concept will be introduced as an “easy piece” in extending OD’s reach-and-grasp. As will be shown, FWH constitutes a kind of medium for OD’s message. Thus, building regenerative systems of interaction helps channel FWH applications, thereby reducing the probability that license would result as opposed to increases in responsible freedom at work. In addition, FWH is both a palliative of some common effects of bureaucratic life, as well as a useful precursor of explicit and extensive post-bureaucratic infrastructures. Finally, FWH exemplifies the kind of policies and structures that will have to be developed to reinforce interaction consistent with OD approaches and values. The genre is large, even huge, as Chapters 22 and 23 will help illustrate.

But let us keep our conceptual focus on FWH. In this chapter, special attention will be paid to two deep test borings into the literature and practice: roughly during the decade centering on 1975–1980 as well as the decade preceding the millennium. The focus below is on FWH only, among the broader family of designs often imprecisely lumped together as “alternative work schedules,” or AWS.

Other disconcerting issues will complicate and obfuscate this chapter but, nonetheless, the effort has major redeeming features. First, the cost/benefit ratio of FWH applications seems quite attractive. The big payoff in this first sense should motivate more FWH applications.

Of special significance, both practical and social, FWH also seems associated with improved effects in families, and especially for single parents. The findings are reasonable. Thus, mothers report FWH eases the problems of managing the often-conflating demands of work and family (e.g., Hochschild, 1983). Relatedly, employees under FWH report having more time to spend with their families, and about 96 percent of them report greater satisfaction with their FWH schedules (e.g., Bohen and Viveros-Long, 1981, p. 147). In general, greater use of FWH also is directly associated with a range of positive effects at work (e.g., Ezra and Deckman, 1996, p. 177).

Second, despite its generally attractive consequences, FWH applications have not dominated management thought and action. Much room exists for new applications. Of all full-time wage and salary earners, AWS were enjoyed by a growing proportion over time (Federal Times, 1997a):
1985—12.5 percent  
May 1991—15 percent  
May 1997—more than 27 percent

Other reports feature the same accelerating profile, but with greater percentages of participation (e.g., Ezra and Deckman, 1996, p. 175).

Third, however, observers often fail to distinguish FWH from the much-broader AWS family, including work weeks of four 10-hour days. The curious conflation is consequential. If nothing else, AWS variants do not seem peas in a pod. As noted, the focus below is on FWH only, soon to be defined in detail. Consequently, readers will not have to buy into the assumption of the casual bundling of undifferentiated AWS variants (Government Finance Review, 1995, p. 68). The focus here is on FWH only. Other AWS (alternative work schedules) variants will require separate treatment—by somebody else, in some other place.

Fourth, the analysis below ties FWH directly to a technology for change with values. There is no telling what some AWS, or even FWH, results mean in the absence of clarity about how greater responsible freedom is approached in specific cases with regard to hours at work.

Fifth, FWH can be considered in the policies/procedures category of OD and, whatever success rates are reported for FWH, they should be enhanced by the two other OD domains—reinforcing forms of interaction and types of structures. What has been written about FWH suggests the centrality of interaction improvements, but prework on appropriate changes in interaction is rare. People may have fewer opportunities to meet casually when they are flexing, for example, and here regenerative interaction can be a valuable FWH adjunct in specifying supportive skills and values. Relatedly, the functional model associated with bureaucratic structures can result in problems with which FWH programs must deal, as hosts of innovation at some possible sites of application have concluded (e.g., Dalton and Mesch, 1990, p. 583). A basic issue here is that bureaucratic structures inhibit cross-training, which obviously would facilitate responding to variable patterns of employee arrivals and departures under FWH. Specifically, structures like the flow-of-work model considered at several points above and elsewhere (e.g., Golembiewski, 1995, 1996) would facilitate cross-training and hence probably would increase FWH success rates.

The relevant literature regarding structure is still in its early stages (e.g., Myers, 2000), but the prognosis seems hopeful. For example, the change from classic work teams to autonomous groups seems associated with decreases in burnout (e.g., Novelli, 1989). And although small Ns create interpretive problems, bureaucratic sites and structures seem associated with advanced burnout (Myers, 2000). Hence, the speculative-if-informed case arguing that position seems generally convincing (e.g., Golembiewski, 1996), if incomplete.

Sixth, conceptually, FWH may be viewed as a way of increasing responsi-
ble employee freedom and discretion at work, which has the advantages of being inexpensive while encouraging easy new skills and training. FWH thus represents an attractive intervention.

Conceptually, FWH and its kin are consistent with the “organization psychology” that underlies OD. The reader can easily develop the numerous senses in which FWH fits with various prominent behavioral models: Maslow’s pyramid of needs; McGregor’s Theory Y vs. X; Argyris’ dimensions for self-actualization; and Herzberg’s distinction between motivators and hygiene factors, among many other possible and variously precise variations on substantially common conceptual themes. For example, FWH permits movement from dependence to growing independence, in Argyrian terms. And FWH similarly enhances the work itself as well as the employee’s sense of personal responsibility, which most readers will recognize as Herzbergian motivators.

Although FWH is no longer a hot item in the literature, it has been disseminated quite widely in both public and business organizations. FWH seems to have encountered notable resistance only in the federal government, which studied it with great fanfare but stopped short of systemwide adoption. Up-to-date statistics are not reliable, but perhaps a quarter of all worksites have some FWH options.

FWH applications tend to have positive effects, on balance—indeed, very positive effects, as will be shown. Its relative fall from grace in recent days no doubt has a multifaceted explanation, but a major contributor certainly has been the cutbacks and downsizing characteristic of our immediate past. In sum, these now-passing days were not employee-friendly. In contrast, our contemporary days—featuring selected manpower shortfalls if not deficits in available recruits—may very well apply renewed pressure for a new increase in FWH adoptions in our immediate future.

How, then, will this chapter go about establishing that FWH permits a reasonable approach to generating greater responsible freedom in organizations? Attention immediately is paid to several general FWH issues and then moves on to a survey of public-sector applications, whose success is estimated by hard criteria, and then a detailed view of costs/benefits of FWH applications concludes this chapter.

GENERAL ISSUES AND FWH

Six sections constitute the immediate point of departure. Consideration first goes to the basic FWH model.

Basic FWH Format

Although it has appeared in diverse forms, the fundamental FWH model is quite simple. Schematically, the FWH day takes a form such as the one shown in Figure
Flexible Work Hours as Empowering Potential

21.1. Every employee works the “core hours” each day of the normal five-day workweek. Both arrival and departure times can vary according to employee need or even whim, during the several flexible hours at the beginning and end of each day. Some plans permit banking of hours worked over some accounting period—say, two weeks. If applicable legislation permits, for example, an individual on a thirty-five hour workweek might accumulate thirty-nine hours in the first week and bank the surplus for the second week. The next week that individual would work only thirty-one hours, as by taking off early on Thursday and Friday afternoons—working, let us say, five hours on Thursday and Friday rather than seven hours each day.

One point should be clear already. FWH is definitely not like the 4-by-10 plan, or a “compressed workweek,” which requires ten hours a day for four days. Hence, not every alternative work hours plan is FWH, but many variations exist in specific plans concerning the length of core hours, how many hours can be banked to reduce work hours, what conditions constrain employee choice, and so on. Time-recording devices are also often used, but not always. For some purposes, these differences can be critical (e.g., Harvey and Luthans, 1979). But they get no further attention here.

What else can be said about FWH, as a useful but general introduction? Five more emphases will help, although they clearly fall short of being comprehensive.

Why the Confusion About What “It” Is?
The FWH model seems straightforward, so why the general confusion with AWS? All too often, FWH often gets lumped together with other “alternative work schedules”—e.g., compressed workweeks—having little in common with FWH, as depicted here. This appears to have been a tactic in writing legislation, for example, which may assemble political support but at the cost of combining apples and oranges. Both effects have been observed. See Public Law 95-390, Federal Employees Flexible and Compressed Work Schedule Act. We will be more strict here.
“Steady As She Goes”
Dominantly, perhaps, the FWH literature is upbeat. Halston’s work (1989) is representative of the more methodologically aware FWH studies. Overall, he concludes, his results with two state agencies “are similar to . . . studies in which no statistical analyses were performed,” and they reinforce the “positive halo effect” of the earlier literature (p. 372).

Conditions of Application
Attractively, the baseline of FWH success seems both substantial and subject to increase by modest efforts. Basically, most recent research emphasizes the situational conditions that heighten FWH success, and in this major sense reinforce the use of FWH as a way to extend OD values as established, for example, by a successful interaction-centered design. So, Flexi-Time is not an application that “always works.” Distrust in labor/management histories is an important situational feature that influences success/failure (e.g., Chavez, 1997), to illustrate, and most successful applications will need to deal with this past history by forging new collaborative relationships. Such conflictful histories have complicated not only specific applications, but also bedevil legislation at both federal and state levels. Broadly, McCampbell (1996) provides a long list of such positive situational features including:

The support of top management;
The isolation of jobs that can be included without jeopardizing coverage and service;
Trust between employers and supervisors;
Training for both employees as well as supervisors; and
Integration with strategies—missions, goals, and objectives.

Perhaps above all, FWH is a dynamic relationship between different levels and skills (e.g., Peak, 1994)—a technique in a value-loaded context. So nothing is automatic, although FWH attraction in general seems good enough to motivate a sufficient balance of positive forces.

And Rigorous Research!
More recent FWH research can be labeled more “rigorous” (e.g., Dalton and Mesch, 1996), but a large agenda remains. In part, the widely acknowledged positive consequences of FWH act both to encourage and discourage such disciplined effort. Perhaps, best said, researchers have not yet validated the broad network of variables associated with FWH success/failure but an attractive prognosis seems generally reasonable. For example, the growing proportion of single-
Flexible Work Hours as Empowering Potential

parent families, among other features, implies a great and growing need for family-friendly designs like FWH.

Constraints on FWH Applications, Especially in the Public Sector

There has been an appreciable time lag in public-sector applications of FWH, finally, and this suggests some general public-sector constraints against change. This lag could be established in many ways. But here consider only that FWH applications spread quite widely in Europe by the early 1960s, and the first major longitudinal research study in American business appeared in the early 1970s (Golembiewski, Hilles, and Kango, 1974), but it was not until the late 1970s that the major public-administration journal carried any extensive mention (Rubin, 1979; Golembiewski and Proehl, 1980) of what many consider a major managerial innovation. FWH is even now spotty in local government, even after a multi-year and positive study at the federal level.

Beyond this general time lag, however, FWH appears not to have encountered any special resistances in public-sector applications. The results in business (Golembiewski and Proehl, 1978; Glueck, 1979; Nollen, 1979) seem quite comparable to the public-sector experience (Golembiewski and Proehl, 1980). Moreover, applications in both sectors have similar rates of failure—very, very, low. Finally, prevailing institutions and practices in public personnel administration do not present anywhere near the obstacles to FWH that they pose for other kinds of OD designs. Interaction-centered OD has experienced some resistance, sometimes of a political character (e.g., Warwick, 1975, pp. 59–83), and sometimes because of concerns that “democratic administration” may undercut popular control (Mosher, 1968, p. 374). OD focusing on structure presents a mixed picture. On one hand, autocratic bureaucracy is deeply entrenched in the ideation and practice of public personnel administration (Golembiewski, 1984, 1985)—in structure, position classification, job design, and related manifestations that were originally intended to isolate administration from politics and that have been in place for over a century. This can inhibit structurally focused OD. On the other hand, the public-sector arena is pervasively procedural, and that implies a recognition of (and at times a receptivity to) structural or policy interventions (e.g., Golembiewski, 1962a; Rainey and Rainey, 1986) as well as to policy-and-procedural interventions like FWH.

TWO INTERVALS OF APPLICATION

These comments may appear to be general, and they no doubt create a sense of redundancy. This part of the text seeks to overcome the former by rooting discus-

FWH 1990–1996

No effort will be made here to review this earliest FWH literature. That has been done elsewhere, and at substantial length (e.g., Golembiewski and Proehl, 1980; Golembiewski, 1990, pp. 183–190). Almost everybody loves FWH, and its applications are associated with positive managerially related outcomes. No more will be said here about this early literature, beyond the comments below. See also Exhibit 21.1, and especially cases 1–16.

Basically, this intermediate literature shows only a handful of failures. The intervention seems to provide what most individuals seem to want and what the managements of most organizations are willing to share—a greater potential to control aspects of their worksite, even though the actual exercise of that potential may be infrequent (e.g., Golembiewski, Hilles, and Kajno, 1974). In exchange, the overwhelming experience is that this freedom is used responsibly. For example, indicators of performance reflect either an enhancement, or at least no diminution, of performance (e.g., Golembiewski and Proehl, 1978).

So it appears safe to say that FWH applications—wherever—seem to be relatively easy pieces that can sustain lofty success rates. That bottom line is made even more attractive by several features of the intervention. Thus, the design is broadly applicable, perhaps especially for initial OD interventions in organizations with little experience in participative management. Moreover, modest skills are required to make a go of it, and training costs are minimal. In addition, both management and labor unions (where they exist) may find grounds for common cause in FWH applications, with possible benefits that generalize far beyond the immediate intervention.

The case for enhancing OD success rates also can be made in terms of the preceding analysis, if but briefly and selectively. For example, FWH seems a reasonable low-stimulus design, with attractive uses in general but especially when individuals report advanced degrees of burnout. FWH can be part of a first-stage effort to reduce strain on employees, in short, with the second stage being devoted to later systemic change when individuals develop greater emotional slack. Relatedly, FWH at once rests on and contributes to the kind of group properties associated with low burnout—Supervisory Support, Peer Support, and Task Orientation. Typically, to illustrate, work units have to develop suddenly shifting ways of covering for an employee who is flexing, and that happens most fluidly when Peer Support is high.

More generally, FWH fits best with the regenerative interaction systems at the heart of OD—high openness, owning, and trust, as well as low risk. Indeed, one of the leading applications was presented in terms of not only resting in trust.
Flexible Work Hours as Empowering Potential

but also as contributing to it as well as to a developing OD program (Golembiewski, Hilles, and Kagno, 1974). In the absence of effective and nondefensive communication, in general, FWH applications could quite easily degenerate into misuses of the new discretion. It seems to be a measure of the attractiveness of FWH as an intervention that such outcomes are infrequent, if not rare.

To be sure, the intervention does not come scot-free, but the costs seem generally bearable and might even be attractive to assume. On general grounds, consider only two of these costs—the first relating to a practical issue and the other involving methodology.

Many practical problems can exist in FWH applications. For example, in one FWH installation, a small percentage of supervisors still had a problem with the program after four years. These few supervisors—because they were deeply distrustful, or could not bring themselves to delegate, or could not monitor performance by developing shared objectives—worked the entire FWH day, which in their case was 10.5 hours. This surprising finding in an evaluation on otherwise-positive FWH experience inspired some training costs, but costs attractive both to most supervisors involved and to their organization (Golembiewski, Fox, and Proehl, 1979).

The methodological issues are more intractable, and hence (if anything) stand in greater need of attention. As the summary above indicates, FWH studies are relatively numerous but will win no prizes for rigor of design or method. It will require a real avalanche of sophisticated studies with zero or negative effects to counterbalance the positive tenor of the FWH literature and, although that avalanche may come, it seems unlikely. We also lack an appreciation of the specific dynamics involved in FWH applications, and that needs early remedying. Models for the required work are available (e.g., Orpen, 1981; Ralston, Anthony, and Gustafson, 1985), but the general attractiveness of FWH applications has yet to be sufficiently complemented by detailed study. Paradoxically, indeed, that very attractiveness no doubt discourages rigorous research about dynamics that seem so obvious on their face.

Extending Analysis Circa 2000

No satisfactory update of the early FWH literature has appeared, but several general points from a search for applications around the millennium seem to usefully update the view above. See Exhibit 21.1, and especially cases 17–19. First, the rate of FWH adoptions seems to have decreased, and perhaps the best estimate is that diffusion has extended to perhaps a quarter or even a third of relevant sites. No doubt the largest possible extensions followed three years of “study” in the U.S. federal government. Broad authorization did not result, however, and the main result seems to have been adoptions on their own by many federal agencies. FWH also has gained enough visibility to be a part of the
<table>
<thead>
<tr>
<th></th>
<th>Statistical treatment?</th>
<th>Hard/soft data?</th>
<th>Effects on “hard” productivity indicators</th>
<th>Number</th>
<th>Unionized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>None</td>
<td>H/S</td>
<td>Increases of 10% in “successful contacts made on first call from outside”</td>
<td>100</td>
<td>?</td>
</tr>
<tr>
<td>2.</td>
<td>?</td>
<td>H/S</td>
<td>“Inconclusive results”</td>
<td>200</td>
<td>?</td>
</tr>
<tr>
<td>3.</td>
<td>?</td>
<td>H/S</td>
<td>Measured (but unspecified) increase “in hours of morning service”</td>
<td>260</td>
<td>?</td>
</tr>
<tr>
<td>4.</td>
<td>?</td>
<td>H/S</td>
<td>Increase in indicator of “communication with other offices in Washington and Alaska”</td>
<td>119</td>
<td>?</td>
</tr>
<tr>
<td>5.</td>
<td>None</td>
<td>H/S</td>
<td>Increase of 3% “per net staff hour”</td>
<td>112</td>
<td>Yes</td>
</tr>
<tr>
<td>6.</td>
<td>?</td>
<td>H/S</td>
<td>Increase of 6.9% “over first quarter”; increases of 14.0% “over second quarter” in rated productivity</td>
<td>150</td>
<td>?</td>
</tr>
<tr>
<td>7.</td>
<td>?</td>
<td>H/S</td>
<td>Increase of 11.6% in “office productivity” measured in terms of “median productivity per hour per clerk”</td>
<td>120</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Study Details</td>
<td></td>
<td>Outcome</td>
<td>Change</td>
<td>Notes</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------</td>
<td>---</td>
<td>---------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>8</td>
<td>Social Security Administration (Nollen, 1979)</td>
<td>?</td>
<td>H/S</td>
<td>353</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>Social Security Administration (Nollen, 1979)</td>
<td>?</td>
<td>H/S</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Social Security Administration, Bureau of Disability Insurance (Swart, 1978)</td>
<td>None</td>
<td>H/S</td>
<td>Detroit Engineer District</td>
<td>?</td>
</tr>
<tr>
<td>11</td>
<td>U.S. Army (1977)</td>
<td>None</td>
<td>H/S</td>
<td>Regional Engineer District</td>
<td>?</td>
</tr>
<tr>
<td>12</td>
<td>U.S. Army (1977)</td>
<td>None</td>
<td>H/S</td>
<td>Regional District Activity</td>
<td>?</td>
</tr>
<tr>
<td>13</td>
<td>U.S. Army, Tank Automotive Command (1974)</td>
<td>None</td>
<td>H/S</td>
<td>400</td>
<td>?</td>
</tr>
<tr>
<td>Exhibit 21.1 Continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Statistical treatment?</strong></td>
<td><strong>Hard/soft data?</strong></td>
<td><strong>Effects on “hard” productivity indicators</strong></td>
<td><strong>Number</strong></td>
<td><strong>Unionized</strong></td>
<td></td>
</tr>
<tr>
<td>14. U.S. Geological Survey (1977)</td>
<td>None</td>
<td>H/S</td>
<td>Decrease of 3.6% in Map Production in one area; and increase of 14.3% in another area. Decrease of 11.5% in Map Distribution. Increase of 13.5% in vouchers processed in Finance. Increase of 6.3% in Technical Reports processed.</td>
<td>2700</td>
<td>Yes</td>
</tr>
<tr>
<td>15. U.S. Information Agency (1975)</td>
<td>None</td>
<td>H/S</td>
<td>Increase of 4.5% of vouchers processed in Accounts Payable and Claims Section; and reduction of 16% in backlog.</td>
<td>33</td>
<td>?</td>
</tr>
<tr>
<td>16. U.S. Navy, Finance Office (Lampman, 1975 and 1975a; U.S. Navy Sea Support Center 1976)</td>
<td>Wilcoxon test</td>
<td>H/S</td>
<td>No change in quantity of work of experimental group as compared to control group; “expected increase” did not occur. Significant decrease in error rate of experimental vs. control group.</td>
<td>“Small office” of 100</td>
<td>?</td>
</tr>
</tbody>
</table>
17. U.S. government service agency (1986)  
Two-tailed t-test  
H/S  
Two tasks showed significant improvement; 3 tasks showed random changes; and 1 task reflected increased processing time. Notably, also, significant decreases occurred in both sick and annual leave rates. Employment grew from about 700 to 800.

18. Public utility, nontechnical services  
Least-square  
H/S  
“Gross reduction in” absenteeism after FWH in test unit. Approximately 140 in test unit.

?  
H/S  
Survey in process, but major effects on productivity and use of leave are reported.

agendas of those officials of both labor and management anxious to be family-friendly.

Second, FWH applications suggest a yo-yo: personnel shortages encourage them and personnel oversupply discourages them. Given the growing shortages in specific job categories, the future seems to hold some optimism for FWH as a low-cost fringe benefit. Of course, we are now emerging from a cutback and downsizing era, which implied a low priority for employee-friendly features like FWH.

Third, the past decade or so has seen another kind of zigzag effect: the number of evaluative studies seems to have dropped sharply, but greater methodological rigor characterizes the more recent studies.


FWH AND MEASURED PRODUCTIVITY

Surprisingly, public-sector studies of FWH are (if anything) more concerned about measured productivity than studies in the private sector. Most private-sector studies of FWH report few “hard data” (Golembiewski and Proehl, 1978, pp. 845–847), and the prevailing opinion there generally overwhelmingly inclines to satisfaction with FWH as a low-cost program that has positive reflections in “soft data”—as in self-reports about productivity, cooperation, an enhanced worksite, greater control over one’s work, and so on.

Exhibit 21.1 summarizes a variety of information about nineteen public-sector studies of FWH applications that have reported “hard data” about variations in productivity. Exhibit 21.1 permits two broad classes of summary conclusions—the first urges a variety of cautions about interpreting Exhibit 21.1 and the second class nonetheless presents several probable conclusions about FWH applications in the public sector.

FWH CAUTIONS AND ATTRACTIONS

Exhibit 21.1 supports a generally optimistic conclusion, given major caveats. The balance of attractions and cautions can be suggested by several brief sketches.

Six Cautions About Exhibit 21.1

At least six cautions about the studies summarized in Exhibit 21.1 require up-front emphasis. First, despite major efforts, the original reports of many of the studies were not available for present review. These studies were available in
Flexible Work Hours as Empowering Potential

summarized form in one of two sources (Ronan, Primps, and Cloonan, 1978; Nollen, 1979).

Second, most FWH applications involved units of small size doing clerical or professional work, so questions about the generalizability of results to larger contexts are appropriate.

Third, the nineteen studies resulting from two searches are not a picture of methodological elegance. Exhibit 21.1 does not give the details, but most designs are pre versus post, without comparison groups. This shortfall can have serious consequences, as in Case 10 in Exhibit 21.1. There, similar effects were observed in both experimental and comparison units, which some may interpret as a “Hawthorne effect” rather than a FWH effect. General failure to maintain specific experimental and control conditions precludes such useful interpretation.

Fourth, individual studies measured one or a few aspects of the quantity or quality of production. This implies the possibility that the data reflect “balloon-squeezing”—looking good on some measures of the job but only by neglecting others.

Fifth, the studies typically use only pre versus post comparisons, often with short observational intervals of less than six months, without comparison groups. Longitudinal studies covering a year or more would be helpful in eliminating alternative explanations of observed effects. Use of comparison groups would serve a similar purpose. For example, an apparent change on a short-post observation might constitute only a temporary outlier in a data-set covering a broader span of time. Without comparison groups, similarly, one might miss the fact that an apparent increase might only reflect a general trend and not the effect of a FWH application. Experience also implies that some FWH effects take several months to develop, especially with respect to supervisory behaviors and attitudes. (Golembiewski, Fox, and Proehl, 1979, pp. 248–250).

Sixth, few of the nineteen studies in Exhibit 21.1 use any statistical treatment beyond simple arrays of data. Patently, this implies questions as to FWH applications. One cannot answer this key question by eyeball examination: When is a change “big enough” to be considered nonrandom?

Seventh, despite a few exceptions that seek to get at underlying processes (e.g., Golembiewski, Billingsley, and Yeager 1976; Graf, 1976; Dalton and Mesch, 1990), almost all FWH have a narrow “outcome bias.” Informed speculation abounds as to the specific processes energized by FWH applications, but data are in short supply.

Five Encouragers of a Robust Interpretation

Given such cautions, other factors encourage a relatively robust interpretation of the studies reflected in Exhibit 21.1. First, available studies imply some generalizability. For example, at least over half of the nineteen applications—and no doubt
several others—occurred in union settings. Some of the literature (e.g., Goodman, 1979) suggests that union contexts encourage some optimism about the usefulness of the intervention, despite the pessimism of some observers (e.g., Chavez, 1997).

Second, the later search isolated only a few studies of the productivity effects of FWH, but we know that applications were sharply increasing between the two intervals. No doubt these facts permit several interpretations, at least as first approximations. Reasonably, however, that dual pattern suggests such substantial confidence in FWH effects so that managers need not gild the veritable lily by difficult research. Relatedly, the recent literature reflects no disaffection with FWH by labor or management (e.g., Ezra and Deckman, 1996; Government Finance Review, 1998).

Third, the general thrust of Exhibit 21.1 reflects a definite balance of positive effects on measured productivity. Specifically, a large portion of the “hard” or “objective” measures imply increased productivity, and markedly so in a number of cases. Only three of the nearly 30 measures reflect any drop in productivity post application.

Fourth, this overall thrust of the data is quite powerful when coupled with other trends in effects associated with FWH applications. Specifically, almost all FWH studies include “soft” or attitudinal data as well as “hard” data, which may be considered objective or nonreactive. Multiple measures often permit multiple cross-comparisons to confirm or deny what hard or soft data alone imply. Recent FWH research also is more methodologically conscious in cases (e.g., Dalton and Mesch, 1990), as is reflected even in Exhibit 21.1.

Extensive cross-comparisons clearly reinforce the overall thrust of Exhibit 21.1. These attitudinal data very affirmatively attribute a broad range of favorable effects of FWH on aspects of productivity (Golembiewski and Proehl, 1980; Nollen, 1979; Ronen, Primps, and Cloonen, 1978; Swart, 1978). Moreover, a range of hard data—on absenteeism, overtime, various costs, and so on—provide similar reinforcement for positive effects (Golembiewski and Proehl, 1980; Ronen, Primps, and Cloonen, 1978; Dalton and Mesch, 1990).

Fifth, the trend in Exhibit 21.1 gains added attractiveness in the light of the low cost of FWH applications. Available data, for example, indicate that FWH requires changes in the behaviors of supervisors. But these changes come in an attractive package: they involve competencies that almost all supervisors have or can come to develop in brief time-frames; they require little training and other overhead support; and those new supervisory behaviors or attitudes tend to be ones to which employees respond positively (Graf, 1976; Golembiewski, Fox, and Proehl, 1979).

These low costs for FWH applications and relevant training contrast sharply with the high costs often associated with other OD interventions, especially interaction-centered designs. These often imply costly training in attitudes and behav-
Flexible Work Hours as Empowering Potential

ioral skills, as in group decision-making or intact work units, which often appear in team-building experiences. On the general point, see Golembiewski (1979, esp. pp. 85–132).

REFERENCES


Enhancing the Empowering Potential of the Concept “Workplace”
Flexi-Place as Exemplar

Even casual students of industrial and commercial development are aware of the importance of changes in the concept of the workplace. At one time, for example, cottage industries dominated; then, with the development of centralized power capabilities, it became not only practical but also economical to change the site of work. Many times, those changes were coerced, so attractive were they to economic elites. Suddenly, employees in huge numbers went to their work, often at considerable distances, to labor cheek-by-jowl with their colleagues. Much evidence indicates major dislocations—social, psychological, and familial—resulted from these new ways of organizing wit and will.

This chapter deals with another redefinition of workplace, which applies to growing numbers of today’s employees, if clearly not to most or all. This redefinition encompasses yet another way to increase responsible freedom at work. Here, this redefinition is called Flexi-Place—henceforth, F-P—which can help serve a broad range of interests, personal as well as collective and economic, of the employees as well as of management.

A CONCEPTUAL NICHE FOR IRONY VIII
F-P was neither invented by ODers nor widely accepted by them, but F-P can clearly serve the same values that OD espouses. This chapter devotes attention to that potential as it reviews the relevant literature, not much of it reflecting an
**Exhibit 22.1 Five Selected Domains Facilitating F-P Applications**

<table>
<thead>
<tr>
<th>Individual</th>
<th>Work</th>
<th>Organization</th>
<th>Career</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low needs for affiliation and support or willingness to moderate them</td>
<td>Jobs feature chunks of related activities rather than separate and sequentially interdependent activities in different jobs</td>
<td>Culture features trust in employees as well as between them and managers</td>
<td>Willingness not to be on management track, at least for a time</td>
<td>No preschool children, or adequate child care, constitute the simplest case</td>
</tr>
<tr>
<td>High need for autonomy</td>
<td>If in bureaucratic structures, jobs are loosely linked</td>
<td>Focuses on objectives and results</td>
<td>Strong technical orientation</td>
<td>Where preschool children live at home (20 percent in federal pilot study), or where participant or dependent is disabled (12 percent in federal pilot study), no pattern of adverse consequences was reported, but presumably these constitute possible risk factors that can be compensated for by care concerning other features</td>
</tr>
<tr>
<td>Record of solid prior performance</td>
<td>Permits time at home as well as at a &quot;normal&quot; workplace</td>
<td>Recruits and socializes supportive supervisors, able to monitor out-of-sight employees</td>
<td>Willingness about, and capacity to function with, reduced access to informal networks and schmoozing</td>
<td>Family is supportive and understanding about separating some work/home domains, while integrating others</td>
</tr>
<tr>
<td>Substantial need for balance between home and work demands</td>
<td>Is high on motivators vs. satisfiers (Herzberg, Snyderman, and Mausner, 1959)</td>
<td>Does not use piece rates, at least in a coercive mode</td>
<td>&quot;Comfortably plateaued,“ at least for the F-P interim; doing well and content to keep doing it</td>
<td>Adequate space and conditions for worksite at residence</td>
</tr>
<tr>
<td>Low need for upward mobility, at least for a period of time</td>
<td>Permits reasonable security for data sets</td>
<td>Recognizes dual ladders—managerial as well as technical/professional</td>
<td>Wiling about, and capable of, development in professional and/or entrepreneurial ways, as contrasted with reliance on bureaucratic promotions.</td>
<td></td>
</tr>
<tr>
<td>A self-starter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Flexible Workplace as Empowering Potential 533

explicit OD mode. But that is regrettable rather than unavoidable. Indeed, this
general neglect is a powerful motivator of this chapter—an easy opportunity
to both guide and protect a notion useful for increasing responsible freedom in
organizations as well as to increase the range and scope of OD applications prom-
ising substantial success rates.

Conceptually, F-P is at once a single notion that admits great variety and
even some complexity as well as major doses of practicality. Basically, F-P pro-
vides that some portion of an employee’s workweek be spent at work in a nontra-
ditional setting—as in the employee’s home or in some ad hoc setting remote
from typical employer sites. Little is typical about F-P, but the approach often
provides for some time each week at a central office—to have face-to-face meet-
ings, reinforce bonds between employees, and so on. The recent introduction of
electronic means of work and contact—the computer in a network, FAX, e-mail,
and so on—provides the biggest push for F-P and the search for favorable costs/ben
efit cases to implement.

Exhibit 22.1 provides a useful introduction to F-P as it relates to fusing
important domains of life and work. In an elemental sense, indeed, the character-
ization suggests a return to the past. For most purposes, that exhibit is transparent.
Thus, it provides a mix of costs/benefits for individuals and organizations. Indi-
viduals might suffer some cost (for example) by being “out of the flow” of normal
dynamics at a central worksite while experiencing countervailing benefits—re-
duced commuting time, eased responses to the myriad conditions in individual
families, among numerous other benefits. Exhibit 22.1 also sketches several fea-
tures facilitating levels of aggregation in addition to the employing organiza-
ton—individual employees, their work, careers, and families.

PROFILE OF F-P EFFECTS: PRE-1990, AND ESPECIALLY
IN FEDERAL GOVERNMENT

A decade or so ago, F-P was seen as the answer to many needs. Thus, futuristic
human resource planners anticipated telecommuting impacts at multiple levels;
individual and familial, in responding to the issues associated with great and
growing proportions in the workplace especially women and single heads of
households. The win-win features at multiple levels include:

- **Individual**, as in increasing the range of choices while decreasing some
costs of conventional employment
- **Organizational**, as in reducing rent and other costs of conventional work-
sites
- **Social**, as in countering the growing alienation between work and family

Boas Shamir (1992, p. 275) gets to the heart of F-P’s perceived leverage. He
observes: “Here was an opportunity to rectify [major problems], achieve a reinte-
gration and reduce some of the tension attributed to the organization of work in industrial society.” That possible leverage encourages expansive thought. Shamir adds (1992, p. 275), “it was believed that working at home would ‘free work from the constraints of time and location’, to use a phrase coined by Bailyn (1998), and increase the quality of working life.”

Despite the Johnny-come-lately status of F-P compared to business and entrepreneurial applications (e.g., Fleming, 1990; Mills, 1993), let us concentrate on a federal F-P initiative launched in January 1990 as a pilot study. Federal F-P may be described as (Joice, 1993; p.i; emphasis added): “a government wide, nationwide project which allows Federal employees to work at home or at geographically convenient satellite offices for part of the workweek. . . . FP was established to improve the government’s ability to recruit and retain capable employees, to improve employee quality of life, and to reduce Federal operating costs.”

EXHIBIT 22.2  Some Major Features of Flexi-Place Application in Federal Government

Over 90 percent of F-P employees report that their performance was either unchanged or improved, with some variation for different aspects of performance. Nearly 85 percent had prior job ratings of either “outstanding” or “exceeds fully successful” (Joice, 1993, pp. iii–iv, 12, 16–19, 21)

Disability of either participant or a disabled dependent was characteristic of 12 percent of all cases surveyed

Twenty percent of federal participants had preschool children living at home (Joice, 1993, p. 11). Nearly 9 in 10 participants had over 11 years of work experience (Joice, 1993, p. 121)

Some supervisors—but no more than 2.5 percent in the federal pilot study—were concerned about their control having been reduced over the job performance of employees on F-P (Joice, 1993, p. 19)

Depending on the facet of personal life, between 12 and 77 percent of respondents reported “improved” with no more than 3 percent reporting “decline” on any facet (Joice, 1993, pp. 23–24)

Adequacy of and access to materials and equipment was seen as having “declined” for about a third of federal participants (Joice, 1993, pp. 25–26)

Sick leave usage was lower for over 40 percent of the participants and stayed the same in about 45 percent of the cases (Joice, 1993, p. 28)

About 80 percent of supervisors saw F-P as desirable in cases with minimal refinement, as did virtually all participants (Joice, 1993, p. 32)

A full cost-effectiveness estimate of the federal F-P pilot is not yet available (Joice, 1993, p. 41), despite favorable indications re recruitment, retention, and sick leave

Source: Based on Golembiewski, 1996, p. 147.
It requires no special insight to see how F-P could serve several diverse populations. From the start, F-P was seen as helpful in placing disabled employees, who constitute a strategic and visible subpopulation of that far larger cohort of employees having problems with urban transportation. Other probable users include single parents, notably women with young and especially nursing children; those who want to vary their routines; and those who provide home-based care for the aged or ill or for other dependents.

F-P also is responsive to a very large subpopulation—the 43 percent of surveyed federal employees who reported that their most productive periods occurred outside of normal working hours (Joice, 1993, p. iii). This is consistent with the long-known fact that not all people have similar circadian rhythms (e.g., Patkai, Petterson, and Akerstedt, 1973). In short, there seem to be “day people” and “night people,” and conventional worksites limit the adaptations that individuals can make to their own rhythms.

High performers also got special recognition in the federal pilot program. Participants were carefully selected, and care was taken to safeguard the interests of the several participating agencies, the employees, and any labor union or professional associations.

Business or entrepreneurial F-P applications had been ongoing for a decade and more before the federal pilot study began and have a similar profile, motivations, and outcomes. And applications to public-sector application at state and local levels also have occurred, with similar effects.

Conveniently, let us focus on the public side of employment to characterize such common F-P effects. Exhibit 22.2 provides a useful profile of F-P effects, summarizing the effects of large-scale application. The summary speaks largely for itself. In short, F-P has a range of low-cost and positive consequences.

ENHANCED DESCRIPTION, PRO AND CON

It would be a mistake to leave matters with Exhibit 22.2, however. To begin, let us paint with a broad brush. Much of the F-P literature urges the very broad and general balance of attraction versus constraints. For example, several careful observers (Mokhtarian et al., 1998) present this list of positive motivators for Flexi-Place applications:

- Work-related (e.g., to get more work done)
- Family-related (to spend more time with family)
- Leisure/independence-related (to have more time for self)
- Travel-related (to reduce the time or stress of commuting)
- Ideological (to save energy and improve air quality through reduced travel)

Arrayed in contrast, but less than counterbalancing, are two classes of unattractive factors of F-P applications:
External variables, including awareness-related, employer-related, and job-related

Internal (psychosocial) variables such as the desire for social or profession interaction, lack of self-discipline, risk aversion, and household distractions

Awareness-related constraints of F-P include realistic expectations about F-P, as contrasted with viewing it as a kind of managerial wonder drug.

