SOCIAL SCIENCE IN GOVERNMENT

THE ROLE OF POLICY RESEARCHERS

New Edition

Richard P. Nathan

**The
Rockefeller
Institute
Press**

Rockefeller Institute Press, Albany, New York 12203-1003

First edition 2000
Printed in the United States of America

The Rockefeller Institute Press
The Nelson A. Rockefeller Institute of Government
411 State Street
Albany, New York 12203-1003

Library of Congress Cataloging-in-Publication Data

Nathan, Richard P.
Social science in government : the role of policy researchers / Richard P. Nathan.--
New ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-914341-65-0 -- ISBN 0-914341-66-9 (pbk.)
1. Social sciences--Research--United States. 2. Policy sciences--Research--United States. 3. United States--Politics and government--1945-1989. I. Title.

H62.5.U5 N34 2000
320'.6--dc21

99-087567

ISBN: 0-914341-65-0 Cloth
0-914341-66-9 Paper

For Mary

CONTENTS

Tables and Figures

Foreword

More than ten years have passed since Basic Books published the first edition of Richard P. Nathan's *Social Science in Government*. In the intervening decade, the author has been involved in many new public policy research projects that offer rich material — new case studies — to highlight both the pitfalls and possibilities of demonstration and evaluation research. These new cases do not merely pile up examples in support of an old argument. By highlighting the pluses and minuses of this kind of public policy research, they have stimulated a fresh look at Nathan's basic argument from beginning to end.

Once the new version of this book was finished, Dick Nathan was faced with a question about next steps and asked me what he should do. I said he ought to publish this book with the Rockefeller Institute Press. Many of the case studies described are based on research done at the Institute. More to the point, the book's central argument makes the case for the Institute's core mission. Dick agreed, but was unsure about what the process should be. The Institute Press's success depends on its reputation. Manuscripts go through a normal process of anonymous peer reviews, and these reviews can lead to rejection as well as extensive rewriting. There was no way the Press could conduct these peer reviews properly if Nathan were to be the person to solicit reviewers, as he does for other books.

So, for this volume, I acted as publisher during the book's early stages. I solicited reviews, passed them on anonymously, and wrote one myself. These were serious comments that led — as good reviews will

— to considerable revisions. The book is better for the work of the reviewers, and I believe the Rockefeller Institute Press is stronger for having adopted this process. The final product speaks for itself.

Michael J. Malbin
Rockefeller Institute and University at Albany,
State University of New York

Preface to the New Edition

This is a new edition of a book I published a decade ago with Basic Books. Since the earlier version of this book is out of print, I decided to revise and reissue it. The ideas presented are important to me, and I have continued to develop and refine them. The new edition reflects this further contemplation. It also contains several additional chapters that update the earlier edition with new case material to bring to life important questions about how social science can be useful and used in American government.

Besides the five new chapters, a number of the original chapters have been updated. This edition also contains a new final chapter on the limits and possibilities of applied social science. The case material used in this book focuses on major public policy issues, highlighting applied social science research in one policy area — welfare reform and related employment programs and social services. I hope the book can serve three purposes: as a primer for people interested in how social scientists can help social policymakers learn about what works; as a history of welfare policy research; and as a personal retrospective on a career in applied social science in which learning what works and what doesn't has been a central endeavor.

Many people have influenced the ideas advanced in this book, particularly those who worked on the studies used as case material. One of the main sources of case material is ten field network evaluation studies I conducted with colleagues at three policy research centers — the

Brookings Institution, the Urban and Regional Research Center of the Woodrow Wilson School at Princeton University, and the Nelson A. Rockefeller Institute of Government in Albany, the public policy research arm of the State University of New York.

The second major source of case material is research by the Manpower Demonstration Research Corporation (MDRC), founded in 1974. MDRC has conducted a large number of social experiments to test new programs designed to deal with the hardest problems of society's most disadvantaged people. I was a member of the board of directors of the Manpower Demonstration Research Corporation from 1974 to 1997. I admire the work done by its extraordinary staff, dedicated to applied public policy research that makes a difference. At the same time, I should make it clear that I did not participate as a researcher in the conduct of MDRC studies. My discussion of lessons drawn from these studies represents my own personal interpretations.