The early and general over-balance of motivations versus constraints appears in most early sources, including Gray, Hudson, and Gordon (1993) as well as Handy and Mokkartarian (1996). Originally, such expectations were seen as particularly likely in the public sector, but the effects on further consideration seems to be quite general.

Although the balance of attraction versus constraints seems clear enough in the general case, as above, it seems probable that differences can be expected by various segments of the workforce. Relevant work has hardly reached a mature stage, but it is not extravagant to use one study’s results as illustrative, if not prototypical. Thus, Mokktarian and colleagues (1998, esp. p. 1115) illustratively sensitize the reader to several downside features of F-P.

Most broadly, women on average rate the advantages of telecommuting more highly than men—both overall and within each occupational group. Women were more likely than men to have family, personal benefits, and stress reduction as potential motivations for telecommuting, and they also were more likely to reflect the constraints of supervisor unwillingness, risk aversion, and concern about lack of visibility of management. Clerical workers were more likely than managers or professionals to see the family, personal, and office stress-reduction benefits of telecommuting as important, whereas managers and professionals were more likely to cite getting more work done as the most important advantage of telecommuting. Constraints present more strongly for clerical workers than for other occupations included misunderstanding, supervisor unwillingness, job unsuitability, risk aversion, and (together with professional workers) perceived reduced social interaction. Constraints operating more strongly for professional workers included fear of household distractions, reduced social and (together with managers) professional interaction, the need for discipline, and the lack of visibility to management of F-P employees working at a distance, as it were. Key constraints presented by managers included reduced professional interaction and household distractions. Lack of awareness, cost, and lack of technology or other resources did not differ significantly by gender or occupation.

Let us follow this broad-brush treatment with one oriented toward several details. At a minimum, two sets of details are useful: one set concerns useful details supporting F-P, and a second aggregation assembles generalizations concerning why F-P has reduced effects, fails, or is considered inapplicable.
Details in the Supportive Case

Several classes of details support optimism about F-P. First, the setup costs for F-P seem manageable, even easily so. Expenses for a full-scale home office are estimated to be about $5,000 for computer, printer, phone and modem, as well as fax machine (Mills, 1993). In the federal service pilot project, however, the costs of supporting start-up for a single F-P employee seem to be substantially less than $1,000 (Joice, 1993, pp. 31–32).

Second, relatedly, F-P can economically serve various policy initiatives. For example, the Americans with Disabilities Act requires that employers make accommodations for the disabled, if they are not overly difficult or expensive. Flexi-Place can provide a reasonable way to meet these legal requirements while retaining a valued employee and easing the rigors of getting to and from work.

Third, what about the bottom line? The available evidence, though necessarily tentative in the early literature, contributes to a generally positive picture of cost/benefits. Early findings indicate that F-P variants have effects much like those reported in the federal study (Joice, 1993, p. 49):

Measures in areas such as job performance, motivation, quality of life, and cost indications . . . show that Flexi-Place, using employees with proven performance, was a success. Flexi-Place shows promise as an effective mechanism for national efforts regarding work/family, transportation, and energy issues. . . .

Most participating organizations spent very little money on pilot programs. . . .

Indications of improved job performance (productivity), reduced usage of sick leave (benefits), improved health (health care) and reduced vehicle usage (transportation/energy issues) for a significant proportion of the participant group suggest long run reductions in costs.

Fourth, the federal experience seems to be broadly representative of F-P effects. For example, Lis Fleming (1990) highlights these covariants of F-P, wherever—increased productivity; reduced costs of office space, even for larger staff; and decreased absenteeism. And Phillip E. Mahfood balances a case or so of failed F-P with a huge list of benefits (1992, pp. 10–30), as well as with recaps of results like those from a four-month study in San Diego County. Among other effects, Mahfood (1992, pp. 119–22) reports that the county saved $1,440 per year in the costs of office space alone for every telecommuter working at home two or more days per week, productivity increases were up to 40 percent, and participants saved 3,300 vehicle miles of travel during the 120-day study.

Anecdotal evidence supports such descriptions in detail. For example, a few researchers have identified substantial F-P population (e.g., Christensen, 1988); many observers are aware of firms like the F. I. Group, which grew from...
two employees in 1964 to some eleven hundred in 1988 and does perhaps 70 percent of its work in nontraditional sites (Kinsmon, 1987). Other observers see F-P as spreading like wildfire (e.g., Mahfood, 1992), and at least a dozen projects in state and local governments also have been authorized early (Desky, 1993b, p. 1).

However, the overall impression of this author is that attention to F-P diminished after the early 1900s, and for obvious reasons. The rationale for F-P has heavy loadings of a desire to reward and retain valued employees with a status many desire. The Human Resources bias post-1990 or so has turned, and definitely so, to “weeding out the dead wood.” Nonetheless, F-P ways and means are useful to understand: fashions always change; and even if fashions never change the HR challenges may include definite biases, but a diversified portfolio typically is always useful, even if some components are used only infrequently.

Details in the Early Contra-Case

Perhaps basically, some researchers question whether F-P is now sufficiently diffused to warrant serious study (Olson, 1998). Or perhaps it never will be, some suggest. Illustratively, one observer reports that only ninety of nineteen thousand eligibles elected an F-P option (Perin, 1990). Similarly, Aetna at one point had forty-five thousand employees, of whom 70 percent were women, and we are told only that “hundreds telecommute from their homes” (Hymowitz, 1990).

Other critics highlight the common selectivity of participants in F-P evaluations, as was the case in the federal pilot study (Joise, 1993). The concerns of these critics have an obvious base, but only limit F-P’s diffusion rather than deny its usefulness.

Despite the record sketched above, F-P diffusion has not been as rapid or as thorough-going as some had expected—or perhaps hoped is the better word. In any case, no reasonable estimates of F-P incidence were available in the early 1990s. Although it appears that nearly 8.5 million people worked at home at least eight hours per week during the mid-1980’s, most of them were not telecommuting in the full sense (e.g., Shamir, 1992, p. 286).

What accounts for this apparently limited diffusion? Three points provide useful perspective. Three classes of demotivators can be distinguished, all associated with the force field in which a F-P installation exists. In both of the first two cases, F-P can at once contribute to OD and also be served by it. A brief look at some legal issues concludes the contra-case.

Toward Regenerative Interaction

With F-P as well as with an attempt to move beyond bureaucratic conditions, the approach to regenerative interaction seems useful, if not critical. There, recall, openness, owning, as well as trust are high, and risk is low. In general, the under-
Flexible Workplace as Empowering Potential

lying rationale seems apparent, although it has not received direct research attention in the F-P literature.

The contrary case has many exemplars—F-P should fail where degenerative interaction dominated. Generally, where low trust exists, it does not appear that supervision could be exercised at a distance, as it were. And low trust also inhibits the development of objectives that facilitate oversight by results rather than direct observations. In fact, this recognition of the high trust linkage is reflected in the early federal experiment, which assigned F-P status only to those with long histories of superior performance.

Relatedly, some managements have used F-P to create mischief. For example, one business organization decided that all employees were independent contractors which, among other features, would “save” the employing organization the substantial fringe benefits due to all regular employees. The F-Pers complained that the ex post facto reclassification came as news to them, and they successfully sued the employer (Mahfood, 1992, pp. 151–55). Needless to say, F-P had a limited life in this organization, whose interactions systems came to take on ever-more degenerative features, given management’s decision to play coercive games.

Toward Non-Bureaucratic Structures

Bureaucratic concepts do not provide ideal structural fit for F-P, and a sense of this elemental fact often inspires management resistance. For example, F-P variants are uncomfortable with the issue of how one can monitor someone who cannot be seen (Desky, 1993, p. 16), which implies an obvious kissing cousin of the close oversight prescribed by bureaucratic structures. Thus, Constance Perrin (1990) notes that the “principle of continuous visibility” is powerful in the bureaucratic model. In addition, since that kind of control may variously contribute to performance as well as complicate it, F-P opponents have some reasonable grounds for their counter-argument. From bureaucratic perspectives, Perrin observes (1990, p. 22), managers can see F-P as a “repudiation of their inspectional and disciplinary rights.” Hence, the resistance to F-P often encountered in traditional organizations, even as a matter of principle.

As everywhere else in this book, policy innovation needs reinforcement by appropriate structure and interaction. So practical reasons for revisiting F-P can inhere in traditional practices. Looked at from the other side of the coin, resistant managers do not tend to help others to develop goals and norms that will permit effective management without resort to close oversight. Typically, management by outcomes versus close oversight takes off only after cross-training and job enrichment, which is to say, after some experience with post-bureaucratic structures and concepts (see also Golembiewski, 1996). Thus, cross-training provides powerful reinforcement for F-P and, almost by definition, bureaucratic structures either inhibit or preclude cross-training. That is, bureaucratic jobs are often
dumbed-down activities, which are closely linked to other limited activities. Such jobs are not easily unbundled for F-P. Such practices as cross-training free supervisors from hovering at the employee’s elbow, as it were. Training in goal- and norm-setting also is more consistent with flow-of-work structures, as the text above establishes at various points (e.g., Chapter 14).

Several of the observations made with respect to flexible work hours (FWH), in Chapter 21, also apply here. For example, substantial cross-training or job enrichment can be very helpful at both loci of F-P work—at home as well as at a conventional office. That individuals can be in touch electronically reduces the need for any piece of work to be done at a specific site, but training in flexibility as well as enriched jobs are needed to limit the reliance on supervision of details. Bureaucratic structures do not help here.

For such reasons and others, close observers of F-P do not hesitate. Shamir (1992, pp. 281–82) sees the bureaucratic structure as the main impediment to F-P applications. Refining this basic point, Olson (1998, emphasis added) concludes that the traditional model is “a major barrier even when technology is available and cost-effective.” And cases exist in which some F-P applications have been bureaucratized to death, as by attempts at electronic surveillance of the micro-productivity of keyboard operators (e.g. Shamir, 1992, pp. 282–83).

Legal Considerations in Early F-P

Adopting organizations also will have to consider various legal prescriptions (e.g., Mahfood, 1992, pp. 99–116), which are real but do not seem onerous. At an extreme, for example, the legal system could sharply limit F-P applications. Consider a simple decision rule: that all “places of work” require the same or similar features. For example, similar medical and emergency features might be required. Or similar entrance/exit facilities might be mandated. Or similar sprinkler systems to inhibit fires might be specified. Clearly, what is appropriate for an office of (let us say) 25 people might be uneconomical and arguably unnecessary for someone working in a spare room in his or her usual residence.

In any case, such legal decisions would clearly limit the cost savings to management that inhere in F-P offers in residences. On occasion, indeed, the U.S. Department of Labor flirted with such draconian requirements for F-P but quickly backed down.

RECENT EXPERIENCES WITH F-P

The early literature on F-P, ending with the report on the 1990s federal pilot study, does not have to be qualified with lessons learned the hard way in more recent experience. Although the F-P research literature remains a slim one—as viewed from the perspectives of Ironies I (1990) as well as of Ironies II (2002)— we can gain some sense of the staying-power of the validity and reliability of
early claims. The pages above constitute a short post test, as it were; and this section adds a longer post test, about five years beyond the first assay. One government report, 1990–93, is summarized in exhibit 22.2 and it presented a first federal overview (U.S. Office of Personnel Management and U.S. General Service Administration, 1993); a second, more ambitious and more methodologically aware report (U.S. General Accounting Office 1997) seeks to subject the results of the earlier report to strict analysis. The GAO (1997) report is not an assemblage of rigorous results but instead applies the experience of usual accounting standards to evaluation of F-P applications. It involved twenty-one site visits, reviews of policies, and interviews throughout the public service, irrespective of the F-P site at a residence or some ad hoc F-P center that agencies might share. The 1993 study was basically a survey of the first cohort of F-P applications, who worked only out of their homes.

The 1997 study supports eight illustrative generalizations, which in cases extend the 1993 conclusions but in no case require a substantive revision of the original optimism. First, later observers reflect a basic confidence that F-P has a positive-enough balance of consequences that it should not be regarded as a reward for past performance but should be seen as another alternative for managing work so as to increase responsible freedom. In the federal pilot study, in contrast, F-P participants were hand-picked for major achievements in performance.

Second, despite the strongly positive tone of the 1993 report, no basic F-P legitimization now exists in federal law. Thus, one early objective has still not been met. Nonetheless, F-P has received numerous boosts in the public sector, as by calls for more family-friendly workplaces in presidential executive orders.

Third, the self-selection dominant in early F-P evaluations has been remedied only in part. Directly, the 1993 study used a small and selected population—as in a strong preference for disabled workers—who all worked out of their residence; and the 1997 study adopts a notable but limited relaxation of such limitations, and hence its results are less subject to criticism as only narrowly generalizable. For example, the 1997 study also includes centers established for F-P work, whereas the 1993 F-Pers all worked at home and hence were easy targets for the criticism that apparent cost-savings could be explained by various shortcuts and lack of amenities that might have important consequences for employee health and safety (U.S. GAO, 1997, pp. 34–38).

Fourth, F-P cost/benefits have remained at least as attractive in the 1997 study as they were in the 1993 study and in the earlier pilot study. Indeed, the attractiveness may have increased, although specific managerial training may have been required to spread the word (e.g., U.S. GAO, 1997, pp. 14–15). For example, over 77 percent of responding supervisors noted that managing F-P employees presented no greater challenges than traditional worksettings.

Fifth, much room exists for future F-P applications in the federal government, despite the positive reception of the 1993 and 1997 studies. By the end of
1996, the federal goal was 60,000 F-P users (U.S. GAO, p. 3). In contrast, a July 1996 estimate by the President’s Management Council put F-P federal users at approximately 9,000 (U.S. GAO, 1997, p. 4). For example, F-P applications typically are limited to “professionals” (U.S. GAO, 1997, pp. 12–13), which sharply limits adoptions. The GAO report sees this limitation as too restrictive, by far (1997, pp. 9–11). Even so, matters had been worse: the federal pilot study admitted only those with exemplary performance records, a bias also dropped in many later applications identified by the 1997 report.

Sixth, this modest record of diffusion—especially in the context of the positive results attributed to F-P—implies several major inhibitors. These include the following:

Agency resistance seems clearly the major restraining force, since F-P has a poor fit with traditional bureaucratic structures, policies and procedures, or attitudes (U.S. GAO, 1997, pp. 8–10, 14–15).

Sharp restrictions on the jobs made available for F-P usage.

Seventh, the GAO report (1997) emphasizes that most constraints on F-P are surmountable, even easily so. Indeed, such resistance may have no roots in bad experiences: difficulties with F-P applications seem rare (U.S. GAO, 1997, pp. 4–6). Moreover, demonstration sites have been made available to provide in vivo examples of F-P applications. And the GAO report encourages much broader applications than many agencies permit. Even in agencies now accepting F-P, for example, typically 90 percent or more of all positions may be excluded as possibilities for F-P.

More broadly, F-P applications apparently could profit from OD guidance. Directly, F-P is poorly suited to degenerative interaction, and degenerative features are to be expected as common consequences of bureaucratic ways-and-means.

Eighth, F-P practice has successfully incorporated relevant stakeholders in negotiating F-P installation, as in the case of employee unions. That right, indeed, is guaranteed by federal law (U.S. GAO, 1997, p. 5).

**SCHEMA OF A CONTEXT FOR MUTUAL BENEFIT**

In conclusion, F-P experience and theory provide some evidence of a major win-win potential, but we must proceed with caution about how far to push this claim. Overview could be enlarged, even substantially, but the present point should be clear enough. F-P is not for everyone in every work situation, for certain, but it exemplifies another easy piece that has a reasonably broad applicability. In many cases, it can provide win/win capabilities—meeting a range of conventional as
Flexible Workplace as Empowering Potential

well as idiosyncratic employee needs while also meeting organizational needs for efficiency and economy.

Let the text be specific about what might profit from curbing. Some P-F supporters seem too exuberant about the potential of Flexi-Place to generate balanced effects, and this section provides necessary contrast in two particulars. First, as is always the case, the role of values is great in Flexi-Place applications, and here OD as both technology and as normative infusion can be of great value. For example, many early applications echoed the familiar OD theme of greater responsible freedom at work, as in the resolve to develop leaders rather than mere monitors, to deal with results instead of controlling workers’ behavior or time (Joise, 1993). Here F-P applications include such advantages as employing disabled workers. On occasion, however, F-P has been guided by values that narrowly served management and disadvantaged employees—as in the goals of undercutting labor unions or depriving employees of once-held benefits or status.

Second, more broadly, balanced F-P applications also can profit from contexts influenced by technologies-cum-values such as OD. For example, Joice (1993, pp. 56–57) highlights the following characteristics of a “well-designed program,” among others:

The basic design takes into explicit account costs/benefits to all stakeholders.
Employees need to be provided with equivalently equipped work stations.
Mutual adjustments of all stakeholders are anticipated and worked through.
Flexi-Place employees do not suffer from reduced benefits and opportunities.
Flexi-Place functions in a manager’s environment that emphasizes results rather than close observation by supervisors.
Job analyses should determine the jobs to be flexed.
Some-in-office time seems required to preserve communications and keep flexing employees aware of opportunities.
Trust and confidence between stakeholders are critical in attaining and preserving regenerative interaction at work.

Also broadly relevant are various risks of Flexi-Place. Useful detailing is available elsewhere (e.g., Mormer Solomon, 2000, p. 58). Mormer Solomon emphasizes the character of the workplace, employee health, information security, and personal safety. Recent policy flip-flops at the federal level highlight important areas (Mormer Solomon, 2000, p. 56; Sudbury and Towns, 1997).

All these risks and benefits will require mutual adjustments between stakeholders. Useful experiences with such adjustments exist (e.g., Nguyen, McShane, Willmering, and Harrison, 1995), and they imply that common ground can be found without herniating stakeholders.
REFERENCES


Enhancing the Empowering Potential of the Concept “Development”
Demotion as Exemplar

This chapter may seem like preaching to a ghostly choir. After all, “development” seems inherent in empowering, and vice versa; So, why gild that lily? The rationale is compelling, to be sure, but it is not entirely obvious.

Well, the purpose here is quite direct and necessary. It seems both convenient and useful to seek “development” in adverse personnel actions such as demotions, in addition to the more conventional and positive senses of “development” in growth or upgrading. Why bother with demotion? Paramountly, perhaps, we will see more adverse personnel actions in the future than occurred in the past. Certainly, the ideal psychological contract about employment is changing. Not so long ago, that ideal was oriented toward career-long employment by an employer, given reasonable performance—see AT&T, General Motors, Smith Kline & French, Coca-Cola, and so on. Nowadays, very few employers are so bold. At most, a few aspire to the ideal of keeping employees continuously employed, somewhere (e.g., Golembiewski, 2000).

By implication, today’s less forgiving times require that we have to get better at “development,” in both positive and adverse varieties. That compound point has not penetrated deeply enough into practice, and hence this chapter. Speaking metaphorically, effective development must relate to “rain” as well as “sunshine,” and for one basic reason: we seem to be in for more turbulent weather, organizationally speaking; and even if that does not prove to be true, you need both sunshine and rain to make the flowers grow.
“RAIN” IN DEVELOPMENT

There is no author’s originality in this metaphor. Goodwin Watson was a wizened and wise facilitator of developmental experiences for individuals, and he was fond of noting, “It takes the rain as well as the sun to make flowers grow; too much of either can be damaging”—and Goodwin was so correct.

OD devotes far more attention to the “sun”—via encouraging trusting interpersonal relations as well as by building regenerative interaction, among other efforts. Thus, OD evolved in basic response to the challenges of organizational growth—quite attractive challenges, as that genre goes. Despite the real risks, these challenges have definite practical and conceptual upsides. They focus on expanding competencies and opportunities, and so on. As such, they attract—more than beat on—the consciousness of readers like your author.

In general, organizational studies have not strayed far or often from the sunny side of the street. Although it is doing better, OD has given too little attention to the organizational rain—to downsizing rather than growing, to contracting rather than to expanding opportunities. For some early exceptions, see Golembiewski (1979, 201–211). The Organization Behavior literature has a similar aversion, although that also has changed some (e.g., Sutton, Bruce, and Harris, 1983).

Let us be a bit more specific here. No doubt exists on one point. Manila (1967, p. 161) observed that “demotion [is] a generally neglected topic” and also provided confirming data that gets support from recent years (Kohl and Stephens, 1996, pp. 75–76). This record exists despite the fact that solid majorities of managers (e.g., 75 percent in Stephens and Kahl, 1989, p. 39) indicate that demotion is a useful tool. Nonetheless, to choose one example, managers in the oil and gas industry assert that 60 percent do not plan on using it as a human resource tool; and that perhaps only a tenth of these managers relied on demotion in the past two years (p. 39). The researcher More (1967, p. 213) stresses that this is an awkward combination and is likely to become more so.

In sum, this chapter sees an open opportunity in demotion to enrich practice via an OD design that seems quite effective. In this sense, then, demotion can at once inform practice and perhaps motivate more OD applications with already substantial success rates.

Broadly, this chapter proposes that this bias toward the sun is awkward, and that a greater balance is needed. Essentially, “rain” there will be, and neglecting it is pollyanna-ish. Moreover, despite the dangers, organizational rain provides real opportunities for moving toward the same OD values that can be more pleasantly approached via sun. This chapter continues on the theme of easy pieces—relatively convenient designs, which require only journeyman intervening skills, and which seem to have a high probability of inducing intended outcomes for which there is an increasing need.
Demotion as Empowering Potential

A CONCEPTUAL CONTEXT FOR IRONY VIII

The focus here is on the demotion design, whose initial application (Golembiewski et al., 1972) came before its time.* Nowadays, such an initiative could occur in many loci. Then, the locus was a corporate division in difficult economic straits, even as the national economy moved along at a good clip. The demotion design was developed nearly three decades ago for the express purpose of gaining experience with an adverse personnel action alternative to termination, with a second major goal of extending the values of an ongoing OD program. The challenges involved enlarging the repertoires of employee and managerial responses applicable at a difficult time in order to retain experienced personnel who could provide a valuable and quick boost for organizational regrowth when the economy revived. Paramountly, perhaps, the design ought to acknowledge an alternative acceptable to employees when conditions changed after demotees had performed well enough in the past.

The basic idea did not find an initially receptive local management, often fixated or even frozen during a period while they moved toward making a too long delayed decision. Most experienced observers did not believe the demotion design would work. They doubted that demoted workers could be effectively integrated into the work force, which was the key to retaining their experience and productivity until the economy revived. They saw the exercise more in terms of consequences encouraging later guerrilla warfare by those seeking to even the score.

The results prove otherwise, not only in the initial application but in several replications. The design generates attractive results—indeed, highly favorable ones. The beneficiaries are individuals who served their organization satisfactorily in the past, as well as the managements who acknowledged obligation but who saw no alternative other than generous separation packages.

Bad times came upon the U.S. economy again in the late 1970s and the early 1980s, and then the 2000s, and later still; this time the repertoire of responses was broader than in 1972. Shortened weeks for all in work-sharing programs have replaced termination for some employees; moreover, union contracts have been renegotiated to support participation in some downsizing; again, various job-sharing and part-time arrangements have been utilized to fine-tune the balance between employee needs and economic exigencies, and multiple if variously useful efforts have been made to help employees constructively face loss of jobs from cutbacks, plant closings, and the like. In the case of a Canadian oil and gas firm, all of the employees were brought together in an auditorium, and emerged having met cutback requirements by a variety of self-choices: work

* An earlier version of this material appears in Golembiewski, 1982–83. See also Golembiewski 1977, 99–122; and Golembiewski et al., 1972.
sharing, permanent or temporary reductions in workweeks, and so on. (Golombiewski, 2000). Overviews of such approaches to downsizing are conveniently available (Hall and Isabella, 1985).

DEMOTION AS A “RAIN” INTERVENTION

With this conceptual niche, the present scope narrows in several particulars. This section begins with a rationale for an OD response that includes demotion; later attention goes to various details of one useful design for demotion.

Why Be Ready for Rain?

All in all, demotion can be said to constitute a growing future emphasis in OD’s arsenal, and for at least five reasons. First, national economic recessions or depressions no doubt will always be with us—although perhaps with reduced dislocation, if we are skillful and lucky. Of course, some dour observers see such dark clouds as dominating the economic horizon. But virtually all observers see a future of persisting and increasing turbulence in which many job changes will be necessary.

Second, the value of demotion as a tactic will continue even during periods of economic affluence. Even then, demotion can be part of an organization’s options because, among other reasons:

Some firms or public agencies always will be experiencing bad times.
Any recovery might be shallow and short-lived, and demotion may supply flexibility—e.g., organizations could retain experienced employees for a quick push when the economic climate really improves.
Even in very good times, skills will be obsolescing, employee needs and ambitions may be lowered, and products and missions will change.

Each of these possibilities, as well as others, implies room for demotion among the options available to forward-looking organizations. If nothing else, a continued allegiance to globalism should maintain the supply of such opportunities and probably will increase them.

Third, contrary to the common wisdom, at least some organizations already make substantial use of demotion. In one industrial organization with over 50,000 employees, for example, about half of the managers and executives surveyed saw “at least a pretty good chance of their own demotion” some time during their careers (Goldner, 1965, p. 718). That author saw that bare fact as one of the most important findings in his study, given its divergence from general opinion about this personnel action. Relatedly, emotion shows up clearly in statistical analyses of careers (e.g., Rosenbaum, 1979). This incidence motivates trying to do better with demotion, obviously.
Fourth, substantial evidence implies that major improvements in demotion practices and policies are possible—indeed, are probably necessary (e.g., Hall and Isabella, 1985). In part, this potential derives from the multiple uses of the term “demotion.” Thus, More isolates eleven meanings of the term (1962, pp. 215–216). This does not reflect conceptual carelessness, however, but rather more a general embarrassment, lack of skill in doing a necessary thing, or a native cunning motivated by a fear of diluting or destroying motivation in the demotion process. Thus, Goldner dwells on “the ways in which organizations make demotion socially acceptable” (1965, p. 715), while Veiga concludes of his review of five demotions that “top management” has so sweetened the moves and given such misleading (though well-intentioned) counseling that only one of the five was perceptive enough, or perhaps secure enough, to recognize what had really happened (Veiga, 1981, p. 21).

Fifth, some substantial proportion of jobs is brand new every decade, and that proportion is certainly 10 percent and may be far higher. It is pleasant to think that any individual over a career of 30–40 years will always be promotable and/or trainable in providing necessary adjustments. That is a pleasant but unrealistic expectation, and it may even be a dangerous alternative at several points in a career (Golembiewski, 2000).

These five perspectives, then, contribute motivation to this chapter’s consideration of a pure-vanilla case of demotion—an open situation in which all in the organization were aware of the action, in which all demotees would experience loss of both status and compensation, and in which a deliberate effort was made to aid the coping of both management and demotees.

The following narrative sketches a simple design, first applied in 1972 for aiding in the adjustment of a national salesforce to a severe cutback, and since then applied on several occasions with similar effects. This narrative illustrates the useful effects of such a design, and it encourages raising a number of ethical issues associated with similar interventions in organizations. More broadly, it demonstrates that adverse personnel actions arouse powerful emotions* that can and should be dealt with, directly and as they occur, for the sake of both management and employees. In the context of this book, the demotion design demonstrates another approach to increasing OD success rates in two senses—by enlarging the arenas in which OD values can be applied and by a design that seems to have intended effects in a high proportion of cases.

**Purposes of Demotion Design**

This chapter focuses on an organization in which thirteen regional marketing managers were, as part of a major reduction in force, given a choice—demotion

---

* The point has been neglected, perhaps because it is so painful. For chilling details, see Slote (1977).
to senior salesperson or termination. The managers ranged in age from thirty-
three to fifty-five, they had been with the company from nine to twenty-four
years, and they had served as managers for periods ranging from sixth months
to seventeen years. All had received satisfactory performance appraisals. Most
of the demotees would suffer a major reduction in salary, in some cases as much
as $4,000 a year, plus other perquisites, if they chose to “pick up the bag, again.”
All were males.

Several forces influenced the decisions of the thirteen managers. The gener-
ous separation allowances available to those with seniority encouraged choosing
termination. On the other hand, the job market was chancy and the company was
considered a fine employer. So real counterforces discouraged leaving the firm,
even for those angry because of the adverse decision about them.

All but two of the managers accepted the demotion and, as Table 23.1
shows, they were given an early assignment intended to facilitate their making
the required adaptations as effectively and quickly as possible. It was referred
to as an “integrative experience,” and sought to counteract the apprehensions
induced by the demotions, with emphasis on the sharing of resources in a commu-
nity setting. The hope was that this would increase a demotee’s sense of mastery
over the consequences of his demotion and hence reduce the initial anxiety, de-
pression, and hostility likely to be induced by the personnel action.

The approach sought to reinforce previous developmental efforts in the
firm, which had invested in a major way in an OD effort that featured an off-
site sensitivity training experience as a major learning vehicle. The thrust of the
initial sensitivity training was to help organization members in two ways: in
developing attitudes and behavioral skills appropriate to OD norms, and in build-

ing those norms or values into their relationships at work. Eighteen of the twenty
participants in the demotion design—the eleven demotees and their immediate
supervisors—had such a learning experience, in fact. For them, that integrative
experience was one more extension into work of their off-site sensitivity training
sessions.

The demotion design was in effect an effort at the worksite to act on the
values emphasized in the off-site training sessions. Briefly, those core values
include:

Full and free communication
Expression of emotional as well as task-oriented behavior
Acceptance of conflict between the individual and the organization, and
coping with the conflict—willingly, openly, and rationally

The demotion design is an “integrative experience” in several senses, in
the context of such OD values. Obviously, it brought new supervisor-salesperson
pairs together. Moreover, it sought to relate feelings and action: feelings would
be expressed and worked through, if possible. The working symbolism was the
<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 6</th>
<th>Day 7</th>
<th>Day 45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thirteen managers informed of choices: demotion to salespersons or termination</td>
<td>Decision required Eleven managers accept demotion</td>
<td>Three major activities: 1. Demotees and supervisors respond to MAACL * pre-test</td>
<td>Two major activities: 1. Demotees meet individually with their new supervisors</td>
<td>Tests of persistence changes: long post test</td>
</tr>
<tr>
<td>Employees understand that choice of demotion will require reporting to a Midwestern city for an integrative experience along with new supervisors.</td>
<td></td>
<td>2. Demotees spend balance of day in discussion</td>
<td>2. Demotees and supervisors respond to MAACL * short post-test</td>
<td>Demotees and supervisors respond to MAACL *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Supervisors have briefing meeting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Multiple Affect Adjective Check List (Zuckerman and Lubin, 1965).
cauterization of a wound, perhaps painful but useful and even necessary. In addition, the design sought to avoid obsessiveness and the postponement of facing new work demands, which over time would probably have added to the negative emotional impact of demotion. In the absence of such emotional support—coming from demoted peers, superiors, and the employing organization—negative effects were probable, especially for the more senior men. Were depression to occur, the best-available information warned interveners and management, its effects might “include insomnia, loss of appetite, excessive worrying, indigestion, and decline in energy” (Kiev, 1969, p. 2).

The integrative experience also provided an opportunity for all involved to obtain early data about possible adaptive difficulties—managers, demotees, and a half-dozen consultants who included outsiders as well as in-house representatives from the firm’s human resources unit. The latter were essential since they could provide help if either supervisors and demotees later experienced difficulties at work due to the demotion processes.

**Characteristics of Action Design**

The design had two learning phases. For roughly 50 percent of the time, demotees worked together. Later, individual demotees attempted to work through issues of concern with their new supervisors.

**First Design Components: Diverse Reactions to a Common Fate**

All demotees spent about four hours together discussing mutual concerns and needs, with two consultants available. The intent was to harness emotional energies to organization purposes rather than to merely diffuse them through sheer ventilation. Several important themes were dealt with, beginning with personal reactions and moving toward work-related issues. The process surfaced such themes, among others:

- Comparing experiences, especially about the diverse ways in which various relevant organizational policies were applied in their individual cases;
- Encouraging expression of anxiety or hostility about the demotions themselves or about associated processes—for example, their style or timing;
- Surfacing and resolving (if possible) suspicions regarding management, such as the concern that another personnel purge was imminent;
- Isolating and, as much as possible, working through demotees’ concerns about authority and dependence, as in their complaints that they were not being treated as adults, or that they were “strong enough” to take the demotions without the integrative experience;
- Dealing with a variety of issues in work relations—for example, how to explain the demotions to clients or other salespeople—in order to develop strategies and norms that would reduce the probability of avoiding issues or handling them awkwardly in the field; and
Identifying specific relevant others with whom interaction had been stressful or with whom it might prove to be so, for the purpose of developing strategies for handling such interactions.

Consultants intended to facilitate expression of feelings and reactions, to help reveal the diversity of the demotees’ experiences and coping strategies, as well as to help work toward a successful adaptation to the demands of the new job. In short, the consultants directed attention to both process and content, to use a convenient distinction. In their attitudes, the consultants were neither advocates of management actions nor emotionally neutral in response to the dynamics of the demotions. By conscious decision, they sought to remain in-between and to respond flexibly as individual situations warranted.

The consultants believed some demotees might decide to accept termination after the integrative experience, and that option was emphasized as open, both early and late. This is consistent with the view that OD should increase responsible freedom. But none of the eleven participating demotees took advantage of the offer. In this case, that openness of choice meant multiple opportunities to choose, as an in vivo exercise of the power that demotees did have in a situation in which they might otherwise feel powerless. Virtually all of the managers-becoming-salespersons emphasized the positive meaning of the integrative experience, but one participant derided it as “hand holding” and “coddling.”

The first component of the learning design emphasized some common elements among the demotees, as well as some differentiating factors between them. The former included the impact of demotion on self-image; experiences with important referents, such as spouses, colleagues, or salespersons from other firms; and concerns about taking on the salesperson’s job—“picking up the bag again” to cover a sales territory, participating in sales meetings with peers they have previously managed, and so on. These elements contributed to a sense of shared concerns, as well as building a learning community.

The design’s strong emphasis on reality-testing required a basic challenge to this sense of shared interests. Prominent among the differentiating features was the fact that the demotees included both long-service employees and recent managerial appointees. On balance, the future for the longer-service men was far less bright. Some men professed shock at being confronted with the choice of demotion or termination, in addition, while others maintained they had more or less expected some action, especially because of falling demand in the industry. A few even expressed pleasure that the adverse action did not hit them as hard as it had many others affected by the major reduction-in-force.

Second Design Component: Developing New Supervisory Pairs

The second major design feature involved the reintegration of the individual demotee into a specific worksite. Hence, this second design feature has strong implications for productivity and organizational effectiveness.
Two approaches were taken to building on the first phase, which sought to help demotees reduce their general anxiety as well as to sharpen their more specific concerns about developing new working relationships. First, the demotees’ new managers met for two hours to discuss their role in the design for the next day—which they learned was to feature individual meetings with the demotees. The basic thrust was to sensitize managers about how the demotees felt, so as to help by suggesting ways of channeling those feelings toward making the most successful adaptations possible. The total sense of it was to increase managerial ownership in the design.

Second, the demotees spent approximately three hours with their new managers in one-to-one situations. The consultants sat in on these several meetings, as time permitted. The major concerns in these one-to-one situations included:

Building early supervisory relations, as in mutual pledges to work harmoniously together, which was easiest in those several cases in which demotees were able to choose their new manager;
Discussing technical problems, such as going over sales territories;
Developing strategies by which the manager and salesperson could be mutually helpful, as in discussing ways to moderate the formation of cliques, which the demotions might encourage; and
Isolating likely problems and cementing a contract to agree to meet any problems rapidly and mutually.

Some one-to-one meetings concentrated on a single concern, whereas other meetings dealt with several. Throughout, regenerative interaction was sought.

Measuring the Design’s Consequences

The effects of the action design are judged by changes in the Multiple Affect Adjective Check List (MAACL), developed by Marvin Zuckerman and Bernard Lubin. The MAACL is an instrument for tapping the psychological aspects of emotion, and posits affect as a state, not as a trait. That is, a time referent is specified for respondents, who react as they feel “today” or “now” rather than “generally” or “occasionally.” Zuckerman and Lubin describe the MAACL in these terms:

[It] was designed to fill the need for a self-administered test which would provide valued measures of three of the clinically relevant negative affects: anxiety, depression, and hostility. No attempt was made to measure positive affects, but some of the evidence indicates that the scales are bipolar, and that low scores on the full scales will indicate states of positive affect (1965, p. 3).

The intervenors had direct expectations about MAACL scores. The demotees were expected to have high initial scores on anxiety, depression, and hostil-
ity, which a successful intervention would reduce significantly. The managers were expected to have lower initial scores on anxiety, depression, and hostility, and the post-treatment administration of the MAACL was not expected to reveal any major shifts, except perhaps on anxiety. This anxiety about role was expected to diminish for the supervisors in one or both of two ways: as managers came to influence and own the design, or as a positive experience with the design developed.

The MAACL was administered three times. One administration occurred just before the integrative experience, and a second immediately after. The third administration was by mail approximately a month after the intervention, so as to test the persistence of any changes. One of the twenty-two participants did not respond to the third administration.

CONSEQUENCES OF A CRITICAL INTERVENTION

The effects of the intervention are summarized in two ways. A first section provides a brief review of the measured consequences. A second urges caution about assuming too much and relies on a broad range of evidence.

Five Perspectives on What Happened

What effects did the intervention have? They were almost entirely in the expected directions, as Table 23.2 suggests. For the sake of simplicity, only gross trends are reported, but the discussion will emphasize all statistically significant changes.

<table>
<thead>
<tr>
<th>Table 23.2 Overall Effects of Intervention on Three MAACL Administrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Scale Scores (by days after demotion)(^a)</td>
</tr>
<tr>
<td>Day 6</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td><strong>Demotees</strong></td>
</tr>
<tr>
<td>Anxiety</td>
</tr>
<tr>
<td>Depression</td>
</tr>
<tr>
<td>Hostility</td>
</tr>
<tr>
<td><strong>Managers</strong></td>
</tr>
<tr>
<td>Anxiety</td>
</tr>
<tr>
<td>Depression</td>
</tr>
<tr>
<td>Hostility</td>
</tr>
</tbody>
</table>

\(^a\) Since each scale has a different score, interscale comparisons should not be made. A lower score reflects less of the emotions in all cases.
Five points detail the major results. First, the demotees reacted strongly to demotion. Without going into technical details (e.g., Lubin and Zuckerman, 1969), each of the three psychological states referred to in Table 23.2 has a different maximum. How high were those scores? About one-third of the scores for demotees on the initial MAACL administration surpassed or closely approached the 98th percentile of scores of a large number of past respondents. So the initial MAACL scores for demotees in Table 23.2 are very high but not off-the-scale. The scores for managers approximate normal levels, with a marked elevation only for anxiety.

Since the demotions were announced about six days before the first administration of the MAACL (see Table 23.2), the initial scores suggest that the demotions had a great and persisting impact that was unlikely simply to wither away. No data exist on this important point, but most scores were near the maxima. So the six intervening days did not do much to dampen reactions to the personnel action.

Second, as expected, the initial scores of the managers are significantly lower than the demotees’ initial scores in all cases. Managers’ scores for anxiety also decreased significantly twice—at day 7, the last day of the learning design, and again at day 45—and this implies that the managers were not upset by the integrative experience. Hostility scores rose, but not significantly.

Third, the effects of the design are consistent. The effects vary only randomly for employees who differ in age, years with company, years as regional manager, and loss of salary involved in the demotion. Details are available elsewhere (Golembiewski et al., 1972).

Fourth, the effects on individuals also establish the efficacy of the design. Overall, about 80 percent of all MAACL scores for demotees are reduced, and an additional 10 percent are unchanged, when the initial scores are compared with the second and third scores for each individual. In addition, no demotee reports an increased score on more than one scale. Overall, all average scores for demotees decrease significantly between days 6 and 7, which were the days of the learning design intervention, and the reductions are maintained through day 45. In fact, anxiety scores on the average are significantly lower, comparing days 6 and 7 as well as days 7 and 45.

Fifth, subsequent history suggests the value of the design. All but one of the demotees were on the job three years later, and when interviewed, they reported vividly and positively on the integrative experience. Moreover, five had been re-promoted when the firm’s markets improved. No strict efforts were made to track the demotees beyond the three-year mark but, essentially, the positive record continued. But note that the effects of any intervention probably should not been seen as being retained through all time, with three years representing a substantial persistence. 
Avoiding Foolish Claims

The results abstracted here could be elaborated, but they come to the same point. The design seems to work, and has done so several times. The initial success of the learning design undoubtedly profited from earlier attempts of the host organization to develop supporting attitudes and behavior via OD efforts. Hence, this design may not be applicable in organizations as a first-generation effort. This was a serious consideration when the design was first applied, but the seriousness diminished with the success of other applications that did not have the benefit of such advanced work.

Nonetheless, intervenors should not be incautious. Methodological inelegances prohibit attributing the effects uniquely to the learning design, even though the presumptive evidence is strong, and especially so after several replications have generated similar results. For example, some may attribute the initial reductions in anxiety, depression, and hostility to the passage of time. But the interval between the first and second MAACL administrations was a brief one, and the demotees had patently developed and sustained high scores on the three target variables over the six days between the demotion notices and the intervention. So no easy belittling of the results seems appropriate.

GREATER SUCCESS BY UTILIZING DEMOTION DESIGN

So what can be reasonably concluded about the demotion design on its variants relevant to applications of OD? Two separate points will be emphasized—one dwells on the empirical consequences, and the second focuses more on the moral or ethical considerations. The latter need to undergird applications of the demotion design if they are to rise about merely serving some establishment wishing to rid itself of a ticklish problem.

Probable Empirical Effects

As the findings suggest, the demotion design seems to have benign or positive effects, on balance. A number of replications of the design, using different resource persons in various settings, have had similar effects. This reduces the credibility of two major alternative explanations of effects.

So the demotion design may not be an ideal one, but it seems useful for moving toward OD values. These OD values emphasize human responses to human problems—creating reasonable choices for people that permit them to exercise greater responsible freedom at the website. The design had those effects for both employees and management, who in their different ways saw no attractive choices and yet who were uncomfortable about the firings. These uncomfortable feelings tend to be turned around by the several choices permitted by the demo-
tion design. In effect, the choices give participants the opportunity to commit themselves to the design, and hence to develop a real sense of the ownership of results.

**Moral and Practical Concerns about Cooling Out**

The demotion design does not generate effects in the abstract, but in the context of distinct moral and ethical contexts that deserve extended attention. These moral and ethical issues have more to do with effectiveness than efficiency, of course. But an OD that fails to be alive to its value-filled character probably will not remain successful for very long, and always should be suspect. The demotion design thus serves a useful purpose in providing a context for raising these moral and ethical issues.

Efforts like the demotion design raise tangled ethical and practical issues about which OD consultants clearly need to be concerned. Moreover, these issues also need attention from that broad range of decision makers who might consider or authorize interventions like the demotion option—executives and managers in all functional areas, human resources directors, and so on. Four points relevant to these broad-impact ethical issues deserve highlighting.

**Value Issues Should Be Up-front**

If organizational interventions are seen as a Band-Aid, as an after-the-fact ameliorative effort to salve the wounds of any actions management decides to take, then the organizational intervenor faces a range of serious ethical and professional issues. He becomes a “cooling-out” functionary, a person who merely dissipates the hostility that might otherwise possibly serve as a force for constructive change. In the long run, the intervenor’s boss in this way can despoil a useful resource.

The best way of avoiding the Band-Aid role is to establish when interventions can make a difference, and to restrict applications to those situations. Facilitative situations include those where:

- Management has a real stake.
- Management perceives that only traditional options exist.
- Employees have real choices and can exercise them.
- An alternative has a real probability of working in an exchange sense: that is, the intervention helps management achieve its needs, on balance, while it also serves employee needs, within a normative context defined by the values underlying OD.

The consulting team saw all those conditions in the integrative experience described above.

Some authorities in organizations may desire only a reduction in conflict,
of course, not an increase in employee options. The point here is that, as the intervenor’s ability to generate real options is limited, so also will his or her usefulness decrease in helping deal with conflict. Many managements seem to be clearer on this point nowadays (e.g., Golembiewski and Kiepper, 1988, pp. 216–226) but war stories to the contrary exist, and OD as a profession has only begun to take the first steps in protecting practitioners from undue influence by management (Golembiewski, 1986; Golembiewski, 1988, esp. part 3).

Cooling Out, or Increasing Options?

The context of the present intervention created real confidence that this consulting team was not merely stabilizing the system, and thus subordinating the interests of the demotees to those of the organization. Specifically,

The demotions were part of a broader reduction-in-force delayed by a management that for several years had vainly sought new products to occupy the full salesforce.

Overstaffing, due to rapidly changing market conditions, was an objective dilemma that required confronting.

Three consultants served various levels of management over the extended period during which policies and procedures for the reduction-in-force were developed.

The demotion experience was suggested by the consultants in response to the expressed managerial concern that, unfortunately, there was no alternative to firing thirteen managers.

The consultants basically saw themselves as enlarging the options open to both managers and demotees, both in substantive and, especially, in process terms. Substantially, at a number of points the design provided several opportunities for specific choices by both management and demotees about whether or not to continue. Two of the thirteen men immediately chose termination in preference to demotion at the first of these decision points, which suggests that choice was real. Process enhancement was considered more significant by consultants, however. The purpose was to induce regenerative interaction between demotees, and especially between the members of each demotee-supervisor pair. Such enhanced interaction intended to raise the probability that demotees could really handle choice and change—that they had valid and reliable data, and that they got appropriate help in group settings to make decisions that dealt with central issues. This contrasts with decisions that came unwound or, worse yet, generated worse problems than those they attempted to solve.

SOME CENTRAL CONCERNS

In process terms, the basic decision was to build the design in a conservative way—from lesser discrepancies to probably greater discrepancies. Hence, the
integrative experience began with the demotee meetings, where some differences between individuals were expected but greater commonalities of experience were anticipated. Then the design escalated to encompass supervisor-demotee pairs, where complex discrepancies were expected—between individuals as well as between various areas of an individual’s sense of the situation. See also Chapter 3.

Real Exchange or Tenuous Illusion?

One might argue that the demotees were in no position to engage in meaningful exchange with management and, consequently, that the demotees were powerless pawns. In this view, the demotees could hardly do other than to feign a positive response to an integrative experience that was in reality forced on them.

The consulting team had no doubt, now or earlier, that meaningful choices did exist for the demotees, and their thought processes are revealing. Four factors are most prominent in the case for the position that meaningful choices did exist.

All demotees had been rated satisfactory performers, and all or most saw the action against them as due to market conditions, rather than poor performance.

Management valued the managers’ past contributions and did not relish the negative impact on morale that was expected to follow the outright termination of managers with substantial seniority. But management could initially see no way to retain the eventual demotees without creating managerial situations that they unanimously saw as both unattractive and unavoidable.

Most of the demotees, given a choice, preferred staying with the firm despite real losses in money and status.

The managers’ work experience might later be valuable when market conditions became more favorable.

In the present case, the re-promotions mentioned above represent perhaps the best indicator that more than a tenuous illusion of exchange existed. A real basis for exchange might not have existed if, for example, the demotees had been marginal performers. But they all had been satisfactory performers.

Multiple and Shifting Clients

Multiple and shifting clients clearly existed in the case under discussion. The initial client was marketing management, when the consulting team served as process consultants, and also recommended the demotion design to them. The OD team also was in the long run held responsible by management for the effects of the intervention. Management’s expectation was that most of the demotees (and perhaps all of them) would experience major coping problems, but past
successful experiences with the OD team—coupled with a deep, visceral distaste for the cutback and its human consequences—motivated accepting the consultants’ recommendation for demotion. But the consultants’ reputation clearly was at stake.

The OD team’s client was neither singular nor stable. Once the eleven men accepted demotion, for example, they became the focal clients. Later, when the immediate supervisors of the demotees were present, they also became part of the client population. The data suggest that the OD team served the immediate interests of the supervisors less effectively, less adequately helping them to deal with their anxiety in a situation that was novel and initially somewhat threatening to them.

Two points summarize the consulting team’s biases in dealing with multiple and shifting clients. First, management was encouraged to apply control through results rather than through knowledge of specific details. Such specification of behavioral or attitudinal objectives may seem to put the proverbial noose around the intervenor’s neck, but it has major redeeming virtues. Without such measurement of outcomes, management might be tempted to seek detailed information about “what went on,” with serious implications for trust and learning.

Moreover, whether or not clients might be multiple and shifting, and whether or not client interests are starkly opposed, the intervenor’s basic allegiance is to a relatively clear set of OD values and orienting perspectives. The allegiance should be articulated early and should be preserved even if all else is subject to compromise, blurring, or selective application. To do otherwise is to risk being considered a cooling-out functionary, and deservedly so.

From management’s point of view, of course, this flexibility for intervenors did not, and will not, come scot-free. Only success in producing desired effects did, and should, motivate the necessary managerial adjustments. Hence the criticality of substantial OD success rates, and especially of efforts like the present one to heighten those success rates.

GREATER SUCCESS VIA EASY PIECES, REVISITED

In sum, the demotion design illustrates the usefulness of an applied behavioral science approach to a grave organizational issue. Acting on explicit values, it proves possible to help meet the needs of employees. Simultaneously, management husbands experienced resources for better economic times, while adding substance to the bold organizational proclamation: “People count.” In addition, essentially similar replications have had similar effects. The focus below is on several senses in which the demotion design can contribute to greater success in OD. The discussion above details several significant caveats, and these should tether any overexuberance in what follows.
Paramountly, perhaps, the leverage for cost-effective outcomes seems high, though appropriate calculations pose formidable challenges. Of course, the narrow, intended effects do occur—levels of anxiety, depression, and hostility are not only reduced, but the reductions persist after what seems a precipitous jump. But such effects are certainly too selective, even if representative. In the present case, the past contributions of employees were acknowledged in a very direct way, which the grapevine saw in ways suggestive of future commitment and trust. Moreover, valuable experience was retained for use when the economic times improved, a possibility that was realized for almost every ex-manager in the present population. Economic times did improve, sooner rather than later, and five of the voluntary demotees were subsequently re-promoted and all but one continued working for at least 5 years beyond the demotion.

**Heightening Success in Four Senses**

The demotion design also seems capable of contributing to greater OD success in several broader senses, only four of which get attention here. First, the design adds to the humanistic repertoire of responses relevant to an important type of situation, for organizations as well as for OD. That is, the demotion design vivifies regenerative interaction at a time of mutual need: the design rests on high openness, owning, and trust, as well as low risk, and the design also seems likely to enhance any existing tendencies toward degenerative interaction. In contrast, under threat, people and organizations tend to resort to closedness and coercion. In this sense, the demotion design illustrates OD values in a situation in which it is common to act in opposite ways (Sutton, Bruce, and Harris, 1983).

Second, the demotion design offers both individuals and organization authorities an opportunity to make real and responsible choices. Although extensions of freedom may seldom accompany crisis situations, the demotion design proposes just such an extension. For employees, to illustrate, the design calls for several explicit choices: a choice between the learning experience or termination, a choice between trying to deal constructively with a negative personnel action in community and with organization support, or alone with personal resources, and so on.

Third, the demotion design empowers OD professionals to “be there” at a critical time. This raises their organizational credibility, with management as well as with employees. But the capability also no doubt contributes to the self-esteem of OD practitioners and to their personal growth, as well as to OD’s development as a corpus of values with a technology.

Fourth, the demotion design tests the values and core technology of OD in a crisis situation. There is no alternative to such risk-taking, if one seeks flexibility and a broadening range of effectiveness.
Demotion as Empowering Potential 563

The failure to employ designs like this demotion effort, then, is multiply lamentable. Even though this OD application fits a narrow niche, it relates to a critical time—for employees, their managers, and their employing organizations. And it offers the potential for doing something of value, when that is much needed.

REFERENCES


Defining “Change” as Trinitarian
Estimating Whether “Change” Occurs, and How Much, Requires Specifying the Kind of Change

A truism underlies this chapter. That is, you have to define “change” not only before you can estimate whether or not it occurs in a specific case, but also before you can estimate how much of “it” occurs. In this strict sense, then, the preceding chapters should be viewed tentatively, for neither the several researches nor this treatment yet have been specific about kinds of change. The time has come to begin to rectify that major shortcoming, if only in preliminary ways.

A CONCEPTUAL CONTEXT FOR IRONY IX

If this reasoning is correct, almost all applied research has a serious inadequacy. “Change” typically gets very general treatment, which is bad enough. Moreover, there may well be several kinds of change. So the definitional problems may be quite serious, as can be their implications for anyone concerned with change in human affairs. ODers need to pay close attention, in short.

This chapter urges distinguishing among kinds of change, distinctions that are suggested by experience and that also are supported by evidence generated with exotic statistical and computational techniques (Golembiewski, Billingsley, and Yeager, 1976a; Golembiewski, 1986, 1989, 1997). An immediate payoff of making such distinctions is more definite reliance on existing research findings, whose interpretation is necessarily related to an underlying concept of change. Indeed, as will become plain, no interpretation of much behavioral research is possible without a determination of what kind of change—if any—has occurred.
This seems a radical statement and implies the gravitas of what this chapter has in mind. More central still, the goal is to facilitate the design and evaluation of efforts that seek to improve the human condition and the quality of life, especially in organizations.

Specifically, this conceptual clarification of “change” distinguishes three kinds. Later, data from a study of a successful Flexi-Time intervention will be used to test these conceptual elaborations. Detailed statistical analysis will support the broad position that a unitary concept of change is not only inappropriate but may be seriously misleading.

The goal here is direct. By extending the boundaries of the known, applied research can better contribute to the further development of scientific knowledge, as well as to the fulfillment of its stated goal of improving the immediate human condition. But applied research contributes in both senses only to the degree that one can ascertain whether a particular intervention succeeded or failed. This knowledge not only requires measuring the quantity of change but especially demands confidence in the concept of change that underlies its measurement.

Our focus here is on what is measured in experimental designs, such as those in OD, especially in those designs placing heavy reliance on self-reports. More specifically, this chapter deals with the paradox underlying a dilemma emphasized by Bereiter, who asks:

When scores on a test are observed to change, how can one tell whether it is the persons who have changed or the tests? If the correlation between pre-test and post-test is reasonably high, we are inclined to ascribe change scores to changes in the individuals. But if the correlation is low, or if the pattern of correlations with other variables is different on two occasions, we may suspect that the test does not measure the same thing on the two occasions. Once it is allowed that the pre-test and post-test measure different things, it becomes embarrassing to talk about change. There seems no longer any answer to the question, change on what? (1963, p. 11). [emphasis in original]

By discussing change in OD contexts—via the application of modern technologies for data processing and analysis—this chapter accepts the challenge of Bereiter’s central question. This chapter seeks to provide perspective on persistence and change in human affairs.

Change on what? Change for what? These questions have serious implications for the OD practitioner, who cannot avoid them but usually does. Such value-laden questions get short shrift in technically oriented treatments or as a result of self-interest, as Ross (1971) reminds us forcefully. Yet OD ideologues early emphasized that OD interventions are value loaded (Tannenbaum and Davis, 1970; Golembiewski, 1972, esp. pp. 59–110). The consciousness grows that OD interventions should be less involved with raising the level of indicators
of some relatively stable system than focused on basic change in the quality of organizational life that should and can exist (e.g., Golembiewski, 1988, part 1).

HOW TO ECONOMICALLY BEGIN THIS EARNEST SEARCH FOR "CHANGE"?

Our point of entry to confronting the central, indeed crucial, complexities of change will be conceptual. After defining the three types of change in summary fashion, we further distinguish them by examples. The basic conceptual distinctions follow:

**Alpha change** involves a variation in the level of some existential state, given more or less constantly calibrated intervals for a measuring instrument related to a constant conceptual domain.

**Beta change** involves a variation in the level of an existential state, complicated by the fact that some intervals of the measurement continuum associated with a constant conceptual domain have been recalibrated.

**Gamma change** involves a redefinition or reconceptualization of some domain—a basic change in the perspective or frame of reference within which phenomena are perceived and classified, in what is taken to be relevant in some dimensions of reality.

**Alpha Change**

Most OD discussion, in unison with most approaches to change, recognize only alpha changes, measured by self-reports, and often use pretest/posttest designs, with or without comparison groups. Symbolically, such designs may be described as $O_1 - X - O_2$, where “O” stands for observation and “X” stands for the OD intervention. That is, the typical design selects some frame of reference or criterion, with change being estimated by fluctuations in the levels of self-reports assumed to be triggered by the intervention.

Alpha changes, then, are conceived as occurring along relatively stable dimensions of reality that are defined in terms of discrete and constant intervals. Note that alpha changes can be nonrandom, as established by some test of statistical significance, or they can be random only. And alpha changes may be very large or very small, or anywhere in-between. The only requirement is that the alpha change occur within a relatively fixed system or state, defined in terms of stable dimensions of reality as estimated by a measurement continuum whose intervals are relatively constant. For example, a parent taking a baby to a shoe store is interested in alpha change. The parent’s frame of reference is growth in the baby’s feet between this visit and the preceding one. The crucial measurement of change occurs within relatively fixed dimensions of reality (our conventional concepts of length and width), as defined by indicators whose intervals are more-
or-less constant (the calibrated marks on the measuring rod against which the baby’s foot is compared). $O_2$ minus $O_1$ measurements estimate the magnitude of the interim change in units of length and width.

**Beta Change**

These changes involve the recalibration of some portion of the intervals used to measure some stable dimension of (for example) psychological space in pre-intervention versus post-intervention responses. This contrasts with alpha changes, which are measured along more-or-less invariant intervals tapping stable dimensions of reality.

Again, back to that baby’s feet. If in a post-test those feet fall in that range within which a beta change had occurred, the parent could not know how much the baby’s feet had grown between visits to the shoe store, or if they had grown at all. It would not be meaningful to compare the two measurements because some intervals on the measuring rod had somehow changed. Of course, beta change could occur for intervals along distant components of the measuring scale, and that would not be relevant to the practical issue of growth in the baby’s feet even though that beta change could be crucial for other purposes.

Beta change is perhaps especially a problem in the social and behavioral sciences, although analogs in the physical sciences have existed. A beta change on a rod for measuring feet is not very likely, for example, although such rods do expand and contract some. Social measuring rods can “expand” and “contract” significantly, however, even as their conceptual definition remains the same. Note that the reference here is to a phenomenon beyond test-retest reliability. For change in the measuring intervals is often an intended effect of an OD intervention, as contrasted with some defect of the measuring instrument. That is to say, OD efforts indeed can change the very measuring instrument being used to estimate the change. This complicates interpreting $O_1$ and $O_2$ estimates, of course.

Put too simply, perhaps, instruments soliciting self-reports are potentially more like rubber yardsticks than they are like the “standard foot” in the Bureau of Standards. That is, self-reports are rooted in socio-emotional or cultural definitions and also can be significantly modified by changes in an individual’s knowledge or experiences. In this sense, applying the same instrument before and after a successful OD intervention—while assuming that the intervals along which self-reports are the same or very similar—may be rather like applying a given survey of opinion to several different cultures, as conventionally understood. Any resulting “findings” must be compared and analyzed very carefully (Ward, 1974, esp. p. 199). Recognizing this awkward effect underlies the common recommendation in recent years to perform “confirmatory factor analyses,” to raise confidence that $O_1$ versus $O_2$ measures are sufficiently similar to justify direct comparisons. In percent terms, “confirmatory factor analyses” are intended to establish
whether or not only alpha change has occurred. Or to put it another way, such analyses test whether or not beta or gamma changes have occurred.

To some, it may seem like splitting hairs to distinguish beta change from alpha, because both deal only with changes in condition within a relatively stable state. But beta change does point up a significant and generic problem in interpreting behavioral research, as an extended illustration will establish. Consider the two sets of descriptions of an organization unit in Figure 24.1, pre-test (designated “Now I”) and post-test (“Now II”) bracketing an OD intervention. In a first hypothetical case, assume that only alpha change has occurred. Even here, Figure 24.1 does not support a single or simple conclusion. The Now II mean score is consistent with an OD failure in one sense, as well as with a success in another. The OD intervention may be said to be a failure because Now II is lower than Now I, and OD interventions should induce changes toward system IV. Or, the OD intervention can be taken to be successful because the respondents at Now II have a more realistic view of how things really are, a firmer descriptive base for subsequent ameliorative action.

Matters get more complicated if the possibility of beta change is acknowledged. Consider employees whose average pattern of responses following an OD intervention was like that in Figure 24.1 (Golembiewski and Carrigan, 1973). Yet things in their organization units were not worse than before the intervention, respondents reported in interviews supplementing their questionnaire responses. In fact, they reported that things had substantially improved. In effect, the OD

![Figure 24.1 Pre- and post-test means on a representative item from Likert's Profile of Organizational Characteristics.](image)
experience may have encouraged respondents to recalibrate some of Likert’s intervals after the intervention in at least two ways: respondents made different estimates of reality, given a clearer perception of what exists; and respondents may have reassessed their intervals for measuring value-loaded terms in the instrument, such as “throughout the organization,” “well-integrated,” and “overlapping.”

We hypothesize that the OD experience had “lengthened” the psychological space between some intervals of the Likert instrument while preserving the essential conceptual content of Likert’s “managerial systems.” The OD program, in effect, showed respondents how much integration there was and could be, and they recalibrated relevant portions of the system III and IV portions of Likert’s scale. One respondent added, “I don’t need to be educated about what a ‘3’ score is! I’ve seen that often enough.” Likert’s intervals are like rubber yardsticks, subject to expansion—along a narrow interval or throughout the range—and contraction as personal and group standards are impacted by the OD intervention. Consequently, even though respondents verbally report “more participation” at Now II, their post-intervention scores are lower than for Now I. The content of the Participative Group interval had stretched further than actual participation had increased, as it were. For the technical development of a similar notion, see McGee’s (1966) emphasis on “elastic distances” in multidimensional scaling.

Presumably, if the OD intervention had involved experience with an equivalent of a POW camp, respondents would have recalibrated the system I portion of the Likert Profile.

Numerous other issues relate to, and exacerbate, the interval problem highlighted by beta change. For example, the “response instability” that has sometimes been taken to signal a “non-attitude” in political research (Converse, 1970; Iyengar, 1973) could in fact reflect a beta change. The difference is critical: non-attitudes can be treated cavalierly as opinitional ephemera; a beta change, in contrast, signals subtle differences in the intervals a rater uses to differentiate a given psychological domain. Other significant issues (Pepper and Prytulak, 1974) also seem relatable to beta effects.

**Gamma Change**

This kind of change is conceived as a comprehensive shift in ways of conceptualizing salient dimensions of reality as well as in intervals for estimating their degree. This totally differentiates it from beta change, which refers only to variation in some of the intervals measuring a relatively stable dimension of reality.

This third kind of change involves the basic redefinition of the relevant psychological space as a consequence of an OD intervention. In sum, gamma is “big bang” change. It refers to a change from one state to another, as contrasted with a change of degree or condition within a given state. Thus, “freedom” for
Afro-Americans in 1960 may have been defined, in part, as not having to ride in the back of the bus. By 1970, such freedom in Atlanta seems to have expanded to include success at lowering bus fares, increasing the number of minority drivers, and impacting on the design of urban mass-transit systems. Identifying gamma change is crucial because the pre-intervention instrument is no longer appropriate. The post-intervention response is not only off the scale; it is on a different scale estimated by different intervals.

Thus, if gamma change occurs as the result of an OD program, interpretations of results are chancy in the extreme, and research takes on an Alice-in-Wonderland quality. For example, issues of instrument validity become enormously complicated when phrased in these terms: Valid for measuring which kind of change?

Bowers’s study of fourteen thousand respondents in twenty-three organizations helps illustrate the importance of distinguishing gamma from beta and alpha changes. Among other tendencies, Bowers reports that survey/feedback interventions are associated with “statistically significant improvement on a majority of measures” based on The Survey of Organizations Questionnaire (TSOQ), whereas laboratory training interventions are “associated with declines” on similar measures (1973, p. 21). In the absence of knowledge about the distribution of types of change, however, it is not possible to conclude that Bowers’s results demonstrate anything, let alone the greater potency of survey feedback.

Alternatively, survey/feedback may have triggered alpha change, which TSOQ picked up, while lab training induced gamma in TSOQ. This interpretation is consistent with the different “depths of intervention” associated with survey/feedback and laboratory training, respectively. Moreover, Bowers’s instrument does seem more sensitive to alpha changes, as all such instruments must be. Furthermore, TSOQ is based on factorial studies of organizational samples (Taylor and Bowers, 1967) other than those in the 1973 study, but we do not know to what degree the 1973 factorial solutions are congruent with the 1967 baseline solutions. Nor is it known what degree of congruence exists between factorial solutions of pretest versus post-test responses to TSOQ in the 1973 sample, which could provide a clue about the possibility of beta or gamma change.

Along with other concerns (Torbert, 1973), then, interpretation of Bowers’s results is made problematic by his failure to specify kind of change. In the case of both survey/feedback and laboratory training interventions, that knowledge is profoundly significant in interpreting Bowers’s results.

Typical descriptions of OD that seek to induce a new social order or culture in an organization reflect the fact that gamma changes are not conceptual niceties: instead, they are the prime intended consequences of OD interventions. Gamma implies not only the recalibration of intervals but also new content for concepts describing the quality of organization life. Hence, gamma change can be distorted or disguised by common measuring instruments whose conceptualization and
operationalization are typically rooted in alpha change, as via Cronbach’s well-named (from the perspective of this analysis) alpha estimate of reliability.

Since gamma change may be thought of as reflecting fundamental changes in conceptualizations or expectations, as well as a basic redefinition of the intervals of the referents tapped by measures of organization and individual processes, it severely complicates the interpretation of the results of OD efforts. Most probably, the general failure to identify gamma change—and beta, as well—results in conservative estimates of OD outcomes. These estimates are distorted or camouflaged by beta and, especially, gamma change. If this is so, Chapters 1 and 2 could well deserve having their success rates adjusted upward.

ANALOGIES FROM SEVERAL SCIENCES

Other considerations also encourage the search for such beta and, especially, gamma effects. First, and very briefly, a substantial feature of recent advances in the physical sciences has involved a basic conceptual distinction between changes in condition and changes in state, between what are here called alpha change and gamma change. Consider complex homeostatic systems, for example. They may experience a bewildering variety of changes in their conditions, in highly variable order, and yet preserve their essentially steady state (Ashby, 1954, 1956). On the other hand, common wisdom acknowledges that systems can sometimes be at such a developmental point that even minor changes in condition can induce a profound change in state: hence the expressions “the straw that broke the camel’s back,” “the critical incident that induced a psychotic reaction,” or “the push we needed to get over the hump.” Failure to distinguish the two kinds of change implies inadequate description and, possibly, encourages dangerous prescriptions for action. Making the distinctions correctly promises to conserve much energy and resources.

Convenient analogies also imply the ubiquitous character of the distinction between change in condition within a state and change in state. Consider four known state of H₂O, simplified to consider differences in temperature only—solid, liquid, gas, and (perhaps) plasma. H₂O will remain in one of the states over a considerable temperature range induced by a substantial gain (or loss) of calories. Water in its known forms, in sum, can experience major changes in condition without a change in state. At critical temperatures, the addition or subtraction of even a small number of calories can induce a change in the state of H₂O with little or no effect on its condition as measured by temperature. In sum, a stepwise model applies. The condition (temperature) is linear. Each of the states of H₂O persists over a substantial range of conditions, but “jumps” to different states can occur at certain critical points. The general notion is consistent with the sense of “punctuated equilibria” in geological or in biological development.

Consider the general case sketched in Figure 24.2 for alpha change and gamma change. The larger vector AB in the figure is associated with a major
change in condition but no change in state. This vector may reflect either alpha or beta change; it is clearly not gamma change. In contrast, the smaller vector CD measuring “condition” represents a minor change in condition but induces a major change in state. Here, a small change induces a gamma change, on the order of the prototypic effect associated with today’s popular approach to analysis called chaos theory. A straw can break a camel’s back under the appropriate condition! A very large change also might be necessary to induce a gamma effect, depending on the initial condition of the system. In this sense, the approach via gamma has a conceptual advantage over chaos theory.

Much common wisdom in OD and experiential education also suggests the value of distinguishing changes in condition-within-a-state from changes in state. Thus, considerable time and effort may be expended in team development before “anything happens,” and then quickly a team will “take off” and “go into orbit.”

On one view, this approximates shaking a bottle of ketchup: you know about that bottle—first none will come and then a lot will. Similar notions inhere in common concepts of developmental phases in laboratory education, as in Hampden-Turner’s (1966) cyclical model of individual change, in all stage models such as that of Bennis and Shepard (1956), or in Lewin’s venerable notions of unfreezing leading to choice or change, which can in turn lead to refreezing (Zand and Sorensen, 1975). To illustrate, Lewin’s model clearly implies several different states, and it is important to distinguish which one exists at any point. Specific behaviors or interventions appropriate in the unfrozen state of consciousness might be impactless or even seriously counterproductive in the frozen state.

The graphics in Figure 24.2 assume too much, of course. Analysis of social phenomena would be simple if we really could define and distinguish state I from
state II. But all we have are vectors with known magnitudes—we often lack a sense of direction. As noted, similar vectors might generate any of the three kinds of change, depending on the initial state of the system. Note that the analysis below of phases of burnout will add powerful new capabilities to this analysis.

**IMPLICATIONS FOR OD AND QWL SUCCESS RATES**

Obviously, this introduction to trinitarian change has important implications for the earliest chapters in *Irony II*. Few of the studies there take kind of change into explicit account. The most that can be said here is that, if anything, failure to distinguish types of change probably serves to deflate estimates of success rates. This could occur by either camouflage or distortion, as it were. Actual alpha change would give a direct estimate of effects; for beta, the impact on success rates is problematic, depending on the range of intervals that are recalibrated at O₂. For actual gamma change, the probability seems that contrary or no change would be interpreted.

**EVIDENCE OF GAMMA EFFECTS IN OD EFFORTS**

Today, we simply know too little about distinguishing the three types of change in any reasonably rigorous and consistent way. Too much is left to the imagination (e.g., Armenakis, 1988). Our purpose here is to take a small step toward what we need to know, and the motivation is clear and direct. OD interventions are centrally involved with seeking change in the concepts of the quality of organizational life that should and can exist, and far less concerned with raising the level of relatively stable parameters or dimensions. Thus, gamma changes—not alpha—are the prime intended effects of such interventions.

This chapter draws attention to the need to distinguish gamma, in effect, by establishing the inadequacy of alpha or beta in accounting for the variation in one specific set of data. The data come from an OD structural intervention—the installation of a Flexi-Time system of work hours (Golembiewski, Hilles, and Kango, 1974). Table 24.1 schematizes the research design. Self-reports on eighteen attitudinal items are reported along Likert-like continua of seven equal-appearing intervals, with the extreme positions being anchored by brief verbal descriptions.

**Factor Analysis—Search for Degrees of Congruence**

There exist many ways to operationalize the three kinds of change (e.g., Armenakis, 1988), but no evidence urges discarding the original approach. It is retained here—conveniently, at least. Overall, the search for gamma effects will employ factor-analytical techniques, whose resulting matrices will be tested for congru-
TABLE 24.1 Research Design of the Installation of a Flexi-Time System of Work Hours Schedule

<table>
<thead>
<tr>
<th></th>
<th>Short Post</th>
<th>Long Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Test</td>
</tr>
<tr>
<td>Day 1</td>
<td>O₁</td>
<td>O₂</td>
</tr>
<tr>
<td>Day 15</td>
<td>X</td>
<td>O₃</td>
</tr>
<tr>
<td>Day 195</td>
<td>O₂</td>
<td>O₃</td>
</tr>
<tr>
<td>Day 375</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experimentals  | O₁          | X          | O₂         | O₃         |
Comparisons     | O₁          | O₂         | O₃         |

O = Observations via self-reports on a standard questionnaire.
X = Flexi-Time installation.

ence. The reliance on factor analysis can be explained nontechnically. Factor analysis essentially seeks to isolate the major dimensions of reality necessary to economically account for the variance in scores on some set of variables.

These properties of factor analysis encourage our reliance on it. That is, any major incongruence between the factorial structures representing each wave of Flexi-Time questionnaire data gathered at several points in time is interpreted as a change in the dimensions of reality necessary to account for the variance in the sets of responses. And, of course, gamma is defined as basic change in those dimensions of psychological space.

Specifically, six matrices are involved: one for each of the three administrations of the standard questionnaire for Experimentals only, and one for each of three administrations including all Experimentals and Comparisons taken together. Responses from Comparisons only are not factor-analyzed, since there were only eighteen, which is not sufficiently greater than the number of questionnaire items to encourage confidence in any resulting factorial solutions. The six factorial structures will be compared both “within” and “between”: within each of the three observations for Experimentals as well as for Experimentals plus Comparisons, and between the three individual observations for Experimentals. Two specific statements may be made.

The essential congruence of factorial structures for Experimentals over the three administrations will imply that gamma effects did not occur.

The essential congruence of factorial structures for Experimentals plus comparisons for each of the three questionnaire administrations will imply that gamma effects did not occur.

As noted, it was not possible to compare directly the factorial structures of Experimentals versus those of Comparisons.

Five major points require emphasis before introducing data. First, factor analysis will be used to help isolate gamma change, defined as the substantial
incongruence of pairs of between-wave factorial structures as determined by Ahmavaara’s (1954) technique. Put otherwise, no major changes in factorial structures are expected when only alpha changes occur. Even large changes from the pre-intervention mean may reflect only changes in the condition within a stable state, as contrasted with changes in state. Changes in state, or gamma, generate major incongruence between pairs of factorial structures. Beta changes are essentially intermediate to alpha changes in this regard.

Second, this chapter reports on only one factor-analytical variant, PFA-2. Any evidence here of gamma effects is almost certainly not an artifact of a specific factorial procedure, however. Three other factorial variants generate essentially the same structures as PFA-2, as is established elsewhere (Golembiewski, Billingsley, and Yeager, 1976b).