I also have drawn on the work of other public policy researchers, many of whom are friends of long standing, who willingly (and in some cases without realizing it) helped to shape my ideas. Scores of people, a number of them mentioned along the way in this book, have produced books and papers on which parts of this book are based. I also want to thank colleagues who read the manuscript for the earlier edition of this book and provided valuable reactions and suggestions: Orley C. Ashenfelter, Rebecca Blank, Robert F. Cook, Martha A. Derthick, Paul R. Dommel, Eli Ginzberg, David A. Long, Gilbert Y. Steiner, Donald E. Stokes, Aaron Wildavsky, and Michael Wiseman.

Michael Malbin served as publisher of this new edition and provided valuable advice and assistance. Thomas L. Gais and Irene Lurie are co-authors of two of the new chapters in this book on welfare and employment studies. Robert L. Cohen provided editorial assistance. Howard Rolston of the U.S. Department of Health and Human Services, Rebecca Blank, and Burt Barnow read this new manuscript and made valuable suggestions.

The Ford Foundation and the Florence and John Schumann Foundation provided the financial support for the original version of this book, written during a sabbatical year from Princeton University in 1986-1987. Special appreciation for this support is due to Susan Berresford and Shepard Forman of the Ford Foundation and William Mullins of the Schumann Foundation. Martin Kessler, my original editor at Basic Books, was a wise critic and a reservoir of helpful advice; Suzanne L. Wagner at Basic Books served as project editor for the original book. Nan Nash, my secretary at Princeton University, helped me put many versions of the original manuscript into our word processors; Carol Kuhl patiently followed suit at the Rockefeller Institute for this edition. Jonathan Jacobson and John Lago proofread the original manuscript, and John Lago provided research assistance as well. Michael Cooper, director of publications at the Rockefeller Institute, took charge of producing this second edition assisted by Marilyn McCabe. Irene Pavone ably provided proofreading and research assistance for this new edition.

New chapters in this book discuss two laws passed after the earlier version of the book was written. Fortuitously for me, evaluation studies of these two national welfare reform laws were helpful in updating this edition in a way that carries forward the earlier discussion of this central social policy issue. Appreciation needs to be expressed to two federal agencies — the U.S. Departments of Health and Human Services, and Labor. They funded the evaluations conducted by the Rockefeller Institute of Government discussed in the new chapters on welfare reform. I also greatly appreciate the support provided by the W. K. Kellogg Foundation for the field network research by the Rockefeller Institute on the implementation of the Personal Responsibility Act signed into law by President Bill Clinton in 1996.

Richard P. Nathan
Albany, New York

PART I

INTRODUCTION

1

Applying Social Science to Government

Social scientists who conduct applied research have a role, and it is an important one — to provide intellectual input in order to inform and assist the governmental process. My view is that the proper role for this action research is to educate, not advocate. But this is not an easy role to play. There is an understandable tendency among social scientists to want to use expert knowledge to advance their own ideas and values. The way the media behaves encourages this. Journalists frequently cite experts on both sides of controversial issues, thereby reinforcing the temptation for applied social scientists to act like politicians since their work is so often used as fodder for argument. The result is that too much of the time too many social scientists act too much like politicians.

This is an especially serious problem in fields of social policy. It is institutionalized in research conducted in graduate schools of social welfare, public health, education, and public affairs. In a parallel way, position-taking behavior has become standard and expected behavior in the professional organizations aligned with these institutions. Young people are trained as social scientists with the idea that activism involving the definition of public problems and the presentation of recommendations for their solution is good professional behavior in the social sciences. The net result is that applied research in the field of public affairs has become infused with an op-ed mentality that in an automatic and unconscious way mitigates against research that has a traditional and rigorous knowledge-building purpose.