Third, the OD intervention studied was clearly impactful. That is, substantial patterns of intended effects were observed in comparisons of pretest, with short post and also long post scores, when the base of comparison was

An item-by-item analysis of a standard questionnaire soliciting eighteen separate responses at three points in time;
An analysis of changes along six dimensions apparently tapped by the eighteen items, as determined by factor analysis of responses to Wave 1 of the standard questionnaire, which structure was then used to score responses at all three points in time;
An analysis of changes along six dimensions apparently tapped by the eighteen items as determined by factor analysis of responses to Wave 2 of the standard questionnaire, which structure was then used to score responses at all three points in time;
An analysis of changes along five dimensions apparently tapped by the eighteen items, as determined by factor analysis of responses to Wave 3 of the standard questionnaire, which structure was then used to score responses at all three points in time.

This may seem to be analytical overkill, but this early stage of analysis needs to be scrupulous in eliminating alternative explanations of results. Better to do too much, at this stage of the game, than too little.

Fourth, estimating the congruence between factorial solutions involves one strategic choice, since the several factorial procedures did not generate an equal number of factors in all cases. PFA-1—using Kaiser’s (1960) criterion of an eigenvalue greater than 1.0 to govern the extraction of factors—generated seven factors for O1, six for O2, and five for O3. In all three cases, the factors with eigenvalues > 1.0 accounted for some 70 percent of the variance.

The convention was adopted to use seven factors, even where this meant disregarding Kaiser’s rule. Several attractions motivated this choice.
Fifth, the procedure for comparing factorial structures is straightforward, if complex (e.g., Ahmavaara, 1954).

### Statistical Procedures

Six tests are run to gain perspective on the notion that variation in the Flexi-Time data set can be explained only by gamma effects. Better said, perhaps, it will be shown that alpha or beta effects are not sufficient to explain that variation.

First, the within-wave congruence of the structures isolated by the four factor-analytic procedures is compared. A substantial congruence will make it difficult to argue that any relationships below are artifacts of any specific factorial procedure.

Second, between-wave congruence is determined, using the total batch of Experimentals plus Comparisons. Estimates of the variance in common between pairs of factorial structures will be used to estimate their congruence. Major reductions of common variance are consistent with the interpretation that gamma changes occurred in Waves 2 and 3 as a result of the experimental intervention.

Third, a similar analysis is conducted for Experimentals only. Crudely, lower between-wave congruence of factorial structures for Experimentals (N = 32) is expected. Such an effect is consistent with the interpretation that gamma occurred. That is, the exclusion of comparisons subjects—who experience only random changes over the observational period from Wave 1 to Wave 3—will remove a portion of the common variance that probably served to somewhat inflate the estimates of congruence of factorial structures for Experimentals plus Comparisons (N = 50).

Fourth, the congruence of factorial structures is tested for a randomly distributed N. This permits an estimate of the stability of factorial structures, as well as of their dependence on the size of N. The substantial congruence of within-wave structures will imply that any patterns isolated in the first three analytical approaches are not simply artifacts of N.

Fifth, a test will be conducted concerning the impact on between-wave comparisons of the present convention of setting at seven the number of factors to be extracted. The congruences of structures with five, six, and seven factors are tested. Why be so finicky? If allowing the number of factors to vary does not much affect the congruence of between-wave factorial structures, the implied conclusion is that gamma accounts for any substantial incongruence isolated by the second and third research emphases above. Note that analysis revealed that within-wave comparisons are affected only slightly by the convention of considering seven factors, but no data on the point are reported in this chapter.

Sixth, an estimate is made of the effects of limiting to five the number of iterations of the structures derived from PFA-2. The congruence of a random sample of structures will be ascertained when iterations are set at five versus
ninety-nine. High congruence of pairs of factorial structures will imply that the major convenience of limiting iterations to five did not stop data manipulation when further substantial convergence of individual structures was possible. Hence, any major incongruence isolated in the second and third emphases will be credibly assignable to systemic effects such as gamma changes.

Congruence of Within-Wave Structures (N = 50 and N = 32)

To summarize briefly, it is easy to establish that the 4 factor-analytical procedures described above essentially isolate the same structures within each of the three questionnaire waves (Golembiewski, Billingsley, and Yeager, 1976a). Conveniently, the summary below of r² implies a very substantial intrawave congruence of structures extracted by the four factorial procedures, given that the square of the product-moment correlation coefficient provides an estimate of the common variance between the structures of any two procedures. In sum:

- Wave 1: product-moment r², = .9524
- Wave 2: product-moment r², = .9850
- Wave 3: product-moment r², = .9929

The average r² permits the estimate that the eighteen within-wave comparisons of pairs of structures generated by the four factorial procedures share over 97 percent of their variance in common, on the average.

This pattern holds for an N of 50, and the pattern for N = 32 is similar. For details, again consult Golembiewski, Billingsley, and Yeager, (1975).

This impressive congruence implies that any results below are almost certainly not artifacts of the single factorial strategy focused on below.

Congruence between Waves, All Subjects

The factorial structures change enough between waves to suggest that alpha or beta changes are not adequate to account for the low congruence between the pre-test factorial structure and the two post-intervention structures. For PFA-2 only, the product-moment correlation coefficients in Table 24.2 imply that about 67 percent of the between-wave variance can be thought of as common. The pattern is similar for the other three factorial procedures, although they account for somewhat less of the variance (around 64 percent).

Table 24.2 suggests a qualified but stout conclusion. The interwave differences might be accounted for in several ways—and not only as signs of the fundamental variation in conceptual space here designated as gamma change. To be sure, no absolute criteria establish the point at which interwave changes become great enough to signal gamma changes. Yet the loss of some 30 percent of variance in comparisons of the congruence of factorial structures for the three
waves of questionnaires has a compelling quality to us and, if nothing else, implies the inadequacy of alpha change as an explanation.

Several reasons reinforce this conclusion. First, it is difficult to credit most of the interwave variations to measurement error. Second, the reduction in variance seems large enough to imply major changes in the dimensions of the relevant psychological space, as contrasted with variations in the level of relatively stable dimensions or with variations in the measurement intervals associated with some stable dimensions of reality. The latter are beta changes, and the former are alpha changes. We presume they would not be associated with in the major incongruencies in analytical structures reflected in Table 24.2. Third, the N = 50 batch of subjects represented in Table 24.2 contains both Experimentals and Comparisons. Other data convincingly demonstrate that Comparisons reflected only random variation in their three waves of self-reports. In a major sense, then, the 67 percent estimate of common variance between factorial structures no doubt understates the impact of the experimental intervention in changing the relevant psychological space as measured by comparisons of factor-analytical structures at three points in time.

**Congruence between Waves, Experimentals Only**

Fortunately, it is possible to test this last surmise. Table 24.2 presents some relevant data using only the Experimental subjects. As expected, the estimate of common variance falls, to an average of 51 percent from 67 percent.

Table 24.3 supports two conclusions. The pattern is consistent with the conclusion that gamma changes did occur in the data batch. Moreover, the reduced estimate of common variance when N is only thirty-two is both expected and large enough to further undercut the credibility of the proposition that alpha or beta effects alone are capable of accounting for the diminished congruence of factorial structures. If a 30 percent loss in interwave common variance is not
TABLE 24.3 Congruence of Between-Wave Factorial Structures for Experimentals Only (N = 32, paired-comparisons of three waves, PFA-2)

<table>
<thead>
<tr>
<th></th>
<th>Interclass correlation</th>
<th>Product-moment correlation</th>
<th>Estimated common variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1 vs. 2</td>
<td>.6121</td>
<td>.6533</td>
<td>.42.7%</td>
</tr>
<tr>
<td>Wave 1 vs. 3</td>
<td>.7581</td>
<td>.7768</td>
<td>.60.3%</td>
</tr>
<tr>
<td>Wave 2 vs. 3</td>
<td>.6711</td>
<td>.7023</td>
<td>.49.3%</td>
</tr>
</tbody>
</table>

sufficient to establish the likelihood of gamma effects, that is to say, a 50 percent loss provides very much more formidable support for that conclusion.

Congruence of Within-Wave Structures for Variable N

It may be the case, of course, that the lower percentage of shared variance in the case of N = 32 versus N = 50 is artifact of the variation in N itself, rather than an effect of excluding the Comparisons. This dependence of within-wave factorial structures on the size of N was tested in a very demanding way.5

The test reveals common and stable factorial structures that are substantially independent even of major changes in N. This finding does double duty. It prohibits gaining a cheap victory over Table 24.2 by doubting or denying the stability of the underlying factorial structures. Moreover, the summary data above clearly imply that a reduction in N by itself cannot account for the pattern in Table 24.3.

Congruence of Between-Wave Structures for Variable Number of Factors

As noted, this analysis conventionally compares the first seven factors in all structures, even in the case of an eigenvalue less than 1.0. This convenient convention does not seem to do violence to the data, as a test case implies. To illustrate for PFA-1 only, for Experimentals plus Comparisons, the congruence of factorial structures was established when the number of factors for Wave 1 was seven, for Wave 2 was six, and for Wave 3 was five. Table 24.4 summarizes the analysis.

The estimated common variances are somewhat higher in Table 24.4 than in Table 24.2, but the average shared variance increases only from 67 to 74 percent when the number of factors is varied to include only those with eigenvalues greater than 1.0. The effect is small, and in the expected direction. Consequently, the convention of setting the factors to be compared at seven had the effect of lowering the variance in common pairs of structures, but not much.
TABLE 24.4 Congruence of Structures with Different Number of Factors, All Cases

<table>
<thead>
<tr>
<th></th>
<th>Interclass correlation</th>
<th>Product-moment correlation</th>
<th>Estimated common variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1 vs. 2</td>
<td>.8326</td>
<td>69.3%</td>
<td>82.2%</td>
</tr>
<tr>
<td>Wave 1 vs. 3</td>
<td>.8139</td>
<td>66.2%</td>
<td>80.0%</td>
</tr>
<tr>
<td>Wave 2 vs. 3</td>
<td>.8612</td>
<td>74.2%</td>
<td>65.4%</td>
</tr>
</tbody>
</table>

The test case suggests that very little of the incongruence between structures reflected in Table 24.2 can be attributed to its convention of setting at seven the number of factors to be compared. Moreover, the incongruence is great enough to suggest that alpha or beta changes alone cannot reasonably account for it.

A similar analysis for Experimentals only (N = 32) leads to a similar conclusion about Table 24.3. In this case, the number of factors with eigenvalues greater than 1.0 for Wave 1 was seven, for Wave 2 was five, and for Wave 3 was six. The results indicate that it is possible to account for only a small part of the incongruence between factorial structures in Table 24.3 as a consequence of setting the number of factors considered at seven. Table 24.4 summarizes the results for Experimentals.

Recall that Table 24.3 indicates that only about 51 percent of the variance, on average, is shared in common among the several paired comparisons of between-wave factorial structures for Experimentals only. Allowing the number of factors to vary (as the data in Table 24.5 show) does not much affect the estimate of common variance, which increases, but only to 52.5 percent. This pattern implies that very little of the between-wave incongruence in factorial structures can reasonably be assigned to the convention of setting the number of factors at seven.

TABLE 24.5 Congruence of Structures with Different Number of Factors, Experimental Cases

<table>
<thead>
<tr>
<th></th>
<th>Interclass correlation</th>
<th>Product-moment correlation</th>
<th>Estimated common variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1 (7 factors) vs. 2 (5 factors)</td>
<td>.7741</td>
<td>.7895</td>
<td>62.3%</td>
</tr>
<tr>
<td>Wave 1 (7 factors) vs. 3 (6 factors)</td>
<td>.6852</td>
<td>.7150</td>
<td>51.1%</td>
</tr>
<tr>
<td>Wave 2 (5 factors) vs. 3 (6 factors)</td>
<td>.6223</td>
<td>.6631</td>
<td>44.0%</td>
</tr>
</tbody>
</table>
in most analyses in this chapter. Moreover, the pattern also supports the interpretation that the between-wave incongruence in factorial structures is large enough to suggest gamma effects.

**Congruence of Within-Wave Structures, Five Iterations vs. Ninety-Nine Iterations**

As a final test, it seems obvious that the SPSS convention to limit iterations to five had a small impact on this analysis. Specifically, ten sample tests were run that involved all three waves, an N of 50 and 32, and five, six, or seven factors, as appropriate. The basic comparison involves determining the congruence of pairs of factorial structures when iterations are cut off at five and when they are allowed to run to a maximum of ninety-nine. The ten resulting estimates of common variance cover a narrow and high range, from 96.96 to 100 percent, with a grand mean of 98.80 percent. This very high degree of congruence implies that the convenience of limiting iterations to five is not analytically troublesome.

**GREATER SUCCESS BY SPECIFYING KINDS OF CHANGE**

One cannot predict without qualification that OD efforts will be rated as more successful when they routinely begin specifying kinds of change. But that is possible, as was the case cited early in this chapter with an apparent failure that motivated the test for a plural concept of change.

What can we be relatively certain about? Two classes of conclusions will do, for starters. The first comes from good old-fashioned hindsight, for more than two decades have passed since the original argument above was published. The second class of conclusions dates from the original publication, and most of its components still seem quite serviceable.

**20/20 Hindsight**

The perspective of two decades suggests the usefulness of the effort to test a plural concept of change (Golembiewski, 1986, 1997). Basically, alpha, beta, and gamma changes have become part of the vocabulary of research involved with the study of human change (e.g., Armenakis, 1988; Tennis, 1989). This is surprising, even astounding. For plural change as a concept is a radical one, and it is corrosive of established beliefs and practices in both research and application. Agreement does not yet exist about the ways to measure plural change, to be sure, but energetic effort has been applied to the topic and may yet pay off (e.g., Armenakis, 1988).

The basic distinction between the types of change also has penetrated many of the conventional OD sources of information. Thus, several national conferences of learned societies have featured panels on the change typology. In addi-
Trinitarian Change

The basic conceptual distinctions also have become commonplace in the general OD literature—in various textbooks (e.g., Burke, 1982), in survey articles intended for professional audiences (e.g., Armenakis, 1988), and in other sources.

Why this clear and growing acceptance of a conceptual distinction that is at once a difficult one and also challenges vested interests? The short form of the answer is clear and direct: greater success in OD requires some such distinction, and that point rings true to both scholars and practitioners, even at a face-valid level.

The longer form of a useful working answer has four emphases, which in turn only illustrate a larger possible catalog. First, OD cannot avoid making such a distinction, being involved as it is with choice and change. The plural concept makes general sense, which accounts for its all-but-universal acceptance. Indeed, I know of not even a single criticism of the basic concept (Golembiewski, 1989), which stands in sharp contrast to the often zesty debate concerning ways to measure the various kinds of change (e.g., Armenakis, 1988).

Particularly impactful on OD is the basic notion that the results of studies of change are confidently interpretable only under a strict condition: that the existence of non-alpha change is rejected. Alpha-only change permits conventional interpretation of results, but non-alpha change poses serious and as-yet-unresolved issues. In concept, this sharply reduces the credibility of interpretations of the results of OD applications. The significance of the notion is widely appreciated, as by a large-scale seller of survey studies. In initial response to the concept of gamma change, the seller commissioned a think tank to do a study, which was critical of the method of measurement but nonetheless seems to have basically accepted the tripartite concept of change. Some employees of the surveying institution later advised clients to test for non-alpha effects.

Second, no counter-demonstration provides a credible basis for rejecting the plural concept. Rather, the tripartite notion has diffused broadly: into educational and psychological measurement, counseling, management, and the study of opinions among elites (e.g., Chittick, Billingsley, and Travis, 1988), among numerous other disciplines, fields, and subfields. This diffusion does not settle matters, but it counts for a great deal.

Third, the basic conceptual distinctions have an intuitive appeal. This is particularly the case in OD, which has often proclaimed its mission to create what is comfortably seen as gamma change—as in the development of new social orders at work, or in generating regenerative interaction to replace the degenerative variety. The basic point gets its sharpest expression in the current emphasis on “transformation” (e.g., Adams, 1984), in contrast with incremental development which virtually shouts “alpha change.”

Fourth, gamma suits many of the metaphors common in OD. Thus, a facilitator may exult, “Our group went into a new orbit.” This refers to achieving a different mode of functioning, of course, and suggests gamma.
While the data do lead to conclusions that are intriguing and challenging, much remains undone. Of particular potential value is the use of the phase model of burnout—see especially Chapters 19 and 20—to help estimate gamma change. The argument has been developed fully elsewhere (Golembiewski et al. 1996), so only a sketch will be presented here.

**USEFUL FEATURES OF PHASE MODEL APPROACH**

As an initial inventory, at least six features of the phase model approach have major attractions for research and applications. The first attraction deals with research of special relevance to the phase model of burnout, and the five others have a range of uses.

**Cross-Cultural Relevance of Phases**

As more detailed analysis shows (Golembiewski et al., 1996), perhaps the greatest theoretical and practical attractions of the phase model relate to cross-cultural features. Phase results are quite similar in cross-national applications (e.g., see Chapter 20), and these results suggest the approach’s broad usefulness in several regards, concerning: the need for change; evaluating the degree of change after intervention; and specifying the direction and kind of change that occurred. The phase model assumes at least one case of gamma change—from no/little burnout to a notable amount. Preliminary analysis indicates gamma difference between several or all of the phases and the factorial structures they generate on the Maslach Burnout Inventory items.

In addition, the phase model directly includes a sense of the direction of change, whereas the argument above focuses on magnitude. The latter feature limits the usefulness of the early forms of alpha, beta, and gamma change.

Obviously, such features could powerfully augment both the reach-and-grasp of OD applications and their success rates.

**Five Illustrative Attractions**

In addition, other attractions of trinitarian change in association with the phase model seem clear enough. We encourage moderation in interpretation, but the data do seem to indicate that something like gamma does occur in the populations examined. Moreover, substantial reasons discourage presuming that patterns in the data are determined by other features—by random causes, a single factorial procedure only, dependence on the level of N, the basic convention of considering seven factors of each wave, or the convenience of limiting to five the number of iterations seeking to maximize congruence of each structure.

The results above may reflect the impact of factors not considered here, of course. For example, several other ways could have been used to estimate congruence, even though preliminary work with several requires no modification.
of the present conclusions. Here, note only that specialists of goodwill differ profoundly as to what is the most appropriate measure of congruence (Armenakis, 1988). However, despite interpretive problems, a major test of an alternative way of determining congruence provides no clear support for the hypothesis that the present results are artifacts of the specific measure of congruence used (Golembiewski et al., 1976b). Also, the ratio of subjects to items in the present data batch is at best 2.6 : 1. Normally, this raises a question about the results above, but the patent stability of the factorial structures when N is allowed to vary suggests that the problem is not significant.

PROMISING LINES OF RESEARCH

Reaching this point has been trying enough—especially for the reader impatient with statistical complexity—but extensions of this promising line of research will challenge researchers. Thus, it will be exceedingly difficult to establish the existence of gamma, as well as to satisfactorily differentiate gamma from alpha and beta. But the present results suggest strongly that the effort is necessary, since alpha or beta cannot credibly account for the major incongruencies in structures reported here.

To the degree that the preceding analysis is close to reality, it has profound implications for experimental research designs in the social sciences, as well as OD. Here we recommend other tests of this analysis, as well as an interim exercise of prudence that can ameliorate our present lack of knowledge about kinds of change. Six themes are especially noteworthy in these regards.

First, this analysis implies that prodigious energies in the behavioral sciences have been applied to the wrong methodological issues. Consider the sophisticated but inconclusive effort directed at the question of how change is to be measured (Cronbach and Furley, 1970; Van Meter, 1974; and many others). In contrast, this analysis suggests strongly that the first question should be: What kind of change is being measured? Few studies (e.g., Buss, 1974) deal with this prior question, however. Most studies seem to assume that only alpha is relevant. That assumption seems clearly inappropriate for successful OD interventions, and the same may be true of many natural-state contrived experiments.

“Confirmatory factor analysis” takes us some distance down this long trail. But it is less flexible than the present approach, and also like the present strategy for measuring change tests for incongruence in the dimensions of the perceived psychological domains is by comparing the results of factor analyses before and after an OD intervention. This is relatively simple and avoids many of the formidable problems in calculating meaningful change scores.

Going one step further will present major complexities, to be sure. It is one thing to estimate statistically the incongruence between two structures, and quite another to label and compare the individual factors in those structures, and to specify their direction in relation to specific OD goals. The first task is mecha-
The second task is a major art form in behavioral science. Our study took the easier route. It does not deal with the specific changes in individual factors induced by the Flexi-Time intervention. It was a matter of doing the simplest things first, as it were.

Third, this analysis implies a hard message for much OD research. Specifically, interpreting any results of existing research is chancy in the absence of knowledge about types of change, which is seldom available. Even research designs that surpass the usual norms for rigor and care are suspect in this regard.

Fourth, this focus on types of change implies the strategic value of time-series designs, as well as of such variants as time-lagged designs (Campbell, 1963). These two kinds of designs may be sketched as follows.

**Time-series design:** \(O_1 - X - O_2 - O_3 \ldots\)

**Time-lagged design:** \(O_1 - X - O_2 - O_3 \ldots O_1 - O_2 - X - O_3 \ldots\)

As before, “O” is observation and “X” is experimental intervention. Of course, the explanatory power of both designs is enhanced if controls or comparisons are provided. Beta effects and gamma effects seem far more difficult to isolate in simple \(O_1 = X = O_2\) designs.

The awkwardly named “catastrophe theory” or “chaos theory” also may be useful. It deals with discontinuous change of various sorts, the kind here called gamma change. Catastrophe theory contrasts with most existing mathematics—which, as in the calculus, deals with continuous change—and has received wide coverage even in the mass media (Newsweek, 1976).

Other alternatives might be devoted to subjects as they respond to measuring instruments requiring self-reports. The perception by respondents that they are using “rubber yardsticks” might be variously parlayed into a kind of early-warning system that alerts analysts to beta or gamma changes in experimental design.

Sixth, scaling techniques less sensitive to metric-level assumptions could also be profitably used to seek underlying structures in this type of analysis. Such approaches include smallest-space analysis (SSA) and nonmetric multidimensional scaling.

Clinical attention also seems appropriate, in at least two senses. Thus, specific character type may be associated with specific (and different) phases of burnout. In addition, an individual’s changes between phases may be related to ongoing life experiences.

**NOTES**

1. The terminology relates to Rensis Likert’s (1967) Profile of Organizational Characteristics. The profile seeks two kinds of self-reports: Now responses, which solicit data about how respondents actually see their
organization unit; and Ideal responses, which seek information about how respondents feel their organization should be. The focus throughout this chapter is on purported descriptions of existential states only—that is, on Now responses.

2. To test the possibility that any results are artifacts of a single factor-analytical technology, in addition, this analysis was replicated for four variants: Principal Factoring without Iterations, or PFA-1; Principal Factoring with Iterations, or PFA-2; Alpha Factor Analysis; and RAO, or Canonical Factor Analysis. For a detailed description of the differences between these techniques, see McDonald (1970). All four techniques are conveniently available in the Statistical Package for Social Scientists (SPSS).

Results will be reported here basically for PFA-2. For evidence of the very substantial congruence of the structures generated by PFA-2 and the three other factorial variants, see Golembiewski, Billingsley, and Yeager (1976).

3. To briefly describe PFA-2, or Principal Factoring with Iterations, the main diagonal of the correlation matrix is replaced with the communality estimates, or squared multiple correlations between the variable and the rest of the variables. The estimates of communality are iteratively improved by factoring again with the calculated communality estimates derived from the preceding solution. This process is continued until the estimates of communality converge, or until the differences between successive estimates are negligible.

The SPSS program is set for five iterations or less, because the CDC 6400 implementation of SPSS used here capitalizes on the 60-bit accuracy of the 6400’s word size. A note on the printout alerts analysts if five iterations do not suffice for absolute convergence.

As a major convenience, most analysis in the text deals with structures for variants which have been iterated five times or less.

4. This convention had several motivators. Primarily, the convention facilitated the convenient replication of the results reported here. In addition, it reduced the potential for error when comparing matrices of larger or smaller size, respecting the advice found in Kaiser, Hunka, and Bianchini (1971, esp. pp. 411–12 and 421). Moreover, the choice of seven factors permits the inclusion in Waves 2 and 3 of several items that load heavily on factors 6 and 7 only. Finally, had the decision been to consider only the first five factors generated by each procedure, substantial portions of variance would have been lost in analyses of Waves 1 and 2. Data concerning the effects of the decision to focus on seven factors will be presented below. For now, note only that the convention was not analytically troublesome.

5. Subjects were randomly eliminated so that five independent subpopula-
tions of thirty-five were isolated for each of the three waves. PFA-1 was then applied to these fifteen subpopulations as well as to the total population \( N = 50 \), and the resulting factorial structures were compared for within-wave congruence.

The results of this severe test strongly imply that variation in \( N \) itself does not determine the pattern reflected by Tables 24.2 and 24.3. To illustrate with one of the fifteen comparisons:

<table>
<thead>
<tr>
<th></th>
<th>Interclass correlation</th>
<th>Product-moment correlation</th>
<th>Estimated common variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1, ( N = 35 ), vs. Wave 1, ( N = 50 )</td>
<td>.9427</td>
<td>.9437</td>
<td>89.1%</td>
</tr>
<tr>
<td>Wave 2, ( N = 35 ), vs. Wave 2, ( N = 50 )</td>
<td>.9415</td>
<td>.9427</td>
<td>88.9%</td>
</tr>
<tr>
<td>Wave 3, ( N = 35 ), vs. Wave 3, ( N = 50 )</td>
<td>.9734</td>
<td>.9737</td>
<td>94.8%</td>
</tr>
</tbody>
</table>

In sum, over 90 percent of the estimated variance in this illustrative case can be considered common in the underlying structures. The average for all subpopulations tested was about 88 percent, which is impressive congruence, especially given the severity of the test.

REFERENCES


Perhaps the most useful yet simple metaphor for a healthy area of inquiry and application is that of a vital center and an expanding periphery. Why? In the absence of an expanding periphery, to illustrate, the danger is that the center’s validity cannot be validated by testing; nor is a growing and enthusiastic cohort of workers-in-the-vineyard probable in the absence of an expanding periphery. Hence, lacking an energetic periphery, even a valid core is likely to suffer, and perhaps stagnate. At an extreme, indeed, the core may in time collapse on itself. Or, in another form, if real testing of the core by extending the periphery does not occur, the core and its peripheries may drift apart, as it were, thereby impoverishing both.

The experience in OD—or ODC, if you will—has been ironic. Thus, great activity has existed in the form of what were initially extensions of the central core, but they inadequately served to test the validity of the core. Either the extensions were inadequately described in conceptual terms that could be tested for congruence with the central core, or the extensions received modest empirical testing and hence could make only a largely ideological claim to continuing attention. In general, strategic choices were neglected, despite high success rates. See Chapters 1 through 3. Neglect of cumulative testing tended to dominate, and forays into the peripheries tended to be more alluring than the less-romantic and detailed testing of the central core. For whatever reasons, the proposed peripheries tended to float away from the central core rather than contributing to its vitality via accepting or rejecting aspects of the core.
The best of intentions often were involved. For example, consider the earlier efforts to expand OD’s usefulness by enlarging the range of values admitted by the core (e.g., Beer, Eisenstat, and Spector, 1990). The basic problem? Paramountly, some normative extensions might result in an area of application that was “no longer OD.” Significantly, also, the implicit assumption often was that care about extending the core was not necessary because prevailing success rates were modest. Hence, the difficulties encountered by the success rate studies of the 1980s in getting a full and fair hearing. The claims of “easy creativity” dominated for a decade or more, without impactful objection; and it was only with the later 1990s that OD consensus took reasonable account of the growing success rate literature—which had expanded in major ways, but whose bottom line remained much the same as in the evaluative studies of the 1970s and earlier. See again, the first three chapters of this volume.

The focus in this chapter is on one such “periphery” that at once can add to the reach-and-grasp of OD but also might serve to undercut OD—appreciative inquiry, or AI. It is singular in that no other extension of OD has achieved an even remotely robust momentum. Hence, AI has the potential for really testing the validity of OD’s central core, with a range of possible consequences that are as yet undetermined. Thus, AI may best serve as disciplined by the normative contexts discussed in Chapters 7 through 9, which tie AI to OD development while enriching both. Alternatively, properly developed, AI may be shown to be a basic successor to the OD central core, a core fundamentally different from its predecessor. In such a case, both AI and OD may prove to be different realms with separate claims to status as models to guide application under different conditions.

What is AI, to be more specific? Two exhibits below sketch the conceptual boundaries of the basic AI model to transcending the alleged limitations of OD and QWL approaches. The basic difference is both direct and even compelling, at first glance to observers like this author and after intense thought and experience by AI aficionados. Basically, OD is seen as “problem-oriented,” and hence as implying too much about limitations in nature and deficiencies in people. Basically, also, AI springs from a basic growth orientation in people, which typical OD allegedly impedes if not perverts into defensiveness often or always.

Exhibit 25.1 summarizes the basic AI model in terms of a few propositions. Fuller details are available, both from AI adherents (e.g., Cooperrider and Srivastva, 1987) as well as from those like this author who classifies himself as a constructive and intendedly friendly critic.

Exhibit 25.2 provides a necessary sharpening of the usual AI model, as it is circumscribed in Exhibit 25.1. In sum, AI ideation rests on what are here called “core assumptions.” For the basic AI model to apply, certain important things must either exist or can be made to exist. Without such “core assumptions,” the basic AI model stands in danger of being a set of attractive ideals that inade-
EXHIBIT 25.1  A Primer on Appreciative Inquiry

AI is a form of social constructionism, a belief system proposing that people create all/most of their realities by their own enactments, especially by confrontation or dialog, whether conscious or unconscious.

AI ideology sees almost all people as capable of consciously enacting or dialoging alternative and more desirable realities.

In AI, participants are oriented toward enactment or dialog by targeting “positive” aspects of their life experiences relevant to some goal—e.g., building a congenial team using past peak experiences in teams as a template to guide that constructionism.

In AI, major consensus(es) will develop around these positive themes, and the “work to be done” will emerge, possibly vetted by knowledge but more likely due to the now-unfettered insights of the participants.

AI promises an aura of optimistic discovery, of movement toward creating a valued common ground.

In AI, the anti-goal is clear: do not introduce “negative” materials, let alone dwell on them. Why? The rationale for avoiding such negative materials is expressed in a cluster of attributions-to-be-avoided: that there is “something wrong” with existing actors or institutions; that this “something wrong” must be remedied or “fixed”; and that problem-solving is frame-repair, at best, rather than frame-transcending.

In AI, the goal seeks to be transformative rather than incremental.

EXHIBIT 25.2  Some Core Assumptions of Appreciative Inquiry

AI taps the “life-giving forces,” which can be elicited only or best by “positive stories.”

AI, “as a process of valuing, consists of a rigorous ability to dissociate all seeming imperfections from that which has fundamental value . . .” (Rainey, 1996, p. 35), and this unique ability becomes the source of its power to gain transforming insight. By implication, that ability is very widely distributed, or perhaps is universal, in normal human beings.

AI inherently is benign, even magical: it encourages “wonder,” as distinguished from knowledge; it releases the “child” as the agent of inquiry, the “everlasting beginner” rather than the skilled diagnostician; and it otherwise motivates the “spirit of inquiry.”

Consistently, AI advocates relish the possibly special usefulness of children in asking positive questions and also value inter-generational dialog (Cooperrider and Associates, 1996, pp. 7–8).

Consequently, AI places no special value on diagnosis, or on role-specialized diagnosticians. Indeed, on balance, AIers see diagnosis as a limit on its processes. On occasion, signal, AI advocates propose that the only worthy role-specialization is that of developing the “unconditional positive question” (e.g., Cooperrider, 1996, p. 8).

AI is uniquely suited to “accelerate anticipatory learning involving larger and larger levels of collectivity . . . , widening the circles of dialogue to groups of 100s, 1000s, and perhaps more—with cyberspace relationships into the millions.” (Cooperrider, 1996, pp. 5–6).
quately provide the reach-and-grasp necessary for effective planned change. At
the very least, AI proposes features that either reject OD or are at substantial
odds with it. For example, much OD places substantial faith in the capacities of
ordinary members of organizations, just as AI does. However, OD also often
places substantial importance in the “OD intervenor” and, for good or ill, that
role virtually disappears under the corrosive impact of several of the assumptions
in Exhibit 25.2.

TWO CRITICAL PERSPECTIVES ON AI

The available evidence does not permit an easy choice of AI over OD; indeed,
that evidence encourages doubt whether any such demonstration will ever be
available, or even attempted in any determined way. That is, this section approxi-
mates the summary of a paradox. Far more extensive analysis is available to
support this section (e.g., Golembiewski, 1998).

Let the paradoxical challenge be presented simply. AI has been the most
prominent new approach to planned change for over a decade, but perhaps the
most telling things that may be said about AI is that it has received little of what
may be called “really appreciative inquiry.” Hammond (1995, 1996), a friendly
observer, conveys the total sense of this in his whimsically but accurately titled
The Thin Book of Appreciative Inquiry. Of course, this shortfall would be reme-
died by such add-on features to a clearly attractive AI core:

Subjecting AI’s often-attractive notions to situational analysis—that is,
which conditions contribute to success, or failure, in what proportions,
for which AI variants?

Comprehensive and comparative testing, both internally and relationally,
to estimate the efficacy of versions of the target treatment—to estimate
the “within” efficacy of each version of the basic AI treatment, as well
as to estimate the “between” efficacy of a version of the target treatment
as contrasted with alternative treatments.

But this shortfall remains and, indeed, seems beyond AI. A simple contrast
frames the basic shortfall. OD is today at a central stage in its development, and
will either build on past progress or fade away. As Church (2001, pp. 18–29)
describes the early steps in the professionalization of OD, and then prescribes
that it “. . . is time to professionalize and parameterize OD . . .” or have it fritter
away. This presumes higher standards for AI than for OD as it was at early
developmental stages. Indeed, AI seems more in the class of “forever promising.”
As one evidence (Fitzgerald, Murrell, and Newman, 2002), friendly observers
still see AI as “the new frontier” in its second decade of prominence.

Why does AI seem a case of perpetual promise? Why, indeed, may no
other outcome seem likely? A complete answer will always be elusive, although
this author (Golembiewski, 1998) elsewhere presents a lengthy effort in that direction. Nonetheless, a useful precis of an answer can be made here. Specifically, three related perspectives on AI receive attention:

AI as a conceptual and operational island
AI as discouraging inquiry
AI as neglect of values

**AI AS CONCEPTUAL AND OPERATIONAL ISLAND**

Perhaps basically, AI aficionados portray it as a kind of island, both conceptually as well as in its broad theoretical framework. That is to say, many supporters see it as *sui generis*. In Kuhnian terms, AI is portrayed as the ultimate paradigm in two senses: as post-paradigmatic successor to all earlier approaches to change like OD, and as the neo-paradigm to replace all deficiency approaches.

If at the risk of being too broad, but only at a small risk (Golembiewski, 1998), most proponents differentiate AI from most or all other approaches to planned change. For example, AI supplants approaches that feature problem-solving, and to AIers that is *the* fundamental advantage. That is, we are told, problem-solving is inherently deficiency-oriented, encourages the locating and placing of blame, and (in the consequent bargain) tends to induce resistance as well as to create additional problems while always failing fundamentally. OD is always problem-centered, in this view, and AI never is. Let us simplify in the service of getting a working if incomplete sense of AI. At its heart (e.g., Cooperrider and Srivastva, 1987), AI proposes to get organization members talking about—appreciating in dialog, as it were—“working” features in responses to “positive questions,” such as these put to a population of school principals (Cooperrider, 1997, p. 9):

As you reflect on your career, can you tell me the story of a high point, a time you felt most alive, most impactful, most successful in terms of the contribution to this school and community?

Consciously, AI not only eschews a problem-solving orientation, which is regarded as including defensiveness, but AI also seeks to build a real-time organization out of the “life stuff” implied by answers to questions about high points and preferred dynamics. The ambitions are clear enough, and available ways-and-means indicate how AI proposes to move beyond such essentials while building on them (e.g., Cooperrider, 1996). See also Exhibits 25.1 and 25.2.

For good or ill, in addition, this basic AI position is seldom qualified. Far more rather than less, consequently, the basic AI position is apt to be exaggerated and that may be costly. Let three points illustrate the broad argument—indeed, the usually unbounded argument.
AI Overstates Real but Qualified Differences

Certainly, AI stands by itself in its unqualified emphasis on the “positive” in change. Just as certainly, however, AI ideation usually overstates the degree of the differences between it and other approaches to change. Note the word “usually.” Especially in recent years, some AIers seek linkages with OD, as in the effort to insert elements of AI into a “traditional problem solving” design for change (e.g., Newman and Fitzgerald, 2001). But such efforts succeed mostly in demonstrating how little is known about such linkages of OD and modified AI. Modest optimism seems appropriate that this situation will change soon, however, as the next major subsection below suggests.

Let a few illustrations do the job of highlighting the overstatement in the typical AI rejection of all problem-solving approaches as negative and deficiency-oriented. Thus, every survey/feedback design with which I have been associated was “problem-solving” but also was built around some version of a balance: for example, what are the positive things in this organization?; and what things need remedial attention, and in what order? Relatedly, the basic contracting mode in many OD designs implies a similar balancing:

- What things can be continued, pretty much as is?
- What things must be started in this organization?
- What things should be stopped?

This diagnosis then becomes the basis for bargains and exchanges. The prototypic question is: What degree of beginning Act A by some people will be required to arrange for stopping Act B by others? And so on and on.

In the vast majority of planned change designs, “balance” similarly dominates, which raises questions about AI’s basic claim. It is not clear what AI’s usual insistence on singularity implies for both theory and practice, but those issues definitely will not be solved by setting up conceptual straw men to be gravely demolished. Moreover, the basic AI charge is bold in two senses. AI in effect finesses the substantial success rates of conventional balanced designs, now apparently well established by the early work in OD, as well as by the 20-plus independent surveys of evaluative studies of OD and QWL efficacy that have accumulated since then. On this critical subliterature on success, see especially Chapters 1 through 3.

AI offers no similar appreciative evaluations of its own efficacy. Nor does its development promise great hope of such an emphasis appearing in the absence of fundamental changes in how AIers see AI. A detailed proof of both positions appears elsewhere (e.g., Golembiewski, 1998). A later subsection, “AI As Discouraging Inquiry,” sketches the case for pessimism while implying how AI could fundamentally learn from OD experience. The immediate focus is narrower.
Lack of Linkages to Other Theory and Practice

Relatedly, and at great cost, AI seldom is linked with other theory and practice like OD relevant to planned change, and AIers experience no great angst about this central lack of linkage. AI usually remains comfortably *sui generis*, that is to say.

Some may consider this only your author’s problem, but let me suggest why the concern is more than self-referential. Consider the well-established treatment progression in third-party consultation or peacemaking (e.g., Walton, 1969). Walton advises an initial emphasis on differentiation—to explore and provide context for the issues-in-contention. Then comes integration, building on but beyond the products of the first phase to manage conflict, or even resolve it. Neglect of either emphasis, Walton proposes, implies a high probability of a failure to manage or resolve conflict. In this regard, Walton stands for virtually all those working with conflict.

Now AI’s resolution-by-definition clearly enough rejects Walton’s view, and in several senses. To illustrate, AI proposes that Walton’s basic sequencing is deficiency-oriented, and hence not ideal or perhaps even effective. But this easy rationale surely needs elaboration, even if correct. Are Walton and those of like mind simply wrong? And what to make of the substantial practice and research that Walton and others see as useful? Or is AI better for some type-cases, and Walton’s approach always suspect or perhaps useful in specified other type-cases?

Even more troublesome are AI’s unexplored interfaces with Jerry Harvey’s (1988) “Going to Abilene,” a favored (if oversimple) concept for description and prescription in OD. How would AI deal with associated phenomena? I do not know; and AI does not appear at all concerned about the point, even though the tsunami of attention to Abilene over the years encourages attention, even requires it.

My best guess? AI could only worsen such a condition, and for two reasons. Even raising the possibility of Abilene seems awkward in AI since it is a “problem” or “problems,” and hence is out-of-bounds for AI. Moreover, successfully diagnosing Abilene would presumably be a “negative.” Hence, AI probably would not help isolate Abilene dynamics. Indeed, AI might make matters worse by reinforcing by neglect the fear of loss of an attractive membership on which Harvey’s type of Abilene rests. The significance of such lacks of linkage do not inspire confidence, even though no one can predict the incidence of Abilene. My own practice suggests Abilene is much more frequent in government than business, but no reliable estimates of incidence exist.

Further, Abilene is only one version of several other related types of “crises of agreement.” To variable but noteworthy degrees, each conceptual subtype
marches to its own drummer (e.g., Golembiewski, 1999, pp. 133–143). So the lack of specific linkage is not trivial. Far more expansively, similar shortfalls in AI seem inherent, or at least deep in the vitals of current AI formulations. Commonly, that is to say, AI discourages analysis. Indeed, on apparent principle, some prominent AIers assign a very low priority to diagnosis and analysis, and some even seem to disparage them (e.g., Cooperrider, 1997).

**AI AS DISCOURAGING INQUIRY**

This last emphasis will profit from greater attention, if only because the claim is so outrageous on its face. The example not only broadly characterizes AI, but no easy way beyond that neglect of diagnosis and analysis suggests itself. It is curious to even write the title above, but AI takes a curious approach to “inquiry.” Indeed, perhaps better said—and even at the risk of a charge of preciosity—AIers have a basic predisposition to fail to appreciate inquiry. Two selected points illustrate this basic contributor to the paradoxes that this section highlights.

**Sparse Research Literature**

Revealingly, AI has not inspired much of a research literature in its nearly two decades of prominence. Applications there have been, and beaucoup; but the existing empirical literature numbers only a few contributions, and those are far-between. In sum, a comprehensive review of AI (Golembiewski, 1998) required two conclusions: the research cupboard was substantially bare and the results of that slim literature are mixed, despite the boosterish tune of AI commentary or anecdotes, which proliferate. Curiously, the sparse AI research is at times very good research (e.g., Jones, 1999). But let us move on.

This position, in my mind, requires few qualifications, resting as it does on both my interpretations as well as those of AIers (e.g., Newman and Fitzgerald, 2001, pp. 37–38). My views on a comprehensive survey of published and unpublished AI sources (Golembiewski, 1998) has been reinforced by a recent updating of published sources.

Whether my own conclusions or those of AIers, the generalization forces itself on the observer. The search of AI sources yields so little research that the key issue drifts away from research. Instead, the question became: what factors contribute to the rarity of research about an approach that is so positively recommended, by so many, with so few qualifications, for so long?

I can only speculate about a working answer, but perhaps usefully. Paramountly, the term “appreciative inquiry” may be curiously disarming for people like myself who are intrigued by AI but remain external to it. The external “constructive critic” may be cautious about unfairly applying to AI conventional criteria of reliability and validity. I found myself being unusually tentative, not wish-
Extending OD Technology/Values

ing to be what AI proscribes—that is, deficiency-oriented, judgmental, unresponsive to the highest aspirations in the human species, and critical of the inevitable bobbles along that way. I finally convinced myself that really caring about AI required subjecting its literature to just such conventional scrutiny.

Compared to those inside the AI tent, I may be less qualified to provide critical analysis, but it seems unlikely to come from within. Thus, AI ideologues often have announced a commitment to “social constructionism,” in most forms of which my sense of “rigorous research” is seen as the imposition of a one-time-elite not aware enough to recognize its self-servingness, and too self-referential to realize that its time of hegemony has not only passed but also that its conventions have become seriously limiting.

In the time since I completed my 1998 survey, to be as inclusive as possible, I have become aware of AI research in the pipeline or in recent publications (e.g., Newman and Fitzgerald, 2001). Perhaps the best work of this kind is from dissertations that hold much promise for the future. Some of the designs coming out of the OD doctoral program at Benedictine University (e.g., Jones, 1999) show great promise but also much room for improvement. In the Jones work (1999), to illustrate, it remains unclear to me as to how AI concepts are operation-alized in practice. Indeed, I have been privileged to see some of this work in various stages of development (e.g., Robinson-Easely, 1999; Jones, 1999), and my opinion stands.

Lack of a “Critical Imperative”

Relatedly, the AI approach does not provide a clear, let alone honored, role for what I would consider constructive criticism of AI applications. Indeed, only a small handful of AI citations fit this category, with Head (1997), Bushe and Coetzter (1995), and Golembiewski (1998, 1999) being representative as well as conveniently available. This aversion seems to characterize both people outside as well as inside the AI tent. Outsiders like myself may be inhibited about raining on someone else’s parade, and the very positive character of general AI commentary also may inhibit the blight of the critical impulse, which obviously does not fit AI logic or values.

Some expressions of a null set for “constructive criticism” in the AI literature are disarming, if not startling. Thus, one major AI ideologue has proposed, as noted above, that prior diagnosis is unnecessary and perhaps even counterproductive (Cooperrider, 1977). At other points, the same AI architect proposes that AEErs are not responsible for “downstream effects.” This nifty doubleplay obviously leaves little room for fair chase by constructive critics.

Perhaps best said, in any case, my personal view of knowledge development features support and opposition—not on the model of the barbarians at the gates, slashing at defenders, but certainly extending to a robust acerbity under
provision. Positively put, I have in mind the institutional efforts of Lee Bradford in the early days of “the laboratory approach” to seek involvement of a broadening range of interests and viewpoints, based on the possibility of learning from them about how to improve or about the need for fundamental revision of the core mission. Of course, Lee was mindful that converts also could result from the process.

Also in OD, major attention and honor went to those—Jerry Harvey and Marv Weisbord come to mind first—who functioned in the role of constructive critic, and typically in the mainstream media like the *OD Practitioner*. As an important aspect of both contributors, each featured a prominent wit—often self-deprecatory, and at times cutting. I envision no comfortable reception for such a role-play in AI, nor does any AIer fit that bill.

**TWO FUNDAMENTAL CONCERNS ABOUT NORMATIVE NEGLECT**

Finally for present purposes, and perhaps basically, this writer has two concerns related to normative issues. Two questions frame these concerns: Why AI? and when AI?

**Why AI?**

In a few words, what values define the appropriate zone for AI applications, and especially as they are potent and “work?” To illustrate with a question based on empirical shortfall introduced earlier, are AI appliers responsible for “downstream consequences” when, by basic design, “upstream” applications do not direct attention to “negative” or problematic concerns? Consider only two type-conditions that raise associated normative implications:

A board of directors is to exercise its legal and professional obligation to review the activities of an executive group.

Many orphans are available for adoption after experiencing extended periods of being warehoused under need-depriving conditions, with some early resistance existing among local health professionals as well as adopters.

Is AI appropriate under the two conditions, given that the focus on “appreciation” might create a euphoria that does not recognize, let alone encourage, a balance of positive and negative features? For example, some probability exists that a focus on both “positive” and “negative” features could be useful or even necessary in both of the two situations above. That is:

The executive group might be involved in well-camouflaged conflicts of interest.
Some substantial proportion of the orphans, because of their harsh childhood treatments, might be expected to experience major developmental post-adoption problems.

Prospective parents as well as authorizing professionals might find this probability of some relevance. Now, as I understand it, AI would neither encourage nor test for such a positive/negative balance. Basic AI questions for inquiry, that is to say, might well have a biasing effect on the associated decision-making, both “upstream” before the basic choice-points as well as “downstream” after those choice-points have been left behind. The key questions to build on and toward as “life sources” might well be:

What are the best conditions that describe the association between a board and its executives?

How would an adopted child add to the full experience of a married couple with no children, or without enough children as the couple sees it? And in what ways would adoptions ease the lives of the children and also of local healthcare and political authorities?

One can easily develop a case for the power of such biasing questions. To speculate about the first case, the board might be subtly discouraged from raising serious “upstream” questions like this one: given the clearly attractive ideal state, what confidence does the track record of the executive group imply about acting on the ideal?

In the present view, AI is 0-for-2 in the first case. Any designs should take both positive/negative features into account. Moreover, intervenors are responsible for downstream effects, and especially if AI ideology in fact biased the decision-making.

In the second case, we can speak of possible dysjoints with greater confidence. Thus, a strong “upstream” dose of AI might encourage both possible adopting families as well as local health professionals in getting with the program. Newly fulfilled families and reducing a major social problem have powerful attractions, that is to say, especially when unbalanced. The myopia might be crucial. Many observers can have substantial (if not complete) confidence that negative “downstream” effects are probable, perhaps inevitable. Thus, other experiences with children deprived of early childhood nurturing and tender human contact lead one to expect major developmental difficulties for the children (e.g., Davis, 1940; Freud and Burlingham, 1944). Hence, also, adopting families might face unexpected challenges.

In both cases, dual central concerns require attention in the present view. But AI does not seem to provide the necessary attention. In sum:

Would AI approaches encourage the “upstream” blending of positive” and “negative” materials? Apparently not.
Would AI intervenors have “downstream” responsibilities for any neglect or shortfalls in such “upstream” blending into the decision-making processes? Apparently not.

When AI?

Relatedly, but with a more decidedly operational emphasis, the author is convinced that AI has some useful applications, but too little clarity exists about when, where, and how. The AI literature does not seem much concerned about this second query, however.

The following chapter, among other features, attempts to provide preliminary perspective on this clear opportunity, if not need. Broadly, Chapter 26 sketches numerous ways in which the full list of ironies can be acted upon, as well as in which AI can be better targeted.

REFERENCES


Reducing Ironies and Increasing Success Rates
Tactics and Strategies

So, this volume is done for now, a second time around; and this author believes it leaves a much enlarged foundation on which to build. It is not a “new OD,” but one definitely enlarged and enhanced.

Conveniently, this constitutes a good place to review the bidding on this volume. Initially, the focus falls on three purposes of this work, and then later highlights several tactics and strategies for moving beyond the present ironies while increasing success rates.

THREE UNDERLYING PURPOSES

This book highlights several ironies, each and all for a trinity of purposes. Both for convenience and because of their clear content, three purposes may be segregated as:

- The sense and scope of this effort
- The conviction of this volume
- The belief underlying this volume

Sense and Scope of the Effort

The basic sense and scope of the volume propose that OD interveners are coping tolerably well, even with several gaps and unclarities in theory and philosophy.
here in the United States as well as elsewhere. Separate headings will help make the point and, when appropriate, attention is paid to the ways in which the second edition extends beyond the first, in the essential particulars of expanding the range of applications as well as in increasing success rates. The main emphases are:

- Broader range and higher success rates
- New issue arenas of application
- Conceptual expansions
- Large systems as opportunities
- Broader contextual differences
- Differences between people
- Several “easy pieces”
- Greater clarity about change
- An integrative chapter

Recall that a major way of organizing chapters is in terms of “ironies.” The general format: the irony is that success rates are substantial, even given gaps or missed opportunities in knowledge or application. Commonly, the other side of the irony is that, with moderate effort, the gaps or opportunities can be exploited.

**Broader Range and Success Rates**

This second edition begins with some good news that substantially surpasses the initial edition. As Chapters 1 and 2 show, success rates are high in North America as well as abroad in what may loosely be called “developing” nation-states. And Chapter 3 adds the important caveat that success is not substantially attributable to positive-response bias or inadequate methodology. Moreover, OD capabilities are growing.

Granted, these success rates may surprise some observers, but agreement is growing and it gains powerful reinforcement from the broad range of political jurisdictions that have hosted OD or QWL applications. The overall picture suggests the view that ODers who care enough about their practice to write about their experiences—whether for publication or just for a record to learn from—are able to make the subtle and multiple adjustments at the many points at which the arts of consultation have to transcend what is known rigorously and scientifically.

The initial trio of chapters moves substantially beyond the first edition. Basically, the earlier edition had to rely on a modest success rate literature, especially two surveys of the effects of applications. Now, about two dozen such surveys exist, and they cannot be marginalized by charges that they reflect the enthusiasm of a few zealots. In addition, major confirming evidence from QWL studies is presented. All of this updates the first edition, and some of the research comes from recent days.
The case for robust effects of OD and QWL applications is substantial, then. Given OD and QWL resources, we know some places to go to increase applications and to at least increase the base for success rate estimates. Some are “easy pieces” (e.g., Chapters 21, 22, and 23), and some are more challenging (e.g., Chapters 4 through 6). But many attractive opportunities exist. Hence, the labeling of the first three chapters as the irony of ironies: substantial success rates, with substantial possibilities of raising them while expanding the kind and character of OD and QWL applications.

Broader Issue Arenas

Great increases in what may be called “issue arenas” also are possible on the face of compelling evidence, then, as Chapters 4 through 9 demonstrate in two different ways. Thus, the first three chapters in this sextet dwell on “replications,” as broadly designed. And the second trio deals with the normative textures of cultures or normative profiles of potential sites of applications. Both sets have clear expansionist tendencies: in the first trio of chapters, for example, the “replication” is lodged in a policy arena that can profit from guidance; the second trio seeks multiple senses of why and how OD/QWL can apply in Western settings as well as (for example) Confucian settings. Applications have a great potential for expansions in both clusters, but especially the latter.

These radically expanded zones for applications have clear attractions for expanding the reach-and-grasp of OD, as here conceptualized. Given sufficient wit and will, obviously, Chapters 4 through 9 not only will enlarge the territory for attention but also will enhance our ability to do the job. Hence, the cataloging of chapters 4 through 6 as an irony of substantial success rates while missing many opportunities for replication; and hence, also, the view of chapters 7 through 9 as OD being characterized by high success rates while there also exist possibilities of extending OD and QWL to a broadening range of cultural features here seen as “work ethics.”

Conceptual Expansions

Chapters 10 and 11 also suggest the usefulness of many possible ways of increasing applications and enhancing their ability to target intended effects. The foci there deal with a learning theory and better fit of designs to situations, neither of which has been a hot area of inquiry.

This second edition may help considerably in directing attention to several conceptual expansions. Their relevance seems clear enough, giving rise to the irony that the present diffusion of OD is great and growing, even as progress about OD as a method seems possible and can serve as a vehicle for more informed and muscular applications.
Large Systems as Opportunities

Chapters 12 through 15 show that huge additions to OD’s reach-and-grasp are possible via vigorous extension to large systems. Not that work in such territories is as rare as some ODers think. Examples above come not only from business, government, and health care but also from the politics/administration interface known better in government than in business, but likely to appear in related forms in many loci.

This quartet of chapters in one sense expands the sense and incidence of large-system interventions, but more centrally it brings to the attention of ODers that more has been going on in large systems than is commonly acknowledged. This is a useful service even though it cannot claim finding new territories to exploit. Rather, Chapters 12 through 15 require only an expansion of consciousness among both ODers and potential clients. Hence, Irony V draws attention to the substantial OD practice and success with large-system applications for which useful models already exist.

Emphasis on Contextual Differences

Irony VI shows important areas in which greater specificity is possible in OD applications that, in turn, should expand both the range of applications and their probable success rates. Two relevant chapters deal with kinds of crises and groups as well as individual properties. Reasonably, continued failure to distinguish crises and group/individual properties should result in making OD applications more of a crap shoot in a black box than they need to be.

The detail in this second edition, especially in Chapter 17, as aided and abetted by the next trio of chapters, adds materials that were largely unavailable at the time of release of the first edition. Hence, those chapters are worthy of special highlighting.

People, People, People

Irony IV focuses on one contextual feature in OD—the differences between people. This emphasis was not entirely absent in the 1990 edition, but it was substantially missing.

Chapter 18 draws attention to one way of sharpening focus and interpretation of OD efforts. This involves specifying differences between people, and it is basically all gravy.

Two other chapters help expand on the irony that OD applications have an attractive profile, yet miss an obvious bet: that differences between people can sharpen applications and interpretations.

Chapters 19 and 20 deal with people as a major element in contextual differences in several ways, with the emphasis on burnout dominating. Largely, the related theory and experience have developed in the past decade or so, and the
The Future of OD

detail above constitutes a major value-added in the second edition. For example, Chapter 20 illustrates the consequences for health of advanced burnout. That summary should at least surprise, if it does not shape an awareness that was missing in the first edition.

And What of Several “Easy Pieces”?

One develops some convenient rationales for many of the shortfalls illustrated above, but what of three “easy pieces” in Chapters 21 through 23? They constitute no great challenge to implementation, and promise substantial effects, but applications are far more limited than they could be, and easily so. Go figure, as a New Yorker might say. Hence, Irony VIII presents possibilities for quick and substantial leverage.

Clarity About Change

Chapters 24 and 25 add a subtle and powerful issue. Existing applications have been soft on different ways of approaching change, despite high success rates. The previous two chapters address this Irony IX which, in the long run, gives the promise of more targeted effects. The bulk of this volume should not obscure a simplicity, which, for example, plays a big role in the puzzling over “easy pieces.” The various sources of value-added just illustrated are not ends in themselves. Basically, this book seeks to use today’s praxis not as a place to rest but as a platform from which to leap toward greater comprehensiveness and greater success rates. Far more important than the sense, then, the conviction of this book is that we need not simply reconcile ourselves and our clients to the present level of achievement, attractive though that general record now seems. We should do better, and clients should expect us to do better, even though our average performance at this time isn’t at all bad.

An Integrative View

This final chapter tries to bundle the volume, by summary and highlighting. It concludes the review of the first of three underlying purposes of this volume as related to the scope and methods of the second edition.

A Simple Conviction

Let us move on to a second underlying purpose. The five chapters that precede this summary and integration, in sum, have a substantial potential for usefully expanding the consciousness of minds seeking to enlarge responsible freedom in organizations. To be sure, however, the three chapters also highlight a puzzle, with which others will have to deal: why not more of the easy applications? The question is not trivial, but no harm can be done by forcing the issue, as Chapters 21 through 25 intend.
A Foundational Belief

A third underlying purpose anchors this volume. Most immediately, the belief underlying this volume is dual—not only can we reduce the surprises in OD, but proximate ways and means are conveniently at hand; and we have no real alternative to expansion short of a kind of dark ages that could easily develop in our era of “more for less.” Ironies in OD praxis may be all well and good, but we need greater specificity and precision in what we recommended and why. Twenty-five chapters preceding this one do the heavy work of acting on this belief: they specify some convenient ways and means of doing better than average. These chapters do not exhaust available ways and means, but the sampler is representative of what can be done, why, and how.

The alternatives to reducing surprises—especially the unpleasant ones—are not attractive. Along with social and economic effects, and paling in comparison with them, the major OD-related consequences feature a substantially stagnant central core of concepts and ideas, along with overreaching extensions beyond that once-solid base. In short, these alternatives risk allowing OD to fall in on itself, as it were, with the old foundations being progressively less able to support the weight of burgeoning giant steps attempted in practice. This suggests a meta-irony: what appears to be bustling activity on the surface of things could in point of fact be coupled with an increasingly inadequate theoretical infrastructure, and this means growing trouble not only in the present but especially in the near future. This is the sense of Chapter 25.

So this book goes about acting on this sense, motivated by that conviction, both inspired by a belief that we not only can, but must, do something more because the ironies are not likely to take care of themselves. The spirit of this volume is upbeat and even optimistic—when the prescriptions are straightforward and available, when they are subtle and off in some future, as well as when those prescriptions are somewhere in-between.

What are the odds that the surprises can be reduced? This chapter proposes that those odds are pretty good, as it goes about detailing a skeletal agenda for OD.

The ironies may well bury OD, of course. But ODers seem to be coping with them well enough for the present, and substantial enhancements of OD praxis do not seem to require a great leap forward. Many opportunities also seem readily available, at least in substantial part.

THREE REACTIVE/PROACTIVE CLASSES

This volume can only illustrate how to go about acting on this trinity of sense, conviction, and belief, but the several illustrations have real substance to them. Let us summarize, with a real concern about repetition but with an absolute fear that some important things will not get due attention.
Conveniently, we identify three classes of responses that do something reactive about the ironies, or even some proactive things. That is, this chapter seeks to motivate an exploitation of the irony of ironies: how success rates can be so substantial while so many opportunities are available for enhancing experience and theory. The time frames of these classes are substantially identified by their labels: immediate or short-run efforts; intermediate-range amelioratives; and long-range or even far-out thrusts.

Immediate Responses

In the very short run, this volume suggests several reasonable ways to reduce ironies in OD, right now. For present purposes, consider only three immediate responses; taking advantage of some chapters as as-in plug-ins, inducing appropriate expectations, and making changes in the packaging of OD applications.

Taking Advantage of As-Is Plug-ins

A number of the chapters detail opportunities that require little or no further development before being put into more general use. Scarcity is not the issue. Indeed, chapters 4 through 6 and 12 through 21 can go now, more or less as-is. The point of as-is applicability may be established briefly and conveniently. The focus is on Chapters 18 through 24. Let us illustrate how and why they are available:

Chapter 18 requires only a minor technical adjustment in the way most surveys are processed, an easy accommodation using available computer programs, although the ethical and philosophical issues definitely need careful review. Performance appraisals constitute only one breakout variable that can be used to search for patterns in survey data and, in practice, they require the same kind of analysis as that applied to the demographic variables typically included in surveys—age, sex, and so on. As Chapter 18 suggests, however, performance appraisals may be a crucial breakout variable. Specific issues involve issues of confidentiality; respondents in the present study provided permission for researchers to access organization archives.

Chapters 20 through 23 similarly present as-is opportunities, and all seem to be associated with high success rates. Flexi-Time applications in the hundreds have been written about, and one has to persevere to find an overt failure. But only perhaps a quarter of eligibles in the workforce are covered, even now. Flexi-Place has made much less of an organizational splash, especially when disciplined by a committed search for greater responsible freedom. The demotion design has been applied several times, to my knowledge, and the effects have been all but uniformly
like those in the original applications. That design seems dramatically underutilized.

It requires only a little more of a stretch to see how other chapters can be brought into direct play in OD applications, in three distinct senses. First, elementally, Chapter 19 illustrates how the phases of burnout are associated directly with important deficiencies or deficits at work, and measuring the phases can provide important data for a profile of an organization’s climate and culture as well as for a profile of its individual members.

The phases may provide a revealing and strategic measure, as well as a convenient one. To suggest the point, Janz, Dugan, and Ross (1986) report a very high correlation (≈ .82) between a summary measure of an organization’s culture and the percentage of employees in the three most advanced phases of burnout. Other observers report that performance appraisals fall as burnout phases approach VIII (Golembiewski and Munzenrider, 1988, pp. 85–93); and mortality levels in health-care settings have been associated directly with the proportions of staff in advanced phases, so this author has heard from reliable sources.

Second, and more expansively, the burnout chapters present far more opportunities for as-is applications. Specifically, different designs seem appropriate for those in advanced phases of burnout than for those with modest degrees of that dis-ease. Mass team-building projects are in increasing vogue nowadays, for example, and differentiating burnout in teams may be very helpful in fitting one of several designs to specific teams rather than—as is almost always done—using a single design for all teams. A one-size design no more fits all organizational teams than one-size socks fit all feet.

Conveniently, and consistently, teams seem to have an affinity for extreme scores on burnout. Although burnout is an individual property, the typical group tends to have an affinity either for those in advanced phases or for those persons experiencing little burnout. There seem to be few groups that are substantially “mixed,” when it comes to phase assignments (e.g., Golembiewski, Boudreau, Munzenrider, and Luo, 1996). This affinity implies that the OD designer can have his cake and eat it too, in the sense of taking individual burnout into account and yet dealing essentially with groups rather than with isolated individuals. Both practically and theoretically, this suggests major points of leverage.

What is the extent of such organizational applications of burnout? No one can yet be definite about that: To suggest that magnitude in a single particular only, however, thousands of organizations conduct “stress management workshops” for (probably) millions of employees. Typically, learning designs do not differentiate differences between learners, when the case seems strong for the proposition that those in Phase I (for example) require a different kind of learning design than those in Phase VIII. Relatedly, burnout seems to have important
The Future of OD

group associations, but almost always only individuals are offered stress workshops, as opposed to intact work groups.

What keeps ODers from acting on such easy as-is opportunities, and in large numbers? The paradox is that clear analogs already exist, as Chapters 19 and 20 suggest. Beyond this benchmarking work, the hopeful (or even impertinent) short answer is that perhaps more will be done after people read this book. Much of the material has been published elsewhere over the years, but this reworked reappearance between two covers might leverage the impact of individual chapters. That is my hope, in any case.

Third, of course, this short answer will not do, for the question poses complex issues in the dissemination and diffusion of knowledge. In sum, your author hopes this book provides maps to useful territory, but these maps also leave much unspecified, not to mention undecided and even obscure.

Nonetheless, all or most of the chapters here provide opportunities for numerous OD applications, and (at least) they do not threaten the formidable success rates in Chapters 1 through 3. In short, OD values and approaches seem broadly applicable—in different working environments, as well as in different issue arenas. This hardly dots every i or crosses every t in the matter of diffusion, but the cumulative evidence clearly does not imply exotic values and fragile approaches.

The last word on the central issue of greater applicability with maintained or greater success rates has hardly been written. Usefully, here, Chapters 7 through 9 provide one perspective on this hardness of OD approaches and values, while much else differs.

Fourth, consider also Chapter 18, which poses no technical problems that force ODers to reevaluate long-standing preferences and practices—some matters of mere convenience and others that are determinedly philosophical and deep in a person’s being. Many canned programs provide ready access to such issues, but two tendencies at odds do exist in that chapter: the technical search for the highest-quality data available, and the populist empowerment of all respondents by assuring that their voices are safely heard, especially in large aggregates. The view in Chapter 18 is that, at least in the long run, the second goal will be achieved only as the first goal is increasingly met. So investing in a little additional data processing constitutes a cheap price and a tolerable risk, in the present view. Also, the risk will be most manageable in regenerative interaction systems, which is what OD is about in its most essential sense, as far as this author is concerned.

But some ODers disagree. That is, some reasonable people see the situation in another light, even some who have the data processing skills. They want to reduce specialist inputs in the surveying process, thereby increasing ownership and involvement by nonspecialists. Moreover, they want to maximize participation by absolute anonymity. The intent is a radically populist kind of action research, with maximum involvement and participation of “just people.” These
goals reduce the probability of—but do not eliminate—analyses like that illustrated in Chapter 18 and elsewhere.

The other view is seen as too limiting, as a general rule, even though I understand and applaud the insistence on the argument under certain repressive conditions (e.g., Brown and Tandon, 1983). What may seem to be harsh medicine to some is viewed here as in the enlightened self-interest of ODers, in addition to being in the long-run interest of other stakeholders as well as of broad social relevance.

Think about the choices for a minute, in response to pressing questions.

In the present view, at least in the long run, Chapter 18 and its kin (like the far-more subtle Chapter 24) deal successfully with the issues of leveraging data in OD efforts, and that is in everyone’s self-interest, overall. Who will pay good money and spend precious time to generate data that may be unrevealing, or even corrosive of one’s self-interests? Only the uninformed or perhaps the manipulator will use a survey as a public relations effort, with no serious intent to act on the data. Apart from earning a fee, relatedly, what virtues inhere in having such clients? And even considering the fee, what is the impact on the derivative credibility of OD practice and practitioners, at least over the long run? Moreover, no one will be served by potentially misleading data. And whom will such faulty data empower, in more than a temporary and ephemeral sense? In the long run, at least, probably neither clients nor respondents will profit. Hence, doing better need not sacrifice ranges of interests, although it may put some in jeopardy.

Inducing Appropriate Expectations

Again at the as-is level, this second edition also can help immediately moderate OD ironies by using our existing base of knowledge to create appropriate expectations. For openers, this discussion suggests three approaches to a more reasonable setting of expectations.

To begin, Chapters 1 and 3 will help create realism about overall success rates. Moreover, motivation to do better than average comes from all of the chapters, this author hopes. This may seem to be a motherhood-and-apple-pie recommendation, but it has a dual edge. Noting tolerable success rates may seem like so much horn-tooting, but the claim can become a standard that the client expects the ODer to achieve or surpass, and the ODer also has to meet a reasonable standard. These dynamics can be powerfully useful. Unrealistic assessments of success rates do no one any long-run good. The estimates in Chapters 1 through 3 set no easy standard, clearly enough.

One severe complication needs note here, as well as much future attention in OD. These success rates are not free-floating and untethered, of course. Thus, Chapter 24 urges, gamma change requires expectations that set high standards in OD. OD applications often seek gamma change—or changes in state rather
than in degree—and this implies much for the future’s evaluations of OD applications. It does no long-run good confusing gamma- with alpha-change conventions. Indeed, one can be seriously misled. To be sure, we suspect that failure to recognize gamma results in classifying more applications as failures than successes. But no one can say with certainty. This is an elemental, if speculative, sense in which success rates can be increased by paying attention to Chapter 24.

Now, this point about gamma may seem like a hard sell for both consultant and client; and as-is, right-now asks for too much. But the point nonetheless has its clear virtues. The data-processing costs and skills have to be taken into account, but only the assessment of kind of change permits an informed judgment about whether or not a particular intervention worked. Indeed, the discovery of gamma was inspired by a case in which the numbers indicated an intervention failed even as participants saw success. When alpha-only was assumed, the intervention seemed to be a definite failure. But gamma change seems to have occurred, and that required a reevaluation of effects. A “big bang” change had occurred, given a plural change view of the effects, while a conventional view indicated no intended change. Indeed, viewed as alpha change, the data suggested a regression.

In sum, specifying kind of change has real consequences. Illustratively, few of the evaluative studies represented in Chapters 1 through 3 distinguish kinds of change, and that constitutes an important source of tentativeness about the evaluations. It is not absolutely clear whether specifying (for example) gamma change will raise or lower success rate estimates, but the probabilities favor distortion or camouflaging of real effects.

So both client and intervenor have a real interest in gamma change, independent of the longer-run benefits that awareness of it may generate. Without testing for gamma, neither client nor intervenor will know whether an intervention worked. Only fads will exist without testing for their effects. We can all be thankful when it does not take very long for fads to run their course. We should adjust our thinking so as to make it occur sooner, and testing for kinds of change may be very helpful in that effect.

Even if reasonable, this recognition of gamma will seldom come easily. One large commercial survey operation may be prototypic. The initial reaction to gamma change was to fund a sizeable study by a think tank to trash the concept, perhaps even to deep-six it. The motivation may only be guessed at, but gamma did challenge the survey organization’s normal way of doing business. The second-look view? The later reaction was to recommend that clients consider using a routine test for gamma to gain better perspective on surveys in longitudinal designs. This author cannot be certain about the present state of affairs.

Also note that Chapters 7 and 9 provide a direct message about appropriate expectations. OD designs should have higher success rates when they fit the contexts in which they are applied.
Enough said, in general. Chapters 10 and 11 focus on one specific aspect of possible differences between individuals and groups, and suggest how the present core insight can be extended in a specific application.

**Making Changes in OD Packaging**

Finally, for the present short list of immediate ways to reduce ironies in OD, intervenors might well change some policies and procedures. For example, some OD contracts might well carry a 5 or 10 percent add-on to advance the science and the art, in addition to the costs of the specific objective sought. Third-parties might well specialize in putting such add-on funds to good use, providing at once an objective evaluation of an intervention that tests or even extends existing theories.

Such a policy would not only add to our knowledge base, and in that sense “put something back in the pot” to acknowledge a debt to the earlier investments of others, but it would also “pass on the favor” to future ODers in much the same way as today’s ODers are gifted by their predecessors. In addition, the use of third-parties has methodological advantages—or at least avoids implications of direct self-interest in cases where OD intervenors do their own research.

One can also conceive of policy-making that constitutes far more of a stretch. Consider what might be called the Fund for Displaced ODers. The idea is not a new one, and similar institutions exist for such an unlikely gaggle of professionals as chemists, pastors, and federal officials. The core dilemma is common: what of those professionals who are tempted, or ordered, or even coerced, to violate “standards of acceptable conduct” lest they lose their jobs or are otherwise punished?

The purpose of the fund would be to provide support for those ODers who are in trouble with their employing agency for doing the proper things. Such a fund could reduce the probability that individuals will cave in to repressive forces and, in the longer run, also might reduce the likelihood that some employers would try to apply muscle to their employees. For example, a person serving in an OD role might get some information, given in confidence, which some management group wishes to track down to its source. To inform might save the ODer’s job, but at great personal cost to self and perhaps others. If ODers were helped in facing that music on their own, matters might be improved all the way around (e.g., Golembiewski, 1986b).

**Intermediate-Range Initiatives**

Ironies II does not exist only in the here-and-now. Several of the chapters suggest targets that will not be available in the very near future, and yet do not have Star Wars time frames. A quartet of such immediate targets receives some attention. Discussion deals with, in turn, exploring the issue of the fit between design and
context, avoiding the reputation and reality of OD as a “mature product line,” and enhancing OD’s foundations.

Fit between Design and Context

Several chapters relate to the issue of the degree of fit, or congruence between specific OD designs and various contexts of application, though in different ways. Chapters 7 through 9 received some earlier attention, so let us focus on Chapter 10.

Chapter 10 deals with the issue of an “optimum prior discrepancy” in OD, which typically requires that people “unfreeze” as a first step in increasing their consciousness about the appropriateness of choice or change. Various discrepancy-highlighting approaches have been used in OD—feedback in a supportive context like a sensitivity training group, the sharing of 3-D images for purposes of directing confrontations, and the use of surveys.

Clearly, some degrees of discrepancy can be too much, and others too little. Greater clarity about which degree of discrepancy is appropriate under what conditions seems like a high-priority item for the research agenda of the near future, and as a guide to applications not very far down the track.

Ironies II presents multiple opportunities to check for degree-of-fit. Thus, Chapter 17 highlights the generic issue of which designs seem most appropriate under which specific environmental contexts or conditions. As yet, we have no comprehensive taxonomies for describing significant contexts. Moreover, useful taxonomies probably will be quite detailed—for example, some studies indicate there may be twenty or so distinct types of small work groups (e.g., Bowers and Hausser, 1977), a dozen or so “organizational archetypes” (e.g., Miller and Friesen, 1984), and perhaps two dozen types of individuals seem in the cards, given advanced research (e.g., Stokes, Mumford, and Owens, 1994). This catalog poses a real challenge for intermediate-range initiatives.

But no one can claim that nature promises a rose garden. Certainly, no credible argument can be made today for using simple and single interventions just because they are more convenient that way. The initial chapters on success rates imply that ODers do a tolerably good job of establishing the fit of design to context, in general. Nonetheless, this encourages confidence to still move forward, as contrasted with an encouragement to rest on OD laurels. The effort no doubt will require major tooling-up, and results may come slowly. But as we gain experience about the more precise tailoring of designs to the major varieties of contexts, so also will success rates increase. This provides powerful motivation for proceeding, posthaste, to test where good theoretical reasons exist for believing we have the beginnings of a taxonomy for strategic differences in context, as in Chapters 9 and 10. Nor is it necessary to wait patiently for conclusive research. Some situational differences—e.g., the phases of burnout (Golembiewski, 1997)—seem to be “centroids” in nature and hence are related in regular
ways to large catalogs of variables. While falling short of taxonomies, burnout phases may help powerfully in characterizing loci of application.