It is not a simple standard to say that public policy research should be evenhanded and dispassionate and that policy researchers should not take sides. In fact, there has been an almost constant debate in academic circles over whether it is possible to be evenhanded as a social scientist. In one sense, social science, believing as it does in the scientific method and in rational analysis, is a point of view. My reference here to evenhandedness is meant in the ideological sense of not being liberal or conservative, centrist or decentralist, coddlers or admonishers, in public affairs.

Politicians have a very different role from public policy researchers, involving two incompatible jobs — to serve as transmitters of values, and to advance their own values and ideas. The way politicians carry out this delicate balancing act is a good basis on which to judge whether they are good politicians. Politicians make decisions on several grounds — on the ground of what their constituents want, on the ground of what they themselves believe, and on the basis of expertise — that is, the intellectual input to the policy process. This third input to the policy process is the focus of this book.

Politicians also and increasingly vote their pocketbook, reflecting the views of large contributors who finance ever more costly electoral campaigns. I hope this problem of the undue influence of large campaign contributions will be alleviated despite the fact that it stubbornly resists reform efforts.

A newspaper column by Michael M. Weinstein of *The New York Times* on how economists view the issue of school classroom size showed how hard it is to sort out social science knowledge and opinion in the political process. Weinstein juxtaposed the work of two economists whom he referred to as an "odd couple."[1] The occasion for his article was the long-standing debate on President Clinton's proposal for new federal spending to put 100,000 more teachers in elementary and secondary school classrooms. Democrats supported him. Republicans, while not opposing the $1 billion plus in new funding to be provided for education, argued that it should be appropriated flexibly to the states, since more teachers may not be every state's top priority.

One member of Weinstein's odd couple was Alan Krueger, an economist at Princeton University. Krueger, according to Weinstein, produced "some of the research results that the Administration uses to bolster its case for smaller class size." The other member of the odd couple is also an economist, Eric Hunushek, a professor at the University of Rochester, who was described by Weinstein as publishing "one study after another arguing that additional spending on schools wastes taxpayer money."

Neither Krueger nor Hunushek are offenders in the sense just discussed of abandoning impartiality and joining the political fray in their research activities. In fact, it is hard to see why they are an odd couple at all, as they are very much alike. Both are respected policy researchers and empiricists. Weinstein's column, published under the rubric "The Economic Scene," depicted the two scholars as agreeing on only one thing. They agreed, said Weinstein, "that Congress should resolve the many unanswered questions by running careful demonstration projects to figure out whether a national program to cut class size can work."[2]

For me, the moral of this tale is twofold. One moral is that it is very hard for even the best applied social scientists to avoid being drawn into the political thicket. The second moral is more subtle and pertains to the quality and nature of proof in social science. I am not sanguine that what Weinstein calls the "careful demonstration projects" advocated by Krueger and Hunushek can ever fully resolve questions like "whether a national program to cut classroom size can work." Applied social science by public policy researchers can aid the political process by amplifying issues and elucidating options, but it can rarely unequivocally and conclusively settle the most emotional, big-stakes political issues.

The Point of View of This Book

A common abbreviation used in filmmaking to show how the camera is positioned is "POV," standing for point of view. Movie scripts are not easy to read. The camera is always moving. It is much easier to get the picture on film than by reading a script on paper: What is the camera looking at? Who is doing the looking? As a teacher, I have found this convention a useful one in the classroom. I tell students studying American government, especially students interested in public service careers, that they should always be mindful of their own POV. And more importantly, they also and always should be mindful of the POV of the people they are dealing with.

It is a good idea in writing to state one's point of view clearly and up front. The purpose of this book on the role of policy researchers in government is to present lessons I have learned in a career in applied social science. The focus is on the work public policy researchers do outside of government. The book also discusses the role of applied social science inside government. Many of my colleagues among policy researchers, as has been my own case, move in and out of government as "inners and outers."* My experience in the federal government helped me develop ideas that influenced my choices of research subjects and the conduct of studies I have worked on; however, the largest part of my career has been as a political scientist engaged in policy research outside of government.

Most of the research projects I have worked on are field network evaluation studies. My role in this kind of policy research began with a national study of the effects of the general revenue sharing program enacted in 1972. (The idea of revenue sharing was to provide flexible grants-in-aid to states and localities on a basis that enabled them to set and carry out their own priorities.) Previous to 1972, I had directed domestic policy studies for Nelson A. Rockefeller's presidential campaigns, which included the fields of federalism and intergovernmental relations. Later on, during the first term of the Nixon administration, I served as an assistant director of the U.S. Office of Management and Budget, and in this capacity participated in drafting federal revenue sharing legislation. Shortly after Nixon was elected for his second term in 1972, I left government and moved back to the Brookings Institution, where I had previously been a research staff member from 1966 to 1969. At the request of the Ford Foundation, I organized a

* I believe this experience is beneficial on both sides, and have urged academic colleagues to get a taste of government as a way both to enrich their scholarship and enhance its usefulness.

nineteen-state evaluation at Brookings of the effects of the new revenue sharing law.

The questions that had to be dealt with in designing this first field network evaluation study help to make my point about how hard it is to study the effects of government policies. Arthur Okun, who had chaired the Council of Economic Advisors under President Johnson, was then a senior fellow at Brookings. He had doubts about the field-research approach. What would you say, he asked, if your mother gave you a check for your birthday and wanted to know what you did with it? His point was that the *fungibility* of federal grants-in-aid (the essential notion being that all money is green) makes it difficult — well nigh impossible, I think he said — to know what happens to such a gift.

My career was at a critical juncture. I responded to Okun that I was confident that studying state and local behavior in policy-making and implementation with regard to the uses of revenue sharing funds would tell us a great deal about what different governments (rich and poor, big and little) do with their shared revenue. Afterwards, Gilbert Y. Steiner, director of governmental studies at Brookings, said he was disposed to march ahead. After all, we had the research money.

The conclusion I reached years later, when we were deeply involved in the field evaluation of the revenue sharing program, was that my instinct in responding to Okun was correct. We did learn a great deal systematically about the uses of shared revenue. Many recipient jurisdictions, as it turned out, were wary of adding these funds to their program base — i.e., using this found-money for ongoing operating purposes. One-time capital purposes were a major use. This was especially the case for small, relatively well off, and fiscally conservative local governments. They feared locking revenue sharing money into their fiscal base and later having to raise taxes or lay off civil service workers when the federal government changed the rules or turned off the spigot, which it eventually did.

Not satisfied with focusing only on this fiscal question, the revenue sharing evaluation also examined the effects of the program in the functional areas where these funds were used for new-spending purposes. We also studied their distribution: Was the allocation of shared revenue redistributive? And we studied their political effects: Who decided on the use of the funds? How was the political role and structure of different types of governments affected?

Later on, the Brookings Institution was approached by the research office of the U.S. Department of Housing and Urban Development to conduct a similar field network evaluation of another form of federal financial aid to states and localities — block grants — specifically in this case, the Community Development Block Grant program. In the same way, we were asked soon afterwards to conduct a study of the public service employment program established under the Comprehensive Employment and Training Act (CETA). Seven other field evaluation studies have been conducted over the years by networks of indigenous researchers in multiple governmental jurisdictions, including:

- ❖ A study of the effects of all federal grants on large cities.
- ❖ A study of the effects of President Reagan's "New Federalism" cuts and changes in federal grant-in-aid programs.
- ❖ A study in New Jersey of the effects of Urban Development Action Grants (UDAG).
- ❖ A study of the implementation of the Job Opportunities and Basic Skills Training (JOBS) program for welfare reform under the Family Support Act of 1988.
- ❖ A study of the start-up and early implementation of President Clinton's Urban Empowerment Zone and Enterprise Community program.
- ❖ A study of the Neighborhood Preservation Initiative sponsored by the Pew Charitable Trusts to aid working-class neighborhoods.
- ❖ A study, still ongoing, of the effects of the national welfare reform law, the Personal Responsibility and Work Opportunities Reconciliation Act, signed by President Clinton in 1996.