To be sure, neglecting such differences in context seems easy to do for certain designs, as in the cases of flexible work hours (Chapter 21), enlarging the concept of the workplace (Chapter 22), or developing new kinds of adverse personnel actions such as demotions (Chapter 23). Relevant designs seem to have a high propensity “to work,” whatever the conditions. But this is convenient only, not a stop order on progress toward comprehending situational differences.

However, even this understandable tendency to go with apparent winners must be resisted, lest we get into bad and irreversible habits. Even in the “easy pieces,” much remains unknown, and success rates seem to be in the position of getting a push from enhancements of straight-vanilla designs. So striving beyond convenience has much to recommend it.

Some tolerably precise sense of fit exists in both easy and complex cases, in fact, and although this is not the place for details, some sense of the possible can be communicated economically. For example, Chapters 17, 19, and 20 provide a working idea of a reasonable fit in cases of advanced phases of burnout, and especially those in the passive mode. Here, low-stimulus designs seem appropriate, if only because individuals classified in advanced phases are at their comfortable coping limits, if not beyond them. Low-stimulus designs include flexible work hour programs, job rotation, and mild role negotiation. High-stimulus designs for learning include interpersonal confrontations, sensitivity training, 3-D imaging, and so on, and these seem more applicable in cases in which individuals are active copers. Interested readers can consult more elaborate reviews of the fit of three types of OD designs to conditions at specific sites (e.g., Golembiewski, Boudreau, Munzenrider, and Luo, 1996, pp. 222–238).

In short, really satisfactory solutions to anything will involve major progress on everything. This need not be intimidating, but it might well be. Directly, we do not have to do everything at once, but the sooner we get started in earnest, on the more strategic fronts, the better.

This does not mean that we should proceed willy-nilly. Indeed, two intermediate-run priorities seem clear enough: expand the conceptual boundaries of OD practitioners and their theories and deepen the foundations of OD, especially in small-group analysis. In turn, the three subsections below focus on these priorities.

Avoiding the “Maturing Product” Label and Reality

OD will be well served by an expanding periphery, although not an exploding one, as well as by that solid central core rooted in process analysis and the values associated with the laboratory approach (e.g., Golembiewski, 1979, vols. 1 and 2). Basically, this expanding periphery requires a growing set of testable and tested designs for specific learning purposes, and it also implies growing knowl-
edge about the set of contexts in which those designs are applicable. Progressively, the expanding periphery implies such integrative thrusts:

- From interaction designs to structural and policy designs
- From affluent contexts to impoverished settings
- From growth to cutback or stagnation
- From high-technology industrial settings to rural subsistence settings
- From executive and managerial levels to workaday worlds at bench, shop, and assembly-line levels.
- From a fixation on management and executives to collaboration between all stakeholders: labor unions, management, consumer groups, and so on
- From work settings to family and leisure settings
- From organizational problem-solving and conflict resolution to national and global peace
- From North American and western European to global settings

Several clusters of chapters also illustrate strategic targets for expanding OD’s base. Thus, Chapters 7 through 9 in effect take a geographical approach to expansion, and Chapters 21 through 23 emphasize conceptual expansionism.

Several other clusters of chapters above illustrate possible growth-thrusts for OD. For example, Chapters 4 through 6 focus on several kinds of replications, for which useful models of application and analysis already exist.

One must nurture such initiatives and not merely announce them. Why? The motivation for the ODer is enlightened self-interest. Krell (1981) has usefully pointed up the several self-deflating features of maturing product lines, and these must be avoided and managed. Mature product lines tend to overstate efficacy, claim spurious or at least superficial differentiation of products and services, and cannibalize the research of the past rather than invest in the research of the future. Basically, mature product lines are milked for profits rather than serve as investments for the future. Only an expanding periphery will help avoid these dire concomitants of matured product lines.

Investing and Expanding OD’s Foundations

In both tactical and strategic senses, movement toward intermediate-range initiatives will rest on concerted research and application pushes to small-group analysis which, in many ways, is at the heart of the OD technology. If this is an essentially correct statement, much work will need doing.

The point is that small-group analysis must be reinvigorated. Consider the oft-heard question: Whatever happened to small-group analysis (Lakin, 1979)? My answer has a familiar ring to it:

The problem of small-group analysis is that it succeeded in inspiring applications substantially before a satisfactory research map was
charted. This is not the worst problem to have, since problems there will be. It is merely a fact that what we learned in small-group analysis through the 1960s was so powerful, however incomplete, that it fueled decades of applications.

So we can answer that small-group analysis is not somehow lost or forgotten. It is manifest in much applied behavioral science, and especially in OD. In fact, it succeeded too well, too fast. It remains our task to assure that exuberant success does not lead to failure.

Although not lost or woebegone, then, small-group analysis was and is stuck. At first, it was quite a comfortable stuckness, given the heady and even intoxicating advances in applications that resulted from the beachheads begun in the 1930s and 1940s (e.g., Golembiewski, 1962a). Of late, the stuckness is far less comfortable, and even downright uncomfortable at times. Consider this contrast. At one time, the achievements of small-group analysis permitted applications, even encouraged them. Increasingly, gaps in our knowledge of small groups act as constraints against the reach-and-grasp of applications and enlarging OD success rates. The critical need is for a taxonomy of groups.

Try an easy example. It is not extravagant to conclude that some groups are highly personal or even “warm,” whereas others can have an atmosphere or climate that is “stiff,” “cold,” or even defiantly counterpersonal. Reasonably, then, the same design seems unlikely to be equally appropriate for groups of these two kinds.

Let us grant that the sophisticated OD intervener working with one or a few groups might well adjust to such differences, even if semiconsciously. The rub? ODers have begun working, in earnest and often simultaneously, with several or even many groups, and that trend no doubt will intensify, if only to make more cost-effective use of expensive OD services. Hence, the contemporary OD intervener stands in growing need of supports from theory and technology. Differences between groups require both prior measurement and classification, and here small-group analysis is no longer adequate.

How to get small-group analysis unstuck? Chapter 16 proposes one approach, which seems generally applicable. It seeks to relate the choice of designs to differences in prevailing crises or presenting conditions in groups, as well as to differing distributions of the burnout phases of group’s members. Similarly, Chapter 17 suggests direct covariation between group properties and burnout, with significant implications for both theory and practice that have been confirmed by other research (Golembiewski, Boudreau, Munzenrider, and Luo, 1996). These findings suggest reasonable ways to get at burnout by restricting the development of group properties that seem associated with the advanced phases—low cohesiveness, high supervisory pressure and punitiveness, and so on. Conveniently, the small-group literature tells us much about how to induce
high cohesiveness, which seems to be a function of member liking, prestige to task, and prestige of membership (e.g., Golembiewski, 1962a, pp. 149–70), in largest part.

Far more expansively, early work (e.g., Bowers and Hausser, 1977), suggests a taxonomy of twenty-plus types of groups. Presumably, relating different designs to individual types or clusters of them will heighten success rates. Much the same seems to apply to the bio-data subgroup taxonomy by identifying individuals (e.g., Stokes, Mumford, and Owens, 1994).

Of course, OD is not only small-group analysis, as central as that area of inquiry is and will remain. For one thing, OD needs a taxonomy of organizations at least as much as a taxonomy of groups. Here again, the research literature contains glistening suggestions (e.g., Miller and Friesen, 1984) that taxonomies exist. Clearly, one structural type does not suffice (Golembiewski, 1962b, 1987; Mintzberg, 1979).

Seeking OD ↔ AI Reinforcements

As Chapter 25 should establish, the interfaces between OD and Appreciative Inquiry seem to promise potential riches for research as well as application, and the present purpose seeks to sketch several opportunities for testing such interfaces in short and especially intermediate time-frames. AI is increasingly recognized and recommended as a vehicle for consultants. Here, the basic question is: given what we know, what seem the safest uses of that technology for learning or change? Specific research is rare, but some useful guides for applications seem reasonable.

In Chapter 25 of this volume, as well as elsewhere (Golembiewski, 1999, 1998), the author takes full opportunity to give really appreciative attention to AI, and this section builds substantially on the last source. The view there is top down, as it were. Viewed from broad conceptual and empirical perspectives, that earlier work evaluates AI as a comprehensive corpus of experience and research. Major shortfalls exist, stereophonically—for AI viewed as theory, research, diagnosis, and application.

On the general proposition that cemeteries are the only human artifacts not characterized by change, this article looks at AI from a particularistic perspective. Given its shortfalls as comprehensive theory/experience, are there uses that can safely be made of some AI approaches or perspectives?

The answer is affirmative and derives from an earlier source that has only narrow circulation (Golembiewski, 1999), and discussion below builds on and extends that source in two major ways of exploiting AI’s value-added.

2 × 2, again Visitors (or conquerors!) from outer space no doubt will be impressed by how often behavioral science is expressed in terms of 2 × 2 matrices. And so it is here.
Exhibit 26.1 provides a useful (if simplified) space for AI applications in terms of four “windows”:

Window I encourages little caution or concern, and there AI may even provide useful leverage. For example, consider the induction of “groupiness” properties such as high cohesiveness in experimental situations, which has many applications in research and practice. AI approaches might well quickly and economically serve such purposes—for example, in testing whether Sherif’s (1935) autokinetic effects will be more marked for AI inductions than for those induced by simply assembling subjects after each individual runs through the protocol. Subjects might be told, sometimes truthfully and sometimes falsely, “you are going to be with people you really will like and will find it pleasant to work with.” Real-AI and pseudo-AI inductions then could be compared with merely assembled clusters of individuals.

Window II need not detain very long, since its high cost/benefit ratio should deter most uses, but examples seem easy enough to imagine. Thus, experiments might try to assess whether subjects high on Internalization should prove more recalcitrant to AI induction than those high on Externalization (e.g., Rotter, 1966).

Such effects would not surprise, of course, and they imply modest theoretical and practical leverage, while having high costs as a downside potential.

Window III encourages polarized reactions. Consider an executive team anticipating a rough review by an active board of directors. Executives...
might well favor an AI application, on the general principle that this would heighten integrative versus punitive impulses in board members—as in board responses to the AI question: Think of really good things that management has done. Combative board members might favor, in contrast, Barry Oshry’s power lab as pre-work to heighten their predispositions toward conflict and contention.

Window IV concerns most, even as it contains the most potential for impactful AI applications.

Questions dominate with respect to Window IV, as in the case of the adoptions by Westerners of (apparently) thousands of Romanian orphans who were deprived of human comfort for extended periods, and more warehoused than cared for. TV dramatized their plight, potential adopters grew numerous, NGOs were anxious to mediate, and Romanian authorities proved willing, and perhaps even anxious to place the children outside of Romania.

Would AI, with its focus on positive and peak experiences, simply heighten these several enthusiasms? And would that occur so much so that, whether by conformist “group think” or even coercion, these AI-induced forces led to the neglect of some dour and even devastating possibilities? Let us grant that AI approaches could generate highly positive reactions in adoptive parents and local authorities by emphasizing the life forces that adoptions might trigger. Basically, in addition, what is the probability that the orphans/foundlings would be unable to develop “normally” in their later lives? And would AI encourage/discourage raising this unappreciative but apparently stubborn reality?

Substantial evidence encourages real caution in this connection, judging from cases of various degrees of isolation of youngsters from human contact. The evidence has a long heritage and covers a range—from “feral children” allegedly raised by animals or otherwise isolated from human warmth and contact (e.g., Davis, 1949; Singh and Zingy, 1943), to the neglected or abused (e.g., Levy, 1937), and including children separated from parents—as by economic privation or war (e.g., Freud and Burlingham, 1944). The chilling prognosis is that some unknown but probably substantial proportion of such children will only partially or even never recover. And this probability, to me, significantly motivates the early and insistent introduction of “negative” materials in the early decision-making by local health authorities as well as by potential adopters to encourage realistic choices.

**Five Possible AI Applications** Within the context of this $2 \times 2$ matrix, what proximate (if perhaps cautious) uses of AI approaches seem usefully employed in approaches to planned change like OD or QWL? Generally, the targets for such AI applications are cells I and III in Exhibit 26.1, and then perhaps cell IV.
Here, Gervase Busche (1998) provides useful guidance. As contrasted with a comprehensive view of “full strength” AI, Busche suggests five “spot uses” of AI that seem broadly applicable in consultation and planned change.

The Power of the Narrative or “Story”  Basically, AI prescribes that people tell stories about personal “high points,” and these in turn can guide action-taking and even reduce “resistance to change” in the sense that “candy helps even strong medicine go down.” For example, the contributors of stories about a “peak experience” in work teams might be confronted with this challenge: “And what have you done recently to try to re-create such a peak experience in your present team?” And in addition, perhaps: “What has kept you from attempting such transplants?”

This suggested design-loop probably does not qualify as pure AI, but two points seem clear enough. Thus, the approach might contribute a powerful impetus to learning, and perhaps especially among wary or even suspicious persons. Moreover, stories are powerful in many traditions of analysis and inquiry—including OD, ethnomethodology, and a long list besides. Stories constitute a kind or storage system for beliefs, fears, prescriptions, ideals, and generally for all that is human. Accessing this “storage system” can highlight expectations about the features of social systems that operate on both the wary and the insensitive; and the focus on “positive stories” may have a soothing effect, and even powerful ones.

However, AI certainly has no monopoly here; thus, reliance on stories need not buy into AI’s epistemology or world view; and the history of stories in social science suggests that the apparently simple technique can have some tricky wrinkles (e.g., Herzberg, Mausner, and Sanderman, 1959; Argyris, 1971). Perhaps more salient, AI seems nowhere very clear about what to do if sharp differences characterize the “peak experiences” or “positive stories of the targeted individuals.” Finally, for present purposes, one cannot simply exclude the learning potential in a little design ju-jitsu, as on the themes: What have your experiences in other groups led you to try to avoid in your new groupings? and What, if anything, have you done in your present group to avoid undesired consequences? And if nothing, why not?

The “Heliotropic Hypothesis”  Many life forms gravitate toward light, of course; and by enthusiastic extrapolation, it may be proposed that social forms and norms gravitate toward the “positive.” Hence, AI might direct energies in just the direction that people prefer, and this might well generate a kind of boost to action. Cooperrider refers to such an effect as the “heliotropic hypothesis.”

This author goes only part of the way with this hypothesis, despite its attractive and uncomplicated view of social norms and forms. To put it colloquially: “You can catch more flies with honey than vinegar.” Agreed, for many situations.

Even if this hypothesis is granted, however, the search for ideal social forms
The Future of OD 625

and norms also faces other and far-larger issues than whether or not somebody’s motivational systems can be engaged in pursuit of attractive ends only. The key issue always is: Motivations for what purposes? The human record seems compelling in this regard. Under some circumstances, to begin, human systems can degenerate to bestial levels (e.g., Chang, 1997) as well as they can rise to the noble. Most of the time, of course, most of us are somewhere in-between the two end points. Put another way, human systems can level “down” as well as “up,” under the same impetus of substantial pressures toward uniform behavior. The issue is not so much a choice between “missiles” as only more or less powerful vehicles. Rather, the choice involves “guidance systems” relevant to direction as well as, if not more so than, magnitude of force. That is, an unguided powerful missile often will be worse than a guided but less-powerful one.

Consultants in planned change are well advised to cultivate a sense of the difference. A powerful learning design may attract, but well-targeted ones make more progress, more safely, more of the time.

AI ideation as guidance system elicits concern, and to the very degree that it induces powerful forces. ODers devote great energies to the normative character of their designs, but AI in its dominant form assumes the social constructionist view that there is nothing special about any social form or norm. That is, goes the social constructionist view, all social artifacts are mere “enactment” that have only arbitrary status, and then presumably only until they are “unenacted” with (in AI) no concern about consequences. This is convenient, no doubt, but also vulnerable to the possibility that nothing special—or even worse—is what one will get from any technology that does not give detailed attention to the character of its guidance system, rightly viewed as direction as well as magnitude.

In sum, even the spot use of AI must be guarded in this second sense. For example, AI’s aversion to “negative” stories may encourage an incautious optimism about the tractability of social facts or beliefs. Oppositely, my sense of prudence suggests balancing the “positive” by the dark clouds often associated with most/all silver linings. Moreover, at various times, some very strange things have appeared “attractive.”

This author is not alone in such concerns. Basically, social constructionist ideas are changing (e.g., Gergen and Thatchenkery, 1996), and no major efforts have been made to build such changes into AI fundaments. Indeed, it seems very unlikely that AI could have developed in the context of these emerging reevaluations of social constructionism. Without a doubt, AI would not have had its pronounced characteristics—especially contra diagnosis and empirical research—in the absence of early social constructionism.

The basic shift in conceptual content of “social constructionism” appears in another place (Golembiewski, 1998, 1999), but a snippet here will be helpful. At its base, two evolving social constructionists urged a three-part revisionism (Gergen and Thatchenkery, 1996, pp. 360, 367):
“. . . although [or, perhaps since] the vast majority of scientists and practitioners see [social constructionism] as tantamount to nihilism . . . we [now emphasize a] reconstructive theme”

“. . . there is nothing about postmodernism that argues against the possibilities of using empirical technologies for certain practical purposes.”

“Although postmodern critique undermines the function of research in warranting truth and shifts the empirical emphasis to more local and practical concerns, it also invites a broad expansion in the conceptualization of research.”

In short, these observers clearly conclude that constructionism can go too far, lofty goals notwithstanding. That also is the present view of the early social constructionist impact on the essentials of AI, where it served to delegitimate empirical research and diagnosis.

This is a critical but costly service. AI correctly perceived that empirical theory needed delegation for their purposes, because “. . . creative theorizing has literally been assaulted on all fronts by proactioners and academic theorists alike . . . [especially by the very insistence on the criteria of ‘normal’ science and research] . . .” (Cooperrider and Srivastva, 1987, pp. 150 ff.) But the associated costs are severe. As Gergen and Thatchenkery observe (1996, p. 365):

[Social construction’s] postmodern critique operates as a major form of delegitimation. In the scientific sphere, [social constructionism] contributes to a loss of confidence in rational theory, in the safeguards of rigorous research methods, and in the promise of a steady increase in objective knowledge.

Highlighting the “Organizational Unconscious” Variants of this useful view have long been with us, and even at the earliest days of OD (e.g., Bion, 1948, 1949, 1950, 1951), at least in principle. Nor have contemporaries failed to reinforce the point (e.g., Kets de Vries and Miller, 1984). Exemplars of this approach urge in common that an “inner dialog” is going on in all organizations, and its dynamics hold the keys to sensitive diagnosis and hence to effective prescription. It is not an extravagant notion that both “stories” and “positive stories” can help gain insight about such internal dialogs, as well as about ways of managing or resolving those dynamics.

How can AI help enrich this long-standing view? Here, basic AI may be put to on-the-spot use but, again, only with caveats. Paramountly, perhaps, other approaches also exist, as in the contrast of espoused and action theories pioneered by Argyris (e.g., 1971). Directly, AI cannot now claim clear advantage over such other uses of stories. Of course, future research may yet justify such a claim.

Indeed, AI’s sometimes remarkable aversion to diagnosis implies that only a gentle reliance on AI is justified in this third possible spot use; that is to say,
for one thing, “organizational unconscious” approaches are into diagnosis in a very big way, and it remains unclear how AI can serve this dominant objective. Far more centrally, no specific organization unconscious has any special claim to truth value, and all must be evaluated normatively. If an “unconscious” exists, in short, should it be supported or resisted? Appreciative Inquiry proposes no special capacity here, or even interest. Directly, organization unconscious can help consultants with “is” issues, but consultants also have to deal with “should” concerns (e.g., Gellerman, Fankel, and Ladenson, 1990).

Helping Resolve Paradoxical Dilemmas Busche (1998) also highlighted how a kind of AI spot use helped a “stuck” team become “unstuck” as its members became able to “discuss the undiscussable.” After a period of going-nowhere-ness in a conventional team-building mode, Busche suggested an AI variant: members should describe a “great team.” Members then were instructed to reflect on features of those “best” groups that had not yet appeared in their present group, and perhaps especially why.

After some flailing about, one member put his mental finger on a transformative recognition in that other group that galvanized members elsewhere but had not yet occurred in the present group. “We recognized that trust costs less than distrust,” the member observed. That powerful generalization from another setting helped resolve an apparent paradox in the present group, where the “stuckness” resulted from a deep distrust that seemed to require such a quantum leap of trust to overcome that members saw it as out of reach—hence the undiscussable that initially resulted in the group’s stuckness.

This is a nice example, and it usefully suggests a mini-use for AI; but at least three cautionary caveats seem appropriate. Other designs could have induced a similar effect—indeed, some conventional exemplars intend to trigger just such effects early in a design for a target group locked into an “avoiding” or “flight” mode (e.g., Golembiewski, 1979, vol. 1, pp. 308–314). These alternatives owe nothing to AI, as well as vice versa. Indeed, such alternatives seem rejected by AI ideology.

Second, it might even help to pose the question of “what is the worst thing that could happen” to a group of individuals if talking about the undiscussable were to occur. This discussion might encourage unstuckness. In addition, uses of “gap analysis” (e.g., a difference between ideal and now scores on the Likert profile of organizational characteristics) can be used to generate just such a “booster shot” as that attributed to the AI variant described by Busche (e.g., Golembiewski, 1979, vol. 2, pp. 18–27).

Perhaps the most subtle issue involves “crisis of agreement” or “going to Abilene.” There, all members actually agree, but group membership being highly valued, all are afraid to test for that agreement lest they be denied membership. All members thus act as if they were the only deviants. Here, AI might well
compound the problem, and especially if the “positive stories” on how great it was to be a group member do not significantly agree, or are superficial or even feigned.

“Tracking” or “Benchmarking” Finally, for present purposes, general AI processes can be engaged with profit, and for several basic reasons: that “you often find what you are looking for”; that “where you look determines what you see”; or that “you often will get what you ask for.” “Expectation effects,” in sum, can play major roles in consultation, as in the often-remarked tendencies of authority figures to accord superior status to those about whom they receive bogus information concerning the latter’s high achievement or potential (e.g., King, 1974). That is, look for the positives in people, and this may encourage people to provide them. Good enough. Great, even.

Care seems appropriate, however, in even this final illustrative sense of AI’s spot use. Such effects as those implied above can be mischievous, and even dangerous if there is also a reality “out there” as well as “in here” in the perceiving person. Much AI ideation assumes only the latter reality, but that seems oversimple (e.g., Sherif, 1935). If “truth” comes in two or more varieties, and their several combinations—“objective” as well as “subjective,” or out there as well as in here—some of the positive support for AI is resolved into evidence about the ease with which people can come to accept simplisms, or even palpable falsities. This has its mixed consequences, of course.

“Finding what you look for,” then, takes on aspects of easily induced and possibly unfortunate outcomes to guard against. In awkward contrast, AI prescribes pursuing just such superficially attractive ways of finessing choices between social forms and norms.

Reprise Appreciative Inquiry is here viewed as a system of thought in need of circumscribing, and at a time when much remains unknown about its features and consequences. The concern is exacerbated by a major datum: the common if anecdotal attributions of powerful effects to AI applications, hence this approach to indicating safe areas of value-added by AI, which can guide consultation until a more final resolution of AI capabilities becomes available.

THE LONGER-RUN VIEW

The discussion now formally changes its time line from the intermediate to the long run. In reality, the discussion cannot be very precise about the cutting points. Small group and, especially, organizational taxonomies no doubt are not around any foreseeable analytical corner; and the same certainly seems true of the practical application of the foci of several other chapters in this book—e.g., trinitarian change.
But there is no doubt about the usefulness of a temporal focus. Briefly, this section tries to pinpoint major transitions toward long-run research and applied objectives in terms of three themes—meaningfully labeled both urgent and long-term. They include the centrality of including structural features in OD/QWL applications, the need for configurational analysis, and the practical urgency of a related and supporting Manhattan project in the organizational scenes. All three emphases appear elsewhere, in diverse forms (e.g., Golembiewski, 1986a), and here both modified and substantially reconfigured for present purposes.

**Toward Encompassing Structural Features**

This first theme will be no piece of cake, despite the undeniable and multiple attractions of viewing OD/QWL as a three-legged stool—as interaction, policy/procedures, and structure. The basic irony is that OD success rates are high, and yet structural change is commonly neglected. Given their clear relevance and power, great leverage inheres in structural initiatives. The attractions/difficulties of reducing this irony are usefully illustrated here under four subheadings.

**The Dominance of Normative Overlays**

Structural emphases do appear in the chapters above, but they obviously do not get coequal emphasis with interaction or perhaps even with such policies/procedures as flexible work hours. The same is true of all OD.

How to gain perspective on this fateful bias? The focus in OD is on what may be called “normative overlays” to encourage attitudes/behaviors that integrate the several classes of fragmentation or balkanization encouraged by the usually present traditional model of organization, or bureaucracy. The point is neither novel nor subtle. Indeed, the basic bureaucratic model clearly emphasizes separation vs. integration. That basic point escapes few observers, early or late. To illustrate, Figure 26.1 structures three activities yielding some product or services. The various separate tendencies there receive detailed attention in many sources (e.g., Golembiewski, 1993, 1995), with basic difficulties entering in the fact that Figure 26.1 structures quickly grow “taller” as employment increases. In technical terms, such structures require large GMUs, or Golembiewski Managerial Units—conveniently defined as that portion of an organization that reports to an integrative superior who can make a reasonable decision about the entire flow of work. That integrative superior is M_{ABC} in Figure 26.1, and the GMU circumscribes a much bigger “basic organization” than the GMU in Figure 26.2, with attendant difficulties in the length of communication chains, in the conflicting self-interests of S_A, S_B, and S_C, and so on.

Degenerative interaction illustrates one outcome of this basic fragmentation. That is, trust between the S units in Figure 26.1 will tend to be low, with the consequences that openness and owning will be low and risk will be high.
**FIGURE 26.1** Bureaucratic model of organizing. Basic departmentation focuses on separate functions, processes, or activities necessary to produce some product or service. Vertical lines or channels of authority dominate. Span of control is narrow—a small number of people report to any superior.

**FIGURE 26.2** Flow-of-work model alternative to bureaucracy. Basic departmentation focuses on aggregating all activities required to produce some product or service. Horizontal relationships reinforce and augment vertical ties. Span of control can be very wide—many people can report to any superior.
Hence the primacy in OD of designs targeting regenerative interaction, either as cultural and skill preparation for subsequent attention to policies/procedures and to structure, or as a sole intervention intended to serve as a crosswalk or “normative overlay” to counteract the fragmenting forces of traditional structures that are not the direct target of interventions. See especially Chapter 14.

“Normative overlays” are common, of course. Programs seeking to enhance upward feedback in organization may prescribe, “Tell it like it is.” Or quality enhancement programs may urge, “zero defects.” Perhaps the most exuberant version is the highlighting of “boundaryless organizations” (e.g., Ashkenas, Ulrich, Jick, and Kerr, 1998). All of these prescriptions are solidly motivated by the fragmenting forces of the bureaucratic model, but some of them do not spiritedly and explicitly provide replacements for the bureaucratic model. In one case, for example, a network approach was used to help interface federal institutions deliberately kept at a distance by the constitutional separation of powers. Similarly, strategic planning (see Chapter 13) often serves largely as a normative overlay intended to lesson the fragmentation of basically bureaucratic models of organization.

The Emergence of “Networks” in Theory and Practice

Because it gets increasing attention nowadays, as well as passing attention in places like Chapter 15, the concept “network” deserves a few words here. Basically, “networks” urge the bureaucratic model to recognize that “no single, comprehensive organization can command all the needed resources” for providing complex goods or services (Kim, 2002). This acknowledges the limitations of all hierarchies or simple markets, but seeks to transcend those limits by building on and beyond the elemental forms.

A body of experience and some theory is beginning to emerge (e.g., Hjern and Porter, 1991; Chisholm, 1998), and it reflects penetration of OD/QWL values and approaches, especially under the rubrics “inter-OD” or “macro-OB” (e.g., Miles, 1980). This does not surprise. For example, regenerative interaction often will help support networks. Chapter 15 reveals substantial success rates for one variety of such work—OD/QWL applications at the politics/administration interface toward the top of organizations, especially public ones. There, many believe that OD applications are low or doomed to failure, as due (for example) to the separation of powers. Special problems do exist at the interface, but attempts have been made to deal with them, even at top levels (e.g., National Academy of Public Administration, 1992; Helley, 1994).

Basic Structural Change at Workaday Levels

Although its volume always could be increased substantially, OD interest in low-level structural change exists in some volume—both early as well as late in OD’s history (e.g., Marrow, Bowers, and Seashore, 1967; Golembiewski and Rountree,
1999), as well as in business as well as public loci (e.g., Rainey and Rainey, 1986). Of course, QWL applications commonly have structural components—as in technostructural designs, job enrichment, and so on.

Basically, such low-to-mid-level applications build around flow-of-work structures such as that illustrated in Figure 26.2, whose basic thrust is integrative. Thus, the performance of $S_1$ vs $S_2$ vs $S_3$ units often can be directly compared, which induces many forces toward self-motivation and self-discipline. This integrative tendency is developed in detail at many other places (e.g., Golembiewski, 1989, 1995b). See also Chapters 5 through 12.

The virtues of a sharp increase in related OD/QWL applications at low-to-mid-levels seems quite obvious. Thus Chapter 14 reflects major benefits to an organization’s feeling-tone as well as to its performance and costs. There, the basic restructuring approach is revealingly labeled “shared care.” Of great practical and theoretic moment, Figure 26.2 structures also tend to be “flat,” with the small number of levels facilitating up ↔ down communication. In addition, the identification of individuals is heightened at low levels. This is a powerful double-play.

Basic Structural Change at Overhead Levels

This fourth category of structural interventions is at once the least common but also the most promising. Thus, several chapters—especially 12 through 15—reflect the advantages of organizing toward (or around) a Figure 26.2 model, at senior and executive levels as well as at low-to-middle-levels. Figures 26.1 and 26.2 apply directly there. In the latter case, for example, the smaller GMU refers to “shared care,” a “strategic operating area,” or a “division.” Contributing to the same conclusion, if from a different perspective, Figure 26.1 structures seem associated with a dour panoply of effects, which Figure 26.2 structures can help ameliorate or even avoid. Chapters 19 and 20 perhaps best relate to this basic message of what to avoid and what to approach in structuring work. There, the focus is on burnout, in which increases/decreases seem associated with the two sets of work structure in Figures 26.1 and 26.2, as well as with the different syndromes of effects associated with each.

Summary

Overall, then, a major irony in OD involves substantial success rates along with variable but always-significant shortfalls in taking structural features into explicit account. Presumably, such shortfalls would show up more frequently in no effects, or fade-out effects, than they would appear as fade-in effects. Presumably, also, such shortfalls in taking structural features into explicit account should reduce success rates, in general. Reasonably enough, the success rate estimates in the earliest chapters no doubt understate the levels of success that more careful OD/QWL applications can attain.
Structural features are a vast terra incognita, on the evidence of several chapters above, and two basic points seem clear enough. We can easily do far better in including structural features in OD design. Moreover, that targeted effort should soon, if not immediately, have the effect of increasing both the reach-and-grasp of technologies for change based on values like those common to OD and QWL.

**Toward Configurational Analysis**

To move on to the second part of this chapter, much organizational analysis and prescription reflect an apparent irony: the more work concentrates on a research area, the less we seem to know. That is, initial studies may isolate robust patterns of relationships, and follow-on work sometimes replicates the initial results. Typically, however, cases of random variation and contrary results soon come to dominate what were once-promising research initiatives. And the search typically moves on, quickly, in search of the next attractive focus for inquiry. The overall sense of it is a boom-town environment associated with discoveries of gold that turn out to be largely dross.

This tendency tests one’s confidence in ever coming to firm grips with nature, and examples clearly abound. Consider the early excitement about the ratio of administrative (A) to production (P) employees as organizations increase in size, which suggested the general rule that A increases at a faster rate as P grows linearly. In this process, organizations have to struggle to support their own escalating weight, or even break down. The general reader was titillated by popular treatments, such as Parkinson’s Law, which (roughly translated) proposes that work expands to fill the time of whatever staff is available. Moreover, those laws propose, staff increases whether the work increases, decreases, or remains constant. That motivated many a chuckle.

One of the greater outbursts of research excitement attended Haire’s (1959, pp. 272–306) reliance on biological models to explain line and staff growth, using Jack and the Beanstalk as his guiding metaphor; but that reaction was soon dulled by the impact of contrary findings. Less than a decade later, twelve A:P studies provided diverse fodder for a review article (Rushing, 1967): four studies showed no significant difference in the A:P ratio in organizations of different sizes, two studies showed increases in the ratio, and six found decreases. Subsequently, Pondy (1969) suggested that inter-industry differences argued against any general rule; and Child (1973) completed the regression by proposing that, in any case, A was too heterogeneous to permit any coherent test of the relative rates of growth with P.

The pattern is repeated, time after time, in many areas. For example, the results of the decade-long fascination with bureaucratization and centralization spearheaded by the Aston Group can be written in similar terms (Miller and
Friesen, 1984, pp. 13–14). Hence, some observers note the boom-and-bust quality of many areas of behavioral and organizational research. Less-respectful souls refer to “this year’s sheep dip,” a view not only discouraging cumulative research but also encouraging cynicism and overselling.

But the irony seems only apparent, and hence pessimism may not be appropriate about coming to grips with organizational phenomena. Mini-histories like that sketched above should be expected, in short. It would be astounding were they not to occur, in fact, given their underlying general approach to research design.

Consider the situation in organization analysis described by Miller and Friesen (1984, esp. pp. 15–17), who emphasize that researchers tend to assume that an organization is an organization. From there, conflicting results are only a matter of time.

Mintzberg’s (1979) five structural configurations suggest why this outcome is all but fore-ordained, so fragile is the core assumption about homology. He distinguishes simple structure, machine bureaucracy, professional bureaucracy, a divisionalized form, and ad hocracy. Miller and Friesen understate the patent implications of these configurations for organizational analysis and praxis, but their main point deserves emphasis:

Now, assuming . . . that at least a good proportion of organizations tended to adopt them or some other configurations of their structural parameters—what would happen if different kinds of organizations were mixed in research samples and then relationships gauged . . . ? We believe that we already have the answers in the [mini-histories sketched above] (1984, 14–15).

Specifically, a machine bureaucracy might well tend toward advanced centralization, whereas a professional bureaucracy might generally favor decentralization, or even chaotic localism. Moreover, both might be efficient, for some times and places, but not for all or most.

The implied lesson for OD is at once simple and daunting. Apparent contradictions in existing research will be reconciled, if at all, only after different organizational configurations are specified. The reader has heard the point before.

Hence the centrality in organization research and applications of “configurational analysis,” which seeks some manageable number of sufficiently distinct clusters of multiple attributes. These may be used, in turn, to guide the search for consistent subsets of covariants. Of course, the persevering reader will conclude that Ironies in Organization Development II is in effect an effort to start some major reasonable next steps toward configurational analysis in OD. You bet!

The last words definitely will not be written here concerning configurational analysis: indeed, little progress has been made between the first edition and this
The Future of OD

update. The jury is still out, for example, concerning the National Organization Survey (Kalleberg, Knoke, Marsden, and Spaeth, 1994: full issue), which seems a substantial effort to shake up the present stuckness in organizational analysis. Analyses like that of Facer (2002) will help evaluate NOS's fate.

Nonetheless, four points about configurational analysis seem safe enough. First, the approach seems to be a reasonable way to overcome stuckness or even despair born of conflicting research results. For example, small-group analysis has treaded water for several decades, and no small part of that record rests on a failure to distinguish families of differently configured groups (Lakin, 1979). Consider only a simple illustration. What kind of intervention or design is best for team building? Task differences loom large, of course. One useful approach reasonably proposes that this choice requires distinguishing two dozen types, and also details statistical procedures capable of generating the distinct clusters of group properties (Bowers and Hausser, 1977).

Second, despite the resounding triumph of "contingency theory" and its compatibilities with configurational analysis, few have become contingency-ists in more than verbal terms. This elemental point gets support from Miller and Friesen's (1984) effort to find taxonomies in the literature. They report: "One of the most surprising results of this survey is the very small number of taxonomies that have been generated" (1984, 35).

This shortfall was critical, and it remains so.

Third, configurational analysis will depart in significant particulars from typical research, as does the work by Miller and Friesen (1984, 18–19 and 34–36). These features differentiate it:

- Deals with many qualities simultaneously to build "a detailed, holistic, integrated image," whereas most existing research is bivariate.
- Analyzes data so as to seek natural clusters of attributes, without the common assumptions that relationships are linear and causation unidirectional.
- Differences—for example, in time, process, situation—are taken into account whenever possible, while most available research fixates on a single point in time.
- Uses anecdotal data to test the applicability of the configurations, whereas existing work typically distances the researcher from the context, as in the reliance on questionnaires.

Similar work has been done by Owens (1971) with the properties of individuals, which has isolated nearly two dozen Bio-Data Subgroups for classifying persons (e.g., Stokes, Mumford, and Owens, 1994).

Fourth, two varieties of configurational analysis may be distinguished, at least. The work by Miller and Friesen (1984) illustrates the reliance on statistical procedures that permit reliable judgments about the number of separate domains
necessary to encompass both differences and similarities in some range of indicators. These researchers define a panel of ten archetypes—some characteristic of successful firms and others of the unsuccessful. The latter include the archetypes Impulsive and Stagnant, and the former include the archetypes Dominant and Niche Innovator. In addition, Miller and Friesen suggest certain transitional patterns sensitive to environmental changes, identifying “likely destinations” (1984, pp. 133–50) into which specific archetypes may transition (Miller and Friesen, 1977, 1980, 1982).

A second variety of configurational analysis may be labeled “clinical” and is illustrated by the five “neurotic organizational styles” proposed by Kets de Vries and Miller (1984, esp. pp. 24–25). They key on the personal characteristics of executives, and particularly on the “shared fantasies” that may develop in a collective exposed to the operating biases and life experiences of particular executives. Their observations suggest five neurotic styles: Paranoid, Compulsive, Dramatic, Depressive, and Schizoid.

These and all other examples of configurational analysis share two significant commonalities. Both have to stand up to empirical test—to determine whether they isolate consistent patterns of covariants that permit successful predictions. As with the statistically derived typology, moreover, the neurotic styles are intended to isolate consistent patterns of organizations. Not surprisingly, since the two books most central to the decision above share a coauthor, the neurotic styles are tentatively associated with several of the Archetypes (e.g., Kets de Vries and Miller, 1984, pp. 214–215). Specific interventions would be suitable for different combinations, in general.

**Toward a Manhattan Project in Organization Analysis and Praxis**

These days of cutbacks in research grants may seem to be a curious time to reemphasize the following proposal, but some things require periodic even if brief emphasis lest we forget their salience. The need for a kind of Manhattan Project for the organizational arts and sciences grows, with the prime alternative being a sense of analytic circularity and even conceptual stuckness. Young readers may need reminding about the original Manhattan Project, whose then-unprecedented scale permitted solving problems in atomic theory that led to a quantum leap—to risk a play on words—in what was known about physics and in what could be done with that knowledge to create both boon and bane for humanity.

The argument is direct, even if the conclusion is intimidating. As noted, part of our legacy of conflicting findings derives from the lack of configurational analysis. Given the probable inclusion in any research area of units of analysis with different characters and characterizations, the research findings is
The Future of OD 637

not surprising. Rather, it is inevitable, except by great good fortune. No obvious reasons encourage one to anticipate that organizations configured as Impulsive will share a range of covariants with those labeled Stagnant. Similarly, “neurotic organizations” might well share some deficiencies in processes and products, but Dramatic and Depressive variants no doubt would differ across many realms of phenomena.

But no narrow success for configurational analysis can be envisioned, given research tools and traditions as they tend to be. As Miller and Friesen emphasize, internal elegance and parsimony will suffice for neither analysis nor, especially, for applications. Instead, any configurations will be useful to the degree that they enhance our knowledge about external relationships and covariants—that is, to the degree that taxonomies covary consistently with clusters of relationships in nature.

Establishing these covariants for configurational analysis no doubt will require a change in the scale and character of research. Patently, such work will need to cope with the welter of operational definitions common in research as well as with similar conceptual targets. Fortunately, considerable conceptual and operational clarification has been achieved in some areas. For example, “cohesiveness” seems in tolerably good shape and has been for some time; and substantial closure has occurred concerning the relative usefulness of the multiple ways of measuring the facets of “satisfaction.” Generally, however, multiple operations are used to measure “the same” conceptual domains, and those operations probably will generate differences in results. Then confusion sets in. For rare exceptions to neglect of operational definitions, see March, 1956, Christie and Geis, 1970, and Altemeyer, 1981.

No wonder, then, that the research findings in an area are typically inconsistent. Again, from the present perspective, we should expect inconsistency and even contradiction between results. Techniques such as meta-analysis may help some (Hunter, Schmidt, and Jackson, 1982), but no technique can replace the laborious, ponderous, and (yes!) unglamorous work of what may be called “comparative operational analysis.”

Now, what does all this have to do with a Manhattan Project? Quite a bit! An introduction to some economics of the research appropriate for configurational analysis has a large price tag. Consider that Bowers and Hausser (1977) isolate twenty-three types of work groups and use them to assess the efficacy of targeting five interventions. Even if nature were so cooperative as to generate equal distributions of the types, the design matrix already has 23-by-5, or 115 cells, and even five entries per cell would require 575 work groups under the improbably favorable condition posited here. Hence the need for an equivalent of the Manhattan Project.

Determined relational analysis sharply escalates the scale of the required optimum research enterprise, in short. Appropriate confidence in results would
be heightened if a range of variables were included in a design and simultaneously applied to one known population. These variables populate at least the five panels shown in Figure 26.3 and their many cross-checks.

Only very large scale designs suffice, patently. And the worse is yet to be written. Huge increases in scale also will result from two elemental data: that multiple operational definitions have been in common use for many of the relevant variables and that at present no substantial evidence permits a grounded choice between most alternative operational definitions.

Note a final point. The unlikelihood of the magnum approach—and, perhaps, the basic stuckness of organization analysis—has encouraged an emphasis on \( N = 1 \) designs. These may be helpful, of course, but emphasis now on \( N = 1 \) designs basically puts matters the wrong way around. The contribution of \( N = 1 \) designs would be enhanced if a substantial set of conceptual-operational pairs were in hand and thus constituted a common vocabulary for research. So, \( N = 1 \) designs will be useful at later stages of development, but today they probably do more to perpetuate than to solve the now-pressing problems of organizational designs. In brief, \( N = 1 \) designs seem more able to profit from a Manhattan Project in the social sciences than to provide a foundation for it.

**ALTERNATIVE TO RESPONDING PESSIMISTICALLY TO THE POTENTIALS OF PLANNED CHANGE**

This second edition, and Chapter 26 to this point, seek to be upbeat about the future, but pessimism is a possible response. Hence, the attention here to two probable sources of pessimism that the progress required to reduce ironies can
be made. The foci are on trinitarian change as well as on visioning, as examples of basic shortfalls in the development of OD.

**Trinitarian Change as Roadblock versus Challenge**

As in the first edition, this update draws attention to trinitarian change, which has powerful implications for OD. See especially Chapter 24. The possible realities are that this concept may result in positive change or merely create a great obstacle. Here, we focus on the possible metaphysical pathos that may result from attention to trinitarian change—a set of attitudes that encourage pessimism and perhaps even despair. Some researchers will be stimulated by plural change to explicate its properties, even though living with the conceptual distinctions is awesome. So what if you can establish that more than one kind of change exists? And so what if you can validly and reliably measure when non-alpha change occurs? One loses the sense of innocence thereby, but wisdom is less assured than ever. In the short run, gamma change taketh more than it giveth in five senses.

The original statistical technology for determining gamma change requires a large N, which limits its flexibility. Most attention to gamma change focuses on aggregates of individuals, which neglects those many cases in which the individual is not the unit of analysis. All technologies for determining non-alpha change require major and debatable assumptions. Available research fixates on testing for non-alpha change between $T_1$ and $T_2$, and neglects the tangled issue of $T_1$ comparisons with $T_2$. For alpha change, $T_1$ can be considered 0, and $T_2$ could be + or -. For gamma, “direction” has no obvious meaning, however. The covariants of the new state indicated by gamma change will be poorly mapped until the properties of alternative states are understood sufficiently to develop a taxonomy of states.

The tripartite model has an awkward profile, then. It plainly corrodes old assumptions but does not yet qualify as constructive in providing new and firm guides for analysis. The stakes are raised in the game of research; once-comfortable rules for interpreting results are less applicable, but the new rules remain uncertain, if rules there will be.

For practitioners or organization interventionists, the pain is greater, if anything. Already badgered with complaints about their know-nothing status as being “not rigorous”—but firmly believing that they know when interventions “work”—their comfort level is unrelievedly buffeted by notions like gamma change, in a catch-22 sense. Interveners typically seek non-alpha change—wit-
ness the emphasis on “inducing a new culture at work.” So beta and especially gamma have an intrinsic interest.

However, even if non-alpha change can be said to occur, no specific way currently exists to accomplish two critical analyses. They involve determining in any single case whether that newly induced dimensional space has the intended character—whether the movement has been in the intended direction and, if so, how far. Often, the same judgment will have to be made in comparisons of two or more variables. In effect, common measures finesse this problem, as in assuming alpha-only change that gets charted along (for example) Likert’s (1967) systems of management—with System 4 typically being considered the ideal of intended change, progress toward which may be estimated by successive waves of self-reports.

So gamma change not only introduces complexity but also precludes relative certainty, or at least reduces convenience. Concepts like gamma change can induce pessimism or lack of interest as an unintended by-product of research that seeks to enrich knowledge.

Managing Pessimism re Types of Change

What to do, then? Three points provide some guidance. While not comprehensive, they provide some direction and even hope. First, some obvious prescriptions apply, although they do not help much in the short run. For example: better to be wary about non-alpha change than wrong about it, even given the problems.

Second, everyday experience contains examples of several kinds of change, and this may embolden us to accept the challenge of understanding plural change so as to use that knowledge to ease transitions. No other viable possibility may exist, to put the point in the boldest terms. For example, trinitarian change basically involves changes in contours of psychological space, and in developed forms this might improve the efficiency and the effectiveness of change. One can either follow the contours of a surface, for example, or jump between the folds in the behavioral space. The former is the long way around, and probably almost more costly in resources. In common, all leaps involve recognizing the contours of transitional stages, and then using that knowledge to consciously short-circuit the full developmental distance along the surface of the contour.

Third, energetic attention should be directed toward testing ways to bridge the developmental gap between present understanding and eventually satisfactory ways of describing particular beta and gamma changes and their properties. Two examples will be sketched here.

Computer Simulations of Kinds of Change

One such initiative at “jumping the developmental gap” has not yet proved useful. Efforts were made to simulate the three types of change on a computer, with the
eventual purpose of experimenting with the simulations and, hence, of economically learning about the properties of different kinds of change. Analogous to thermodynamic systems, for example, “state variables” might be isolated and continuities or discontinuities in curves expressing them might be developed and then manipulated to learn about the properties of the three types of change. Burnout is one major candidate as a “state variable.”

Alpha change constitutes no problems, but about gamma change we learned at great cost only what we already knew: that gamma change means radical redefinition of a structure. We failed to think in “big bang” terms. The initial factorial structure proves very resistant to our incremental operations, which involve randomized applications of various decision rules to individual scores: no change, increases (or decreases) of one or two standard deviations, and so on. Several thousand computer runs proved incapable of generating change in the original structure profound enough to be considered gamma change, as estimated by Ahmavaara’s technique (1954). The apparent moral of the story: once a structure is established, it is hard to destroy by selective changes in scores. Think of it as an army at war. Even if you destroy the first line of combatants, the succeeding lines also have much the same properties.

Burnout as a Surrogate for Gamma Change

A search for surrogate measures of gamma change proves more fruitful. Promisingly, accumulating evidence suggests that the phase approach to psychological burnout may help tap differences in state as well as in degree.

Rationale for Encompassing Trinitarian Change

Competence in dealing with plural change is multiply consequential. To illustrate, one fashionable interpretation of survey findings describes differences in political opinions on surveys over a two-year-period as “non-attitudes,” or cases in which expressing any opinion—however specifically or tentatively held—is said to be preferable to being classified as a person holding no opinion (Converse, 1964, 1970). Researchers propose that test-retest comparisons demonstrate that “even relatively central and permanent issues of American political life have no meaning for the voters, that the issues involved do not form a part of the voters’ conceptual framework. Hence, for a large part of the population, the [survey] answers are essentially random. No real attitudes exist” (Aachen, 1975, p. 1219, my emphasis).

Gamma constitutes a rival hypothesis to non-attitudes, and the implications of the two explanations differ radically in philosophical and practical senses. Basically, gamma change implies data-processing capacities in individuals that support representative ideals. In contrast, non-attitudes imply a very different view of the electorate that can corrode democratic ideals. Aachen observes:
Criticisms... that voters’ preferences on public policy matters are unsophisticated or poorly organized, or that they are without influence in the voting decision [are serious but] they stop short of claiming that voters have no policy views whatever, [as does the non-attitude approach].

The last charge... is far more disturbing than the others. For it is one thing to argue that voters have difficulty connecting their preferences to particular candidates and particular schools of thought, quite another to claim that the preferences are absent from the beginning. (1975, pp. 1921–1928).

What can conceptual distinctions about change add to such debates over non-attitudes? Basically, all parties to the debate allow that three alternative explanations can account for low correlations between successive administrations of the same attitudinal items over a two-year interval. According to Stephens (1976, p. 1224), all observers distinguish:

“Real change,” or a measurable variation in some definite metric
“Measurement error,” or a disturbance imposed on the measurement process by imperfections in the measuring instrument
“Non-attitudes,” or reports by a respondent “who ‘blurts something out’ at random [to avoid] the embarrassment of admitting to no opinion”

The common view of real change seems myopic, however. Previous discussion suggests that three kinds of real change can be distinguished, and two of those varieties reject nonattitudes when explaining low correlations between administrations of the same attitudinal items at several points in time. A skeletal argument follows (Golembiewski, 1978).

The non-attitudes view acknowledges only alpha change, which occurs along relatively stable dimensions of reality defined in terms of relatively discrete and constant intervals. Alpha changes may or may not be random, as established by some test of statistical significance; they may be very large, or very small, or anywhere in between. If alpha-only occurred in the data used to support the non-attitudes view, Converse’s interpretation seems a reasonable possibility.

Two kinds of evidence encourage Converse (1970) to conclude that in effect, alpha-only occurs. He observes low correlations on several items at more than one point in time, and he finds low correlations between different but ostensibly-related items at the same time, which encourages him to reject the several kinds of “real change” as an explanation. Converse may be correct; but he focuses only on differences of magnitude in responses and does not test for any differences in pattern or structure, which alone can establish alpha-only. For him, “real change” seems to mean alpha. Of course, measurement error also could “explain” the two classes of events Converse observes.

The argument can be extended to the second type of change, if with greater
The Future of OD

643

effort. Beta changes involve the recalibration by respondents, between measurements, of some portion of the intervals used to measure a relatively stable conceptual domain. Notice a subtle point. Beta change is not measurement error, such as low test-retest reliability. But without explicit rejection of non-alpha change, beta change could be interpreted as measurement error. That is, the notion of measurement error assumes alpha-only, and in that sense it is context bound.

Occurrence of beta change in Converse’s data requires one condition. For example, low T1 versus T2 correlations are compatible with beta change, given its definition as involving recalibration of some portion of the T2 measurement intervals. Hence, non-attitudes could have a competitor for explaining the low correlations, and a two-year interval provides sufficient time for the development of the evaluational differences implied by beta change.

If beta change occurs in Converse’s data, non-attitudes provide a faulty and even seriously misleading interpretation. Beta change does not require the assumption that voters’ preferences are unsophisticated or poorly organized. Quite the opposite might be the case, in fact. The unsophistication in this case is that of the analyst favoring non-attitudes as the explanation.

The concern about the non-attitudes hypothesis is far greater with reference to the third type of change. Gamma involves a quantum shift in dimensions of reality—the redefinition of both the relevant psychological space as well as the intervals used at T2. In sum, gamma change refers to a shift from one state to another and also implies the serious unreliability of T1 intervals at T2.

The irony only requires stating. If gamma change occurs in Converse’s data, the nonattitudes interpretation is profoundly wrong. Rather than a lumpen striving to avoid embarrassment, in effect, the low correlations observed by Converse might indicate profound shifts among an electorate sophisticated in mental data-manipulation over the two-year period of observation. Such a shift implies complex and patterned decision-making processes, which the non-attitude interpretation rejects. In sum, gamma change explains the low correlations reported by Converse at T1 and T2 as well as non-attitudes, while simultaneously avoiding a profound metaphysical pathos for liberal democracy that comes as a leg-product of the nonattitude hypothesis.

That is, the choice between the two interpretations is consequential. The non-attitudes interpretation implies a view of the electorate that gives little comfort to those holding representative ideals, and it may encourage or even legitimate manipulation and authoritarian practices among political elites. “The people” in this view have the characteristics of a mass waiting to march to someone else’s purposes—eager enough to please an interviewer, but unreflectively so.

Empirical research relevant to making this crucial technical and philosophical choice has barely begun. But gamma effects have been found in survey data earlier interpreted as merely conflicting findings (Chittick, Billingsley, and Travis, 1988).
In sum, success in coping with trinitarian change will have both OD-related and broader effects. The former will result in more reliable estimates of success rates of OD applications, and the latter effects can be far broader as is illustrated above.

GUIDES FOR PLANNED CHANGE VIA OD

OD efforts are not equifinal: not every approach will get you to the same place, let alone to different intended destinations. Oppositely, OD places a premium on clear guides for development. Failure to respect such guides—either by intervenors or in their rejection by those at a particular worksite—can lead to failure, as two summaries illustrate.

Visions of Basic Processes or Dynamics

OD envisions quite specific end-products—a value-loaded set of basic dynamics and processes, if you will. Three relevant areas are typically isolated:

A regenerative vs. degenerative model of interaction, as is illustrated in several places above—especially the Introduction;

Structural features that permit enhanced ranges of autonomy and freedom at work—e.g., as in various non-bureaucratic approaches to work such as those illustrated in Chapters 5, 13, and 14—participative forms of supervision, divisional structures, and so on; and

Policies and procedures that permit greater responsible freedom at particular worksites, including but not restricted to enriched concepts of the workweek, the workplace, and new approaches to adverse personnel actions such as those illustrated in Chapters 21 through 23.

Templates for Designing Basic Dynamics and Processes

OD also can be thought of as being value-loaded or directionally guided in its development of these three basic dynamics and processes sketched above. Commonly, three such templates are distinguished. These include (Myers, 2000):

A normative-reeducative overlay specific to OD—often referred to as the OD ethic (see, especially, Chapters 7 and 8);

One or more contextual overlays, which are also value-loaded; that is, the institutionalization of larger systems we commonly call organizations; the broad normative systems within which organizations exist, one of which is described earlier as the Confucian ethic; and the normative contexts that develop even in small work units;

An Empirical-rational overlay, which uses scientific knowledge about how to approach conditions at work that are considered desired and desirable.
Pessimistic Potentials in Inadequate Fits

Although Chapters 1 through 3 attribute substantial success rates to evaluations of OD applications, there clearly exists substantial potential for inadequate fit between the seven arenas distinguished immediately above. Without any pretense of comprehensiveness, such inadequate fits can possibly come to characterize OD applications in five basic ways.

First, OD applications might not sufficiently integrate changes in interaction, policy or procedures, and structures. Indeed, in practice, this may be the dominant lack of fit. The power and convenience of interaction-centered designs, in short, can discourage reinforcements such as those illustrated at many points below. See especially Chapter 14 for one effort to integrate these OD domains. Lacking such reinforcement, initial progress may take on punishing characteristics—as when greater interactive freedom in principle gets an employee reprimanded by superior levels of authority or by traditional policies and structures that are unaffected by the changes experienced by employees.

Second, the OD ethic may be too differentiated from local contexts of application. For example, some people have made such an argument about OD applications in settings characterized by (for example) the Confucian ethic and its associated values, attitudes, and behaviors. We have little detailed evidence about such a possible dysjoint, and several earlier chapters provide real if not yet convincing progress concerning goodness of fit.

Third, local worksites may be value-loaded in ways that are antagonistic to OD values. Progress toward OD values hence might generate local opposition, even though many or most organization members prefer OD values. Here, possibly, small but powerful cliques might subvert lasting movement toward OD values. Fade-out of initial effects also might well occur. Such effects seem relatively common; or, at least, well-known exemplars exist (e.g., Golembiewski and Kiepper, 1988).

Typically, such effects reflect the persistence of power-coercive strategies

at specific worksites, with the general expectation being that experience in this regard will develop cumulatively—from efforts to involve people at specific worksites in assessing values, issues to be dealt with, and ways-and-means of dealing with them, and then to broadening models of theory and experience, on the general model illustrated in Chapters 12 and 15; and

A power-coercive overlay, with the two templates above progressively helping to isolate areas that are considered consensually legitimate by both management and employees: basically, power-coercion dominates in traditional organizations and OD seeks shifting balances of the four templates in different situations, as many earlier chapters illustrate.
that are common in many organizations. OD change is seldom unanimous, but even substantial majorities in favor of specific changes might be frustrated by determined and well-placed minorities. Basically, this is an operating bias of OD applications, which typically rely on reeducative strategies even when resisting remnants are quite small (e.g., Golembiewski and Miller, 2000).

Fourth, broad socioeconomic systems encompassed by totalitarian ideologies might inhibit the diffusion of the OD ethic, even in the presence of very substantial majorities who prefer the latter normative framework. “We just can’t do that here, even though we would like to.” That summarizes the experience with OD values and approaches in several small populations from totalitarian nation-states (Golembiewski, 2002).

Fifth and finally for present purposes of illustrating how OD applications might fail, the conveniences of long-standing relationships might be considered sufficiently preferable to adaptations better-suited to changing conditions. Something like this occurred in the late 1960s Project ACORD, or Action for Organization Development (Golembiewski and Kiepper, 1988). A cohort of resignations apparently had to occur before substantial change was possible, and those personnel actions took several years.

REFERENCES
The Future of OD


Author Index

Achen, C. R., 641, 646
Adams, 405
Adams, J., 401
Adams, J. D., 583, 588
Agarwal, S., 75
Ahmavaara, Y., 315, 316, 318, 576, 577, 588, 641
Akerstedt, T., 530, 535, 544
Aldinger, R. T., 506, 508
Al-Ebedah, N., 408, 507, 508
Alon, M., 77
Altemeyer, R. A., 637, 646
Anderson, C. M., 430, 447
Anderson, W., 544
Andrews, F. M., 472
Ansoff, H. I., 283, 286, 296, 318
Antoszkiewicz, J. D., 75
Arbose, J., 61, 75
Argyris, C., 9, 201, 202, 205, 208, 213, 214, 220, 228, 277, 348, 349, 350, 364, 365, 375, 379, 400, 401, 402, 624, 626, 646
Armenakis, A., 574, 582, 583, 585, 588
Aronson, E., 430, 448
Ashby, W. R., 572, 588
Ashkenas, R. N., 144, 152, 193, 194, 195, 274, 277, 383, 392, 402, 631, 646
Arthur, W. Jr., 58, 61, 75
Aucoin, P., 136, 162
Austin, J. T., 46
Babcock, R. D., 63, 75
Bacharach, S. B., 44, 153, 229, 647
Bacharach, S. J., 77
Bacon, P. C., 3, 46
Bailyn, L., 534, 544
Bagley, 535, 536
Bamforth, K. W., 370, 380
Barancle, T., 330, 334, 335, 337
Barnard, C., 135
Barra, R., 277
Barsky, J., 506, 508
Bartunek, J. M., 388, 405
Barzelay, M., 139, 153
Bass, B., 83, 93, 469, 472
Basu, K., 58, 61, 76
Bates, J. W., 170, 180
Baxter, D. R., 507, 509
Beckhard, R., 278, 283, 294, 295, 318
<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beer, M.</td>
<td>42, 183, 194, 195, 257, 278, 592, 602</td>
</tr>
<tr>
<td>Beinstock, P.</td>
<td>45</td>
</tr>
<tr>
<td>Bell, C. B.</td>
<td>105, 129</td>
</tr>
<tr>
<td>Bell, C. H., Jr.</td>
<td>3, 16, 43, 276, 277, 278, 450, 472</td>
</tr>
<tr>
<td>Benne, K. D.</td>
<td>192, 195, 281, 318</td>
</tr>
<tr>
<td>Bennis, W. G.</td>
<td>186, 195, 196, 232, 234, 253, 276, 278, 392, 405, 573, 588</td>
</tr>
<tr>
<td>Bent, K. H.</td>
<td>453, 473</td>
</tr>
<tr>
<td>Bereiter, C.</td>
<td>566, 588</td>
</tr>
<tr>
<td>Berg, D. N.</td>
<td>5, 46, 79</td>
</tr>
<tr>
<td>Berg, P. O.</td>
<td>3, 12, 13, 46, 207, 209</td>
</tr>
<tr>
<td>Berlew, D. E.</td>
<td>402</td>
</tr>
<tr>
<td>Berley, S. F.</td>
<td>544</td>
</tr>
<tr>
<td>Bianchini, J.</td>
<td>587, 589</td>
</tr>
<tr>
<td>Billingsley, K.</td>
<td>316, 319, 450, 471, 472, 529, 565, 576, 578, 583, 585, 587, 589, 643</td>
</tr>
<tr>
<td>Bion, W. R.</td>
<td>626, 646</td>
</tr>
<tr>
<td>Blake, R. R.</td>
<td>261, 278</td>
</tr>
<tr>
<td>Blalock, H. M.</td>
<td>472</td>
</tr>
<tr>
<td>Blood, C. L.</td>
<td>418, 428</td>
</tr>
<tr>
<td>Blumberg, A.</td>
<td>7, 42, 97, 129, 195, 412, 425, 427, 547, 556, 563</td>
</tr>
<tr>
<td>Bohan, H. N.</td>
<td>529</td>
</tr>
<tr>
<td>Bolton, A. L.</td>
<td>275, 278</td>
</tr>
<tr>
<td>Boltvinik, M.</td>
<td>43, 76</td>
</tr>
<tr>
<td>Bond, 216</td>
<td></td>
</tr>
<tr>
<td>Bondy, J. S.</td>
<td>45</td>
</tr>
<tr>
<td>Borton, T.</td>
<td>402</td>
</tr>
<tr>
<td>Boss, R. W.</td>
<td>42, 43, 76, 218, 228, 426, 447, 563</td>
</tr>
<tr>
<td>Boston, J.</td>
<td>136, 138, 153</td>
</tr>
<tr>
<td>Bourgeois, L. J.</td>
<td>43, 76</td>
</tr>
<tr>
<td>Bowers, D.</td>
<td>5, 43, 46, 497, 648</td>
</tr>
<tr>
<td>Boynton, K. S.</td>
<td>493, 494</td>
</tr>
<tr>
<td>Bozeman, B.</td>
<td>402</td>
</tr>
<tr>
<td>Bradbury, M. D.</td>
<td>44, 506, 508</td>
</tr>
<tr>
<td>Bradford, L. P.</td>
<td>192, 195, 281, 318, 600</td>
</tr>
<tr>
<td>Bramucci, R.</td>
<td>323, 340</td>
</tr>
<tr>
<td>Brown, F. G.</td>
<td>402</td>
</tr>
<tr>
<td>Brown, J. C.</td>
<td>406</td>
</tr>
<tr>
<td>Brown, L. D.</td>
<td>228, 614, 646</td>
</tr>
<tr>
<td>Browne, P. J.</td>
<td>276, 278, 470, 472</td>
</tr>
<tr>
<td>Bruce, R.</td>
<td>159, 180</td>
</tr>
<tr>
<td>Bruce, R. A.</td>
<td>546, 562, 564</td>
</tr>
<tr>
<td>Bruce, R. R.</td>
<td>402</td>
</tr>
<tr>
<td>Bryson, J. M.</td>
<td>139, 140, 153, 283, 285, 318</td>
</tr>
<tr>
<td>Buller, P. F.</td>
<td>290, 291, 318</td>
</tr>
<tr>
<td>Bullock, R. J.</td>
<td>43, 83, 87, 92, 93, 276, 278</td>
</tr>
<tr>
<td>Bunker, B. B.</td>
<td>281, 318</td>
</tr>
<tr>
<td>Burden, D. W.</td>
<td>276, 278</td>
</tr>
<tr>
<td>Burlington, D.</td>
<td>603, 623, 647</td>
</tr>
<tr>
<td>Burke, R.</td>
<td>488, 493</td>
</tr>
<tr>
<td>Burke, R. J.</td>
<td>273, 280, 500</td>
</tr>
<tr>
<td>Burke, W. W.</td>
<td>4, 5, 47, 65, 76, 254, 334, 345, 381, 402, 427, 494, 583, 588</td>
</tr>
<tr>
<td>Bush, G.</td>
<td>322</td>
</tr>
<tr>
<td>Bushe, G. R.</td>
<td>599, 602, 624, 627, 646</td>
</tr>
<tr>
<td>Buss, A. R.</td>
<td>585, 588</td>
</tr>
<tr>
<td>Bussom, R. S.</td>
<td>64, 76</td>
</tr>
<tr>
<td>Caffrey, J.</td>
<td>589</td>
</tr>
<tr>
<td>Cahoon, A. R.</td>
<td>506, 509</td>
</tr>
<tr>
<td>Calista, D. J.</td>
<td>402, 405</td>
</tr>
<tr>
<td>Cameron, K. S.</td>
<td>43</td>
</tr>
<tr>
<td>Campbell, D. T.</td>
<td>43, 85, 93, 586, 589</td>
</tr>
<tr>
<td>Campbell, K. S.</td>
<td>43</td>
</tr>
<tr>
<td>Campbell, T.</td>
<td>249, 253</td>
</tr>
<tr>
<td>Caplan, R. D.</td>
<td>493</td>
</tr>
<tr>
<td>Carew, D. K.</td>
<td>147, 153</td>
</tr>
<tr>
<td>Carlisle, A. E.</td>
<td>415, 426</td>
</tr>
<tr>
<td>Carrigan, S. B.</td>
<td>44, 129, 139, 153, 194, 196, 293, 299, 318, 385, 403, 404, 447, 473, 547, 556, 563, 569, 589</td>
</tr>
<tr>
<td>Carter, D.</td>
<td>447, 477, 494</td>
</tr>
<tr>
<td>Carter, J.</td>
<td>388</td>
</tr>
<tr>
<td>Carter, S. I.</td>
<td>137, 138, 147, 153</td>
</tr>
</tbody>
</table>
Chandler, A. D., Jr., 283, 288, 295, 318, 370, 379, 439, 440, 447
Chandler, K., 127, 129
Chandler, R. C., 494, 647
Chandrasekaran, K. A., 76
Chang, I., 625, 646
Charith, M., 138, 153
Chase, P., 278
Chavez, L., 529
Charih, M., 138, 153
Cherniss, C., 493
Cherns, A. B., 278
Child, J., 633, 646
Ching, J., 214, 216, 218, 228
Chisholm, R. T., 402, 646
Chittick, W. O., 583, 589, 643
Cho, Y-H., 76
Christi, R., 637, 646
Chun, S-K., 76
Chung, K. H., 93
Church, A. H., 594, 602
Claus, R. E., 76
Clinton, W. J., 3
Cloonan, J., 530
Coch, L., 108, 128
Coetzer, 599, 602
Coghlan, D., 507, 508
Cohen, A. R., 402
Coker, F. W., 146, 153
Collion, M-H., 76
Confucius, 66, 78, 183, 185, 200, 208, 209, 211, 212, 213, 218, 220, 219, 221, 222, 223, 226, 227, 228, 607, 644, 645
Connor, P. E., 273, 277, 278
Conrad, C., 43
Converse, P. E., 570, 589, 641, 642, 643
Cooke, R. A., 406
Cooper, C. L., 504, 508, 647
Cooperrider, D. L., 188, 195, 383, 385, 394, 402, 411, 592, 595, 596, 598, 600, 602, 626, 646, 647
Cotton, C. C., 276, 278, 470, 472
Culbert, S. A., 188, 195
Cummings, T. G., 45, 257, 278, 370, 379
Crawford, K. S., 276, 278
Crockett, W. J., 43, 384, 399, 402
Cronbach, L. J., 71, 479, 585, 589
Crowne, D. P., 24, 248, 254
Daft, R. L., 334, 344
Dahl, R. A., 146, 153, 382
Daly, R., 299, 402
Dalton, D. R., 529
Danials, A., 153
Das, P., 58, 61, 76
Davies, A., 402
Davis, E., 602
Davis, K., 623, 647
Davis, L. E., 278
Davis, P., 355, 376, 379
Davis, P. S., 285, 318
Davis, S., 566, 590
Davis, S. M., 139, 147, 153, 278
De, N., 62, 63, 64, 76
Deal, T. E., 292, 318
Deckard, G. J., 355, 360, 376, 380, 488, 492, 493
Deckman, D., 515, 529
Dempster, M., 511, 508
Delbecq, A. L., 155, 185
Deming, F., 153
Deming, W. E., 131, 153, 327, 344
De Meuse, K. P., 278, 408, 422, 426
Denhardt, R. B., 162, 180
Desky, J., 538, 539, 544
Dess, G. G., 285, 318
Dessca, G., 488, 493, 500
Deutsch, M., 647
Devine, D. J., 273, 274, 278, 544
Devine, D. R., 529
DiBella, A. J., 76
Dietterich, P. M., 318
Dole, E., 333
Dollinger, M. J., 221, 228
Doverspike, D., 58, 61, 75
Driscoll, J. W., 278
Drucker, P., 135
Dugan, H. S., 61, 76
Dugan, S., 612, 648
Duncan, 240
Duncan, J., 51, 78,
Author Index

Dunn, W. N., 3, 43, 65, 76, 258, 260, 269, 271, 276, 278
Duplicea, I. A., 506, 508
DuToit, L., 61, 76
Dyer, G., 103, 128
Dyer, W., 402, 407, 408, 409, 411, 414, 426, 488, 493
Eddy, W., 44
Edwards, J. E., 46
Eisenstat, R., 592, 602
Elsaid, H., 64, 76
Elloy, D., 530
Emery, F., 144, 153
Emery, M., 383, 402
Engdahl, R., 77
Esman, J. W., 402
Evered, R. D., 357, 380
Everett, A., 507, 508
Ezra, M., 508, 511, 515, 529
Faerstein, P. H., 45
Faiferman, V., 506, 508
Fairweather, N. B., 544
Farber, V. A., 379
Farazmand, A., 509
Ferguson, S. F., 598, 599, 603
Festinger, L., 231, 254, 415, 417, 428
Fiadzo, E., 508
Filley, A., 414
Fink, J. J., 276, 278
Fisch, R., 590
Fitzgerald, S., 595, 596, 602
Fleischman, E. A., 43, 143, 153
Fleming, L., 534, 537, 544
Flynn, W. R., 530
Fordyce, J. K., 412, 426
Foulkes, M. B., 495
Fox, R., 529
Frances, C., 257, 280
Frankel, M. S., 282, 318, 627, 647
Franklin, J. L., 76, 247, 253, 450, 472
French, J. R. P., Jr., 128, 493
French, W. F., 3, 16, 43, 276, 277, 278, 391, 450, 472
French, W. L., 105, 108, 129
Freud, A., 603, 623, 647
Freudenberg, H. J., 493
Friesen, P. H., 155, 617, 621, 634, 636, 637, 648, 649
Frucher, M. S., 261, 278
Furley, L., 585, 589
Gabris, G., 162, 180
Gadon, H., 402
Gagnache, J. M., 147, 153
Gardner, J. W., 95, 128
Gardner, N., 179, 180, 402
Garson, G. D., 134, 135, 153
Gavin, J. F., 276, 278
Geis, F. L., 637, 646
Geller, M., 282, 318, 627, 647
Gergen, J. K., 395, 402, 625, 647
Gergen, K., 189, 195
Gibb, J. R., 192, 195, 281, 318
Giblin, E. J., 5, 274, 279
Girard, M., 326, 345
Glen, R. H., 257, 278
Gladwell, J., 369
Gluckstern, N. B., 276, 279
Glueck, W. F., 529
Goldberg, D. P., 484, 494
Goldinger, C., 324, 326, 345
Goldner, F. H., 548, 549, 563
[Golembiewski]
276, 279, 281, 282, 283, 284, 285,
286, 287, 290, 291, 292, 295, 298,
299, 305, 306, 307, 309, 313, 315,
316, 318, 319, 322, 323, 324, 333,
336, 337, 339, 345, 347, 351, 355,
358, 359, 360, 364, 365, 369, 370,
376, 378, 379, 382, 383, 384,
385, 386, 387, 388, 389, 390, 391,
392, 393, 394, 395, 396, 397, 399,
400, 401, 402, 403, 404, 407, 408,
409, 411, 412, 414, 418, 420, 421,
422, 423, 425, 426, 427, 430, 431,
432, 433, 436, 439, 440, 444, 445,
447, 448, 450, 451, 470, 471, 472,
473, 477, 478, 479, 484, 485, 486,
487, 488, 489, 490, 492, 493, 494,
498, 499, 501, 500, 503, 505, 506,
507, 508, 509, 529, 532, 534, 539,
544, 545, 546, 547, 548, 549, 556,
559, 563, 565, 566, 567, 569, 574,
576, 578, 582, 583, 584, 585, 587,
589, 593, 595, 597, 598, 599, 603,
612, 616, 618, 620, 621, 625, 627,
629, 631, 632, 642, 645, 646, 647,
648
Gooden, V., 147, 154
Goodman, E. A., 68
Goodman, P. S., 254, 279, 529
Goodnow, F., 133
Goodsell, C., 45, 146, 154
Goodstein, L. D., 3, 4, 47, 293, 294,
334, 345
Gorbachev, 417
Gordon, G., 536, 544
Gore, A., 333
Gortner, H. F., 143, 150, 154,
Goto, K., 507, 508
Gouldner, A., 287, 473
Graf, L. A., 530,
Gray, M., 536, 544,
Gregor, K., 167, 168, 169, 171, 172,
173, 174, 178,
Guopei, G., 217, 229,
Gupta, A., 79,
Guzzo, R. A., 45,
Ha, M.-S., 218, 229
Haack, M. R., 430, 448
Hackman, J. R., 248, 254, 274, 418,
427
Haire, M., 633, 648
Halberstam, D., 427
Haldeman, J., 257, 280
Hall, D. T., 548, 549, 563
Halley, A. A., 388, 404, 631, 648
Halston, D. A., 530
Hammond, S. A., 594, 603
Hampden-Turner, C. W., 179, 181, 573,
589
Hamilton, M., 589
Handy, S. L., 536, 544
Hanna, D. P., 139, 154
Hardiman, R., 147, 154
Harmon, M. M., 45, 141, 154, 281, 319,
384, 399, 404
Harrick, E. J., 530
Harris, C. W., 588, 589
Harris, R. T., 408, 422, 427
Harris, S. G., 546, 564
Harrison, F., 543, 544
Harrison, R., 232, 234, 248, 254, 414,
424, 427, 476, 493, 494, 498, 509
Hausser, D. L., 232, 247, 248, 253, 407,
422, 426, 617, 621, 635, 637, 646
Harvey, B. H., 530
Harvey, J. B., 45, 369, 409, 411, 413,
414, 415, 416, 417, 423, 424, 427,
597, 598, 600, 603
Hay Group, Inc., 348, 352, 356, 357,
361, 362, 379, 380
Head, T. C., 51, 77, 599, 603
Heider, F., 233, 252, 254
Heimovics, R., 402
Hendrix, W. H., 261, 279, 348
Hennestad, B. W., 379
Hensman, R., 504, 508
Herbst, P., 370, 379
Herzberg, F., 375, 379, 466, 473,
624, 648
Herbst, P., 379
Herman, A., 334
Herzberg, F., 379
<table>
<thead>
<tr>
<th>Author</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hickson, D. J.</td>
<td>649</td>
</tr>
<tr>
<td>Hildredth, W. B.</td>
<td>153, 154, 229, 318, 345</td>
</tr>
<tr>
<td>Hinnings, C. R.</td>
<td>649</td>
</tr>
<tr>
<td>Hjern, B.</td>
<td>631, 648</td>
</tr>
<tr>
<td>Hochschild, A.</td>
<td>530</td>
</tr>
<tr>
<td>Hofstede, G.</td>
<td>45, 58, 61, 71, 77, 216, 399</td>
</tr>
<tr>
<td>Hornstein, H.</td>
<td>254, 427, 494</td>
</tr>
<tr>
<td>Howard, G. S.</td>
<td>317, 319, 493, 494</td>
</tr>
<tr>
<td>Huang, T-S.</td>
<td>507, 509</td>
</tr>
<tr>
<td>Hubert, G.</td>
<td>506, 508</td>
</tr>
<tr>
<td>Hudson, M.</td>
<td>536, 544</td>
</tr>
<tr>
<td>Hulin, C. L.</td>
<td>457, 473</td>
</tr>
<tr>
<td>Hull, D. G.</td>
<td>453, 473</td>
</tr>
<tr>
<td>Hunka, S.</td>
<td>587, 589</td>
</tr>
<tr>
<td>Hunter, J. E.</td>
<td>637, 648</td>
</tr>
<tr>
<td>Hurbert, G.</td>
<td>508</td>
</tr>
<tr>
<td>Huse, E. F.</td>
<td>45, 276, 279</td>
</tr>
<tr>
<td>Hyde, A. C.</td>
<td>135, 155</td>
</tr>
<tr>
<td>Hymowitz, R.</td>
<td>538</td>
</tr>
<tr>
<td>Ignatov, M.</td>
<td>77</td>
</tr>
<tr>
<td>Insel, P. M.</td>
<td>433, 448</td>
</tr>
<tr>
<td>Isabella, L. A.</td>
<td>548, 549, 563</td>
</tr>
<tr>
<td>Itskowitz, R.</td>
<td>77</td>
</tr>
<tr>
<td>Iyengar, S.</td>
<td>570, 589</td>
</tr>
<tr>
<td>Jacobs, L.</td>
<td>217, 229</td>
</tr>
<tr>
<td>Jackson, B.</td>
<td>147, 153</td>
</tr>
<tr>
<td>Jackson, C. N.</td>
<td>506, 509</td>
</tr>
<tr>
<td>Jackson, G. B.</td>
<td>637, 648</td>
</tr>
<tr>
<td>Jackson, S. E.</td>
<td>355, 379, 418, 427, 431, 448, 477, 485, 495, 498, 509,</td>
</tr>
<tr>
<td>Jaeger, A. M.</td>
<td>45</td>
</tr>
<tr>
<td>Jain, U.</td>
<td>75, 76, 77, 78, 79, 80</td>
</tr>
<tr>
<td>Janis, L.</td>
<td>409, 414, 427</td>
</tr>
<tr>
<td>Jantz, T.</td>
<td>612, 648</td>
</tr>
<tr>
<td>Jaworski, W.</td>
<td>75</td>
</tr>
<tr>
<td>Jelinek, M.</td>
<td>179, 181, 283, 292, 294, 295, 309, 319</td>
</tr>
<tr>
<td>Jenkins, J. G.</td>
<td>453, 473</td>
</tr>
<tr>
<td>Jette, R. D.</td>
<td>45</td>
</tr>
<tr>
<td>Johnson, B.</td>
<td>153</td>
</tr>
<tr>
<td>Johnson, K. R.</td>
<td>51, 52, 54, 61, 78</td>
</tr>
<tr>
<td>Johnson, L. B.</td>
<td>37, 141</td>
</tr>
<tr>
<td>Joice, W.</td>
<td>534, 535, 537, 538, 543, 544</td>
</tr>
<tr>
<td>Jones, D. A.</td>
<td>599, 603</td>
</tr>
<tr>
<td>Jorom, M. G.</td>
<td>77</td>
</tr>
<tr>
<td>Jormon, M. G.</td>
<td>78, 79, 80</td>
</tr>
<tr>
<td>Joshi, S.</td>
<td>78</td>
</tr>
<tr>
<td>Kagno, M.</td>
<td>191, 196, 279, 488, 494, 529, 574, 589</td>
</tr>
<tr>
<td>Kahn, R. L.</td>
<td>418, 419, 427, 457, 473</td>
</tr>
<tr>
<td>Kaiser, H. F.</td>
<td>589</td>
</tr>
<tr>
<td>Kaiser, H. G.</td>
<td>576, 587, 589</td>
</tr>
<tr>
<td>Kalleberg, A. J.</td>
<td>635, 648</td>
</tr>
<tr>
<td>Kanter, D. L.</td>
<td>45</td>
</tr>
<tr>
<td>Kanter, R. L.</td>
<td>415, 427</td>
</tr>
<tr>
<td>Katzell, R. A.</td>
<td>45</td>
</tr>
<tr>
<td>Kaufman, H.</td>
<td>45, 51, 78, 406</td>
</tr>
<tr>
<td>Keller, 286, 289, 291, 293</td>
<td></td>
</tr>
<tr>
<td>Kendall, L. M.</td>
<td>481, 428, 457, 473</td>
</tr>
<tr>
<td>Kennedy, A. R.</td>
<td>292, 318</td>
</tr>
<tr>
<td>Kennedy, J. H.</td>
<td>409, 415</td>
</tr>
<tr>
<td>Kets de Vries, M. F. R.</td>
<td>625, 636, 648</td>
</tr>
<tr>
<td>Kettl, D. F.</td>
<td>274, 279</td>
</tr>
<tr>
<td>Khademian, A. M.</td>
<td>136, 154</td>
</tr>
<tr>
<td>Kiev, A.</td>
<td>552, 563</td>
</tr>
<tr>
<td>Kilpatrick, A. O.</td>
<td>430, 448</td>
</tr>
<tr>
<td>Kim, S-E.</td>
<td>211, 213, 217, 408, 631, 648</td>
</tr>
<tr>
<td>Kim, B-S.</td>
<td>444, 447, 487, 494</td>
</tr>
<tr>
<td>King, D. C.</td>
<td>232, 254</td>
</tr>
<tr>
<td>Kinsman, 538</td>
<td></td>
</tr>
<tr>
<td>Kirchenbaum, H.</td>
<td>189, 196</td>
</tr>
<tr>
<td>Kirkbride, P. S.</td>
<td>51, 78</td>
</tr>
</tbody>
</table>
Marin, B., 648
Marlow, D., 248, 254
Marrow, A., 631, 648
Marrow, A. D., 281, 294, 319, 348, 379, 497
Marrow, A. J., 5, 46, 157
March, J. G., 129, 134, 287, 305, 648
Marsden, P. V., 635, 648
Marsak, R. J., 78, 202, 209, 215, 229
Marsick, V. J., 400, 406
Martin, J., 136, 138, 153
Martin, L., 333, 327
Maslach, C., 355, 379, 418, 421, 422, 423, 427, 433, 448, 477, 495, 498, 499, 509
Maslow, A. H., 9, 277, 280
Marsden, P. V., 648
Marsick, V. J., 47
Mathur, S., 78, 78
Mott, M., 79
Mauzer, B., 379
Mausner, B., 466, 473, 624, 648
Maxwell, S. E., 317, 319, 493, 494
McCampbell, A. S., 530
McClelland, D. C., 405
McConkie, M. L., 43, 76, 280
McDonald, R. P., 587, 589
McGee, V. E., 570, 589
McGrath, M. R., 280
McGregor, D., 77, 451, 473
McKenzie, K., 196, 345, 381, 403
McClelland, D. C., 405
McMillen, M-C., 78
McPhail, S. M., 277, 278
McShane, K., 543, 544
Mead, M. R., 547, 556, 563
Meltzer, H., 278
Mesch, D. J., 529
Meyer, G., 3, 4, 47
Michlitsch, J. F., 530
Miles, M. B., 431, 448, 631
Miles, R. H., 647
Miller, C., 147, 155, 329, 404, 633, 646, 648
Miller, C. F., 65, 77, 140, 155, 322, 345, 386, 404
Miller, D., 617, 621, 626, 634, 635, 636, 637, 648, 649
Miller, G. J., 7, 46, 153, 154, 229, 318, 345
Mills, K. L., 534, 537, 544
Milton, J., 192
Mintzberg, H., 140, 147, 155, 621, 634, 649
Mirvis, P. H., 5, 45, 46, 79
Mitchell, E., 82, 93
Moch, M. K., 388, 405
Mohanty, J., 79
Moise, L. R., 494
Mokkartarian, P. L., 535, 536, 544
Molloy, E. S., 257, 278
Moon, M. J., 146, 155
Moos, R. H., 433, 434, 436, 448
More, D. M., 546, 563
Morley, D., 405
Mormer Solomon, C., 544
Morgan, J. M., 272
Morrison, P., 3, 6, 46, 82, 85, 87, 93, 258, 280
Mouton, J. S., 261, 278
Mosher, F., 519
Mossop, J., 46, 295, 388, 405
Moyns, R., 648
Muhsam, H. V., 590
Mumford, M. D., 617, 621, 635, 649
Murai, T., 507, 508
Murrell, K., 595, 602
Murrell, K. L., 79
Myers, M. G., 644, 649
Myers, T., 530
Myers, T. G., 489, 490, 495, 498, 505
Nadler, D. A., 450, 473
Nagel, S. S., 403, 405, 544
Naylor, M., 288
Nethery, K., 179, 181, 219
Neuman, G. A., 46
Newman, H. L., 598, 599, 603
Newman, L., 595, 602
Newhouse, A., 590
Nicholas, J. M., 6, 13, 14, 15 22, 35, 46, 65, 79, 82, 83, 93, 142, 155, 208, 209, 257, 280, 292, 382, 405
Nicholson, J. B., 150, 154
Nie, N. H., 453, 473
Nielsen, W. R., 3, 46
Nieva, V. F., 46, 273, 280
Nixon, R., 409, 415, 428
Nollen, S. D., 530
Nord, R., 278
Novelli, L., Jr., 530
Nguyen, H., 543, 544
Obrastsov, V. L., 79
Oh, S-G., 58, 79, 196, 216, 229
Oldham, G. R., 418, 427
Olson, M. H., 538, 540, 544
Osborne, D., 328, 345
Oshry, B. L., 498, 509
Ostroff, F., 345
O’Reilly, C. A. III, 419, 428, 457, 473
O’Toole, J., 392, 405
Ott, J. S., 135, 155
Ouchi, W., 221, 229
Overman, E. S., 134, 135, 153
Owens, W., 617, 621, 635, 649
Owens, W. A., 635, 649
Packard, R. W., 279
Paine, W. S., 426, 447
Pallot, J., 136, 138, 153
Pollitt, C., 139, 155
Perrin, C., 538, 539, 544
Perry, J. L., 134, 155
Peters, B. G., 155
Peters, T. J., 286, 292, 295, 334, 345
Peterson, E., 535, 544
Pfeiffer, G., 293, 294
Phelan-Carter, D., 427
Pines, A., 421, 427, 430, 448
Pinneau, S. R., 590
Pogoncheff, 291
Polak, N., 76
Pollitt, C., 139, 155
Pondy, L. R., 633, 649
Pope John Paul II, 184, 186, 187, 190, 191, 192, 196, 199, 204, 205, 220
Porras, J. L., 3, 6, 12, 13, 46, 207, 209, 257, 258, 276, 280, 408, 422, 427, 428
Porter, D. O., 631, 648
Posner, B. G., 140, 152, 155
Popp, M., 348, 379
Pozzi, 289, 292, 294
Prasad, B. R., 61, 76
Primps, S., 530
Proehl, C. W., Jr., 6, 7, 12, 44, 46, 65, 77, 82, 83, 93, 152, 154, 208, 209, 279, 280, 292, 299, 319, 382, 396, 428, 427, 529
Prytulak, L. S., 570, 590
Pugh, D. S., 649
Pun, A. S. L., 79
Purser, P. E., 383, 402
Quinn, R. E., 280
Quinn, R. P., 457, 473, 478, 486, 495
<table>
<thead>
<tr>
<th>Author Name</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rabin, J.</td>
<td>153, 154, 229, 318, 345, 648</td>
</tr>
<tr>
<td>Radin, B. A.</td>
<td>138, 141, 155</td>
</tr>
<tr>
<td>Rahim, M. A.</td>
<td>44, 77, 153, 196, 209, 279, 345, 347, 379, 381, 403, 447, 498, 508, 648</td>
</tr>
<tr>
<td>Rainey, G. J., Jr.</td>
<td>390, 405</td>
</tr>
<tr>
<td>Rainey, G. W., Jr.</td>
<td>632, 649</td>
</tr>
<tr>
<td>Rainey, H. G.</td>
<td>46, 134, 155, 390, 405, 632, 649</td>
</tr>
<tr>
<td>Rainey, M. A.</td>
<td>594, 603</td>
</tr>
<tr>
<td>Raju, N. S.</td>
<td>46</td>
</tr>
<tr>
<td>Rao, C.</td>
<td>590</td>
</tr>
<tr>
<td>Rao, L. B.</td>
<td>79</td>
</tr>
<tr>
<td>Rao, T. V.</td>
<td>75, 76, 78, 79</td>
</tr>
<tr>
<td>Raven, B.</td>
<td>409, 415, 428</td>
</tr>
<tr>
<td>Reagan, R.</td>
<td>273, 278, 529</td>
</tr>
<tr>
<td>Reddy, R. K.</td>
<td>139, 155</td>
</tr>
<tr>
<td>Reich, R. B.</td>
<td>328, 333, 334, 345, 389, 390, 405</td>
</tr>
<tr>
<td>Reichers, A. E.</td>
<td>46</td>
</tr>
<tr>
<td>Rice, A. K.</td>
<td>370, 379, 405</td>
</tr>
<tr>
<td>Riecken, L.</td>
<td>415, 417, 427</td>
</tr>
<tr>
<td>Rigg, M.</td>
<td>61, 79</td>
</tr>
<tr>
<td>Ringer, R. C.</td>
<td>76</td>
</tr>
<tr>
<td>Rios, J.</td>
<td>327</td>
</tr>
<tr>
<td>Rivera, T. M.</td>
<td>506, 508</td>
</tr>
<tr>
<td>Roberts, D. R.</td>
<td>46, 257, 276, 280, 419, 428</td>
</tr>
<tr>
<td>Roberts, K. H.</td>
<td>457, 473</td>
</tr>
<tr>
<td>Robertson, I.</td>
<td>647</td>
</tr>
<tr>
<td>Robertson, P. J.</td>
<td>46, 257, 276, 280</td>
</tr>
<tr>
<td>Robinson-Easely, C.</td>
<td>599, 603</td>
</tr>
<tr>
<td>Rooney, W. M.</td>
<td>494</td>
</tr>
<tr>
<td>Rosen, B.</td>
<td>393, 404</td>
</tr>
<tr>
<td>Rosenbaum, J. E.</td>
<td>548, 563</td>
</tr>
<tr>
<td>Rosenthal, R. A.</td>
<td>457, 473</td>
</tr>
<tr>
<td>Ross, D.</td>
<td>328, 334</td>
</tr>
<tr>
<td>Ross, M. S.</td>
<td>612, 648</td>
</tr>
<tr>
<td>Ross, R.</td>
<td>469, 473, 566, 590</td>
</tr>
<tr>
<td>Rothstein, L. R.</td>
<td>140, 152, 155</td>
</tr>
<tr>
<td>Rotter, J. B.</td>
<td>649</td>
</tr>
<tr>
<td>Rouillard, L.</td>
<td>138, 153</td>
</tr>
<tr>
<td>Rousseau, J-J.</td>
<td>188</td>
</tr>
<tr>
<td>Rowney, J.</td>
<td>506, 509</td>
</tr>
<tr>
<td>Ruh, R. A.</td>
<td>457, 473</td>
</tr>
<tr>
<td>Ruh, R. H.</td>
<td>419, 428</td>
</tr>
<tr>
<td>Rushing, W. A.</td>
<td>633, 649</td>
</tr>
<tr>
<td>Rybowiak, J. A.</td>
<td>79</td>
</tr>
<tr>
<td>Saari, L. M.</td>
<td>276, 279</td>
</tr>
<tr>
<td>Santa Maria, R.</td>
<td>507, 509</td>
</tr>
<tr>
<td>Sapolsky, M.</td>
<td>129</td>
</tr>
<tr>
<td>Sashkin, M.</td>
<td>47, 160, 181, 273, 280</td>
</tr>
<tr>
<td>Savaya, R.</td>
<td>80</td>
</tr>
<tr>
<td>Sayles, R.</td>
<td>127, 129</td>
</tr>
<tr>
<td>Saxena, K. K.</td>
<td>79</td>
</tr>
<tr>
<td>Schachter, S.</td>
<td>415, 417, 426</td>
</tr>
<tr>
<td>Schein, E. H.</td>
<td>103, 129, 400, 405</td>
</tr>
<tr>
<td>Schermerhorn, J. R.</td>
<td>64, 76</td>
</tr>
<tr>
<td>Schindler-Rainman, 293</td>
<td></td>
</tr>
<tr>
<td>Schmidt, F. L.</td>
<td>637, 648</td>
</tr>
<tr>
<td>Schmidt, W. H.</td>
<td>590</td>
</tr>
<tr>
<td>Schon, D.</td>
<td>402</td>
</tr>
<tr>
<td>Scholl, R. W.</td>
<td>3, 82, 93</td>
</tr>
<tr>
<td>Schultz, R.</td>
<td>308</td>
</tr>
<tr>
<td>Schutz, W. C.</td>
<td>115, 116, 129</td>
</tr>
<tr>
<td>Schuster, M. H.</td>
<td>257, 280</td>
</tr>
<tr>
<td>Scott, D.</td>
<td>55, 80</td>
</tr>
<tr>
<td>Scott, P.</td>
<td>44, 141, 154, 282, 286, 285, 287, 388</td>
</tr>
<tr>
<td>Scott, W.</td>
<td>135, 155</td>
</tr>
<tr>
<td>Selznick, P.</td>
<td>127</td>
</tr>
<tr>
<td>Senge, P.</td>
<td>47, 400, 405</td>
</tr>
<tr>
<td>Sethi, D.</td>
<td>80</td>
</tr>
<tr>
<td>Seyle, H.</td>
<td>40, 448</td>
</tr>
<tr>
<td>Shafritz, J. S.</td>
<td>129</td>
</tr>
<tr>
<td>Shafritz, J. M.</td>
<td>135, 155</td>
</tr>
<tr>
<td>Shamir, B.</td>
<td>533, 534, 538, 540, 544</td>
</tr>
<tr>
<td>Shani, A.</td>
<td>257, 280</td>
</tr>
<tr>
<td>Sharma, D. V.</td>
<td>80</td>
</tr>
<tr>
<td>Sharma, M. P.</td>
<td>79</td>
</tr>
<tr>
<td>Shearer, J.</td>
<td>488, 493, 500</td>
</tr>
<tr>
<td>Shelley, B. D.</td>
<td>84, 94</td>
</tr>
<tr>
<td>Sherif, C.</td>
<td>234, 254</td>
</tr>
<tr>
<td>Sherif, M.</td>
<td>234, 254, 622, 628, 649</td>
</tr>
<tr>
<td>Sherwood, J.</td>
<td>369</td>
</tr>
</tbody>
</table>

Copyright © Marcel Dekker, Inc. All rights reserved.
Author Index

Shepard, H., 101, 129, 573, 588
Shepard, L. J., 479, 486, 495
Shephard, W., 179, 181, 319
Schachter, S., 426
Schon, D., 400, 402
Simon, H. A., 134, 146, 155, 287, 305
Singh, H., 79
Singh, J. A. L., 623, 649
Sink, D., 3, 44, 45, 65, 77, 82, 83, 93, 152, 154, 208, 209, 279, 292, 299, 319, 382, 396, 408, 427
Sinoway, B., 507, 509
Sipel, G. A., 406
Sinkley, B. D., 63, 68, 80, 280
Slater, E., 186, 196
Slevin, D. P., 111, 128, 129
Slole, A., 549, 564
Smallwood, N., 3, 4, 47
Smith, P., 136, 137, 155
Smith, P. C., 418, 428, 457, 473
Snoek, J. D., 457, 473
Snyderman, B. B., 379, 466, 473
Solomon, M., 535, 536, 543
Sonquist, J. A., 472
Sorensen, R. E., 573, 590
Spaeth, J. I., 635, 648
Spector, B., 592, 602
Stanley, J. C., 43, 85, 93
Steele, T. P., 276, 279
Steinbrenner, K., 453, 473
Stepanova, I., 51, 78
Stephens, D. B., 546, 563, 564
Stevenson, J. G., 418, 421, 427, 431, 432, 436, 448, 473, 477, 492, 494, 500, 509
Stewart, D., 590
Stokes, G. S., 617, 621, 635, 649
Strauss, H., 77
Sudbury, D. A., 543, 544
Sun, B.-C., 45, 47, 58, 60, 65, 67, 69, 77, 81, 82, 92, 93, 94, 132, 142, 144, 152, 154, 255, 256, 257, 258, 260, 271, 272, 279, 280, 333, 334, 345, 396, 382, 469, 473, 507, 508, 509, 563
Sutton, R. L., 562, 564
Srivastva, S., 383, 402, 592, 595, 602, 626, 647
Stokes, G. S., 649
Sussbauer, J. C., 427
Susman, G. I., 357, 380
Sutton, R. L., 43, 546, 564
Svystyntek, D. J., 43, 83, 92, 93
Swart, J. C., 500
Sweet, J., 162, 181
Swierczek, F. W., 3, 4, 43, 65, 76, 258, 260, 269, 271, 276, 278
Synderman, B. B., 648
Taber, T. D., 406
Tandon, R., 228, 614, 646
Tang, S. F. Y., 51, 78
Tannenbaum, R., 566, 590
Tavarashka, T., 80
Taylor, J., 590
Teasley, C. E., III, 139, 155
Tennis, C. N., 319, 582, 590
Terborg, J. A., 337, 319
Terpstra, D. E., 32, 47, 82, 83, 85, 89, 92, 94
Thatchenkery, T. J., 189, 195, 394, 402, 625, 647
Thomas, E. D., 276, 278
Thomas, J. M., 406
Thomas, K. W., 393, 406
Tichy, N. M., 284, 335, 345
Tolbert, P. S., 544
Torbert, P., 5, 44, 47, 77
Torbert, W. R., 232, 254, 473, 590
Torres, D., 44
Towns, D. M., 543, 544
Travis, R., 583, 589, 643
Trist, E. L., 370, 380, 405, 406
Truman, H., 470, 471
Tufte, E. R., 589
Tyabji, H., 155
Ulrich, D., 144, 152, 195, 274, 255, 335, 345, 383, 402, 631, 646
Vaill, P., 286, 290, 293
Van Meter, D. S., 585, 590
Van Slyck, A., 378
Vanek, G. K., 530
Veiga, J. F., 549, 564
Velthouse, B. A., 393, 345
Viveros-Long, A., 529
von Mises, L., 273, 280
Waclawski, J. W., 602
Walker, H., 80
Walker, M., 530
Walsh, P., 136, 153
Walsh, J. T., 406
Wanous, J., 46
Waterman, R. H., Jr., 286, 292, 295, 334, 345
Watkins, K. E., 47, 400, 401, 406
Ward, R. T., 568, 590
Warrick, D. D., 280, 375, 380, 426, 448, 494, 530
Watson, G., 546
Watzlawick, P., 590
Wayman, M., 80
Wayne, S. J., 47, 82, 83, 88, 89, 92, 94
Weakland, H. H., 590
Weber, M., 223
Weick, K., 188, 196
Weigel, G., 186, 192, 196
Weil, R., 412, 426
Weiner, R. L., 493, 494
Weisbord, M. R., 3, 4, 46, 47, 273, 280, 383, 394, 400, 402, 405, 406, 600
Weiss, S. E., 46
Whetten, D. A., 43, 155, 185
White, J. K., 419, 428, 457, 473
Whittaker, J. O., 251, 254
Willmering, B., 543, 544
Wildavsky, A., 155
Wilkins, A., 408, 428
Williams, R. F., 385, 406
Willmering, A., 544
Wilson, H K., 64, 76
Wilson, J. Q., 274, 280
Wilson, W., 146
Winn, A., 97, 129
Winter, D. G., 405
Wolfe, D. M., 457, 473
Wojtczak, K., 75
Woodman, R. W., 47, 82, 83, 85, 88, 89, 92, 93, 94, 153, 181, 195, 280, 402, 405, 406, 588, 602, 603, 647
Woodworth, W., 3, 4, 47
Wootten, J. T., 406
Worthy, J. C., 370, 380
Wright, P. L., 46, 82, 93
Yalom, I. D., 189, 196, 431, 476, 448, 476, 495
Yeager, S., 319, 450, 472, 473, 529, 565, 576, 578, 585, 587, 589
Yoon, J-I., 66, 67, 211, 213, 217, 224, 226, 229
Zand, D., 306, 573, 590
Zander, 306, 307
Zawacki, R. A., 16, 43
Zedeck, S., 544
Zhilina, D., 506, 508
Zikiye, A., 220, 229
Zikiye, R., 220, 229
Zingy, R. M., 623, 649
Zuckerman, M., 551, 554, 556, 563, 564
Subject Index

Action research
and demotions, 545–564
effects of, 155–119
and flexible work hours, 512–530
and flexible work place, 532–545
in health care, 356–378
at interface, 394–395
and regenerative interaction, 110–113
and situational features, 255–280
some features of, 102–105
in team building, 105–128

Alternative Work Schedules (AWS)
applications of, 520–526
and change, 565–586
constraints on, especially in public sector, 519
designs for, 516–519
and flexible work hours, 512–530
and 4 × 10 versions of, 515, 517
and group properties, 520–521
and increased productivity, 526–529
results of, 520–529

Appreciative Inquiry (AI)
and conceptual/operational shortfalls, 595–603
and conflict, 597–598

[Appreciative Inquiry (AI)]
and crisis of agreement, 627–628
critical perspectives on, 593–600
features of, 594–593
and ironies in OD, 623–628
and matrix for applications of, 621–628
and neglect of values, 600–601
and social constructionism, 599–600, 625–628
Assimilation vs. contrast effects, 251–252

Balance hypothesis, 252–253

Burnout, phase model of
and choice and change, 475–495, 497–509, 582–586
concurrent validity of, 501–505
and emotional conditions, 484–485
features of, 430–431
and global replications of, 498–507
and group properties, 429–447
and guides for change, 444–447
impact of OD designs on, 442–444
and ironies in OD, 640–644
measures of, 477–479, 499–501
[Burnout]
and OD success rates, 487–493
and organization structure, 489–493
and personal slack, 475–495
and physical symptoms, 478–484
and research in progress, 506–507
and Work Environment Scales, 433–438, 442–444

Change
analogs about, 572–574
conceptual contents for, 565–572
and optimum discrepancy for, 231–234
and phases of burnout, 582–588
and situational features, 258–280
tests for kinds of, 572–582
as Trinitarian, 567–572

Change, Trinitarian
as alpha, 567–568
as beta, 567, 568–570
and flexible-time, 512
as gamma, 567, 570
and phases of burnout, 582–586
tests for, 572–582

Configurational analysis, as central in tomorrow’s OD, 633–638

Conflict
as agreement, 409–411
and appreciative inquiry, 597–598
as disagreement, 409–411
and “Going to Abilene,” 597–598

Confucian Work Ethic
features of, 213–217
and success rates of OD in Korea, 223–224, 236–239

Crisis, kinds of
as agreement, 409–411
and burnout, 418–421
central roles in planned change, 407–409
and disagreement, 409–411
and success rates, 412–418, 421–426
and team building, 422–426

Cross-training
and bureaucratic structures, 138
in NPM, 138


Demotions
conceptual context for, 547–552
designs for, 551–555
some concerns about, 559–561
some consequences of, 555–558
typical neglect of, 545–546
and values, 558–560

Designs, limited purpose
applications of, 520–526
and flexible work hours, 512–530

Discrepancies in change
roles in OD, 231–233
and success rates in OD, 249–253
tests of, 235–254
three models of, 233–235

Empowerment
as central in planned change, 394–395
and demotions, 545–564
features of, 393–394
and flexible work hours, 511–536
and flexible work place, 531–544

Ethics, alternative work
Confucian, 211–229
Judaeo-Christian, 200–206
Organization Development, 197–207
and OD success rates, 206–209, 223–227
selected similarities between, 220–223

Evaluations of interventions for change
and definitions of “success,” 13–14
and longitudinal designs, 38–41
measurement in, 6–11
and persistence, 36–38
pessimism about, 4–5
and positive response bias in private sector (see also Success rates), 18–19, 81–94
Factor analysis
and consequences of factorial structures, 586–588
and Trinitarian change, 568–584
“Fade–in” effects, 37–38
“Fade–out” effects, 36–37, 393–394
Flexible-place
conceptual niche of, 531–533
designs for, 532–535
and post-bureaucratic structures, 539–540
some consequences of, 533–539, 540
Flexible work hours
and alternative work schedules, 520–526
categories related to, 512–515
as limited purpose designs, 512
and post-bureaucratic structures, 527–529
results of applications, 520–526
Freedom
as conceptual context for OD, 189–229
and defining ethics, 197–229
as responsible, 189–195
two varieties of, 186–195
Group Behavior Inventory (GBI), 115–118
Groups
and burnout, 429–447
and guides for change, 439–441
and OD designs, 436–439
perspectives of, 429–447
and Work Environment Scales, 433–438
Health care
and common contextual features, 348–350
and interaction, 359–360
measures in, 352–356
overall philosophy of intervention in, 350–352
planned change in, 347–380
[Health care]
some results of intervention in, 361–377
“Heliotropic hypothesis,” 624–625
Individuals
and differences in as they impact change, 450–472
features of, 429–447, 449–473
and OD success rates, 466–472
as “one person, one vote,” 449–472
in QWL surveys, 452–466
Interaction, two systems of centrality in OD, some consequences of, 111–113
Ironies in OD
and Appreciative Inquiry, 623–628
and burnout, 641–644
and configurational analysis, 633–638
proposals to deal with, 610–636
re-introduction to, 605–610
and structural features, 629–633
and Trinitarian change, 640–644
Job enrichment
in bureaucratic structures (see also New Public Management), 130–140, 144–150
Korea, OD applications in
and Confucian ethic, 211–229
success rates of, 223–224, 236–237
Labor, U.S. Department of
and accepted principles of change, 333–339
contextual features in, 323–333
design option in, 341–344
planned change in, 322–341, 389–390
Large-scale change
in business, 281–309
general neglect of, 282–284
and generic effects in, 315–317
Subject Index

[Large-scale change]
in government, 158–181, 321–345
in health care, 347–380
and implementation, 288–292
at interface, 382–466
and implementation, 288–292
primer on, 284–292
results in, 158–181, 308–315
some designs for, 292–308
and strategic planning, 285–317
in U.S. Department of Labor, 322–341, 389–390
“Learning organizations,” 393–394
Likert Profile of Organizational Characteristics, 120–122, 235–251, 569–570
Maslach Burnout Inventory, 420–488, 499–501
“Maturing product line,” 618–619
Metropolitan Atlanta Rapid Transit Authority (MARTA)
and OD start-ups, 95, 129
as replication, 158–181
Michigan list of physical symptoms, 478–479
Multiple Affect Adjective Check List (MAACL), 555–560
Networks, organizational, 631–633
New Public Management (NPM)
cacophonies in, 140–148
features of, 133–140
as “liberation,” 133–140
and organization structures, 146–148
and OD success rates, 150–152
as undercut by bureaucratic model, 149–150
Organization Development (OD)
and action research, 102–125
Appreciative Inquiry in, 591–603
and burnout, 425–447, 475–495, 582–586
and change, 231–254, 565–586
and configurational analysis, 633–638
[Organization Development (OD)]
and Confucian Ethic, 66–71, 211–229
and cross-national aspects, 61–63, 66–71
and demotions, 548–564
in developmental settings, 51–80, 211–229
and empowerment, 393–395, 511–530
features of, 2–10, 183–229
and flexible work hours, 512–530
and flexible work place, 532–545
in government, 158–181, 321–345
and groups, 429–447
in health care, 347–380
and individual features, 449–473
at the interface, 381–402
interventions in, 8–10
and Judaeo-Christian Work Ethic, 200–206
and kinds of crises, 407–426
in Korea, 66–71, 223–224, 236–239
and large system change, 361–363, 281–377
and New Public Management, 131–155
and OD Work Ethic, 197–209
and organization structure, 146–150
and personal slack, 475–495
and policy interventions, 441–444
and “positive response bias,” 81–94
and Quality of Working Life, 19–42, 61–94, 255–280
and replication, 157–210
and responsible freedom, 183–229
and strategic planning, 292–316
and team building, 95, 236–247
and transitions, 158–181
and Trinitarian change, 567–572, 641–644
and Work Environment Scales, 436–444
Copyright © Marcel Dekker, Inc. All rights reserved.
Planning, Programming, Budgeting Systems (PPBS) in the U.S. Department of State, 141–142

Politics/Administration, interface of change at, 384–389, 395–401
features of, 382–383
generalizations about, 389–401
kinds of applications at, 383–388
“Positive response bias”
features of, 82–84
and methodological rigor, 84–88
and modest effects on OD success rates, 88–91
Practices, “best” and “good” as undercutting NPM, 144–149

Project ACORD (or Action for Organization Development), 5

Quality of Working Life (QWL)
features of, 2, 255–258
and “positive response bias,” 81–94
and situational features, 255–280
success rates in, 19–42, 69–71, 262–280
types of (see also New Public Management), 255–257, 265–269


Replications in MARTA, 158–181
varieties of, 95, 157–210

Situational features in planned change and four classes of interventions, 260–261, 266–270
hypotheses about, 258–279
in OD, 250–259, 273–280
and OD success rates, 259–273
in QWL, 250–258
Slack, personal
and burnout, 477–487, 497–509
in choice and change, 475–495
and emotional conditions, 484–486

[Slack, personal]
and implications for OD, 476–479, 486–493

Small group analysis, as stuck, 620–621
Social constructionism in Appreciative Inquiry, 599–601, 625–628
and neglect of values, 600–601
Social desirability in change, 249–250
Sociotechnical change, example of, 347–380
State, U.S. Department of and OD in, 140–142
and PPBS, 140–142
Strategic planning in NPM, 139
research on role in OD, 292–316
Structures, organization bureaucratic model of, 146–148, 149–150, 489–493
and burnout, 489–493
division model of, 139, 146–148
and flexible work place, 539–540
and flexible work time, 527–529
flow of work model of, 146–148, 489–490
matrix model of, 139
in sociotechnical change, 347–380
Success rates in planned change at interface, 381–402
and kinds of crises, 407–426
Team-building
and burnout, 422–426
confrontation designs for, 103–108
effects of, 109–119, 125–126
in MARTA, 95–128
at multiple levels, 119–128
and organic managerial systems, 100–102
and success rates, 421–426
Total Quality Management (TQM)
in NPM, 139
Transitions, signals for, 97–101

Values in OD/QWL
in business applications, 158–181, 296–299
in Confucian Ethic, 66–71
and cross-cultural considerations, 70–72

[Values in OD/QWL]
in government applications, 158–165
in health care, 350–352, 359–360
neglect of, 600–601
and social constructionism, 597–601

Work Environment Scales (WES)
features of, 433–434
impact of OD designs on, 437–439