The main units of analysis in all ten studies are *institutions*. The studies have been conducted by networks of field researchers coordinated by a central staff. The focus has been on the *implementation* of new policies, broadly assessing their fiscal, programmatic, distributional, and political effects on state and local governments, nonprofit organizations, and private contractors. Were the policies we studied implemented the way they were supposed to be?

My "POV," which is central to my reason for writing this book, is that these applied social science studies were useful and used in governmental processes, and at the same time that they contributed to scholarship on American federalism. Almost all of the field researchers were professors at universities. In these studies, they answered the same sets of questions in preparing their analytical reports that were then combined by a central staff. While the basic approach is inductive, in the sense of learning as we went along, there tended to be an implicit, and sometimes explicit, set of expectations (which could be called "hypotheses") about likely program effects built into the framework for each round of the data collection for these studies. Most of these studies were longitudinal; in some cases they involved three or four rounds of field data collection.[3] This book, which considers the way these policy research projects assisted governmental processes, also looks at the work of other policy researchers with different research purposes and methods.

This is a good time to reissue the book for two reasons. One is that I have clearer ideas now about the themes developed in the original edition. A second reason is that welfare policy, which is the main subject of the studies used in this book as case material, has changed greatly in the decade since the first edition was written. Two major national welfare reform laws have been enacted, both of which I have studied with colleagues using the field network evaluation methodology. A virtual cornucopia of other studies also have been undertaken on the effects of these two welfare reform laws, so the subject is a big and interesting one.

The first new welfare law passed since the earlier version of this book was written is the 1988 Family Support Act, enacted in the final year of Ronald Reagan's presidency. The bottom line with respect to this law is that it did not make much of an impact. The second national welfare law examined is the Personal Responsibility and Work Opportunity Reconciliation Act of 1996. In contrast to the earlier act, the 1996 act has already had a pervasive impact in American federalism on a wide range of public agencies and nonprofit organizations at every level of government.

The Role of Applied Social Science

As stated earlier, politicians act on the basis of their beliefs and those of their constituents. They also act on the basis of expert knowledge. This, however, is only one — and often not the main — input to public policy-making. One reason for this is that we simply don't have definitive knowledge that would enable politicians to base all, or even most, public policy decisions on uncontrovertible scientific evidence. Over the years, social scientists have developed three bad habits that are important for the discussion in this book of the need for a realistic view of the role of applied social science.

The first bad habit of social scientists is the tendency to want to emulate the natural sciences. Beatrice Webb, who worked with Charles Booth in England on the early development of survey research methods, considered this problem in a book about her life as what she called a "social investigator." Webb's family had a close friendship with Herbert Spencer, whose philosophizing about unfettered capitalism could not have been more antithetical to the career and political path later followed by Webb. In her book, Webb used her relationship with Spencer to illustrate her ideas about the meaning and role of social science. Referring to Spencer's writing, she said, "There was a riddle in the application of the scientific method to human nature which continuously worried me, and which still leaves me doubtful. Can the objective method, pure and undefiled, be applied to human mentality; can you, for instance, observe sufficiently correctly to forecast consequences, mental characteristics which you do not yourself possess?"[4] This is at the nub of debates about the meaning of social science: Can social sciences predict human behavior using "the objective method, pure and undefiled?" My view can be summed up as follows:

> Social scientists should be realistic about the nature of the terrain in relation to the strength of their theories and methods. The data simply do not exist, nor can they ever be collected, which would tell us everything we want to know about every attitude, emotion, and form of behavior of every individual and relevant group in society in such a way that we could use these data to construct models and produce theories that would approach the predictive power of theories in the natural sciences.

The second bad habit of social scientists that is pertinent here is overspecialization. Modern social science is a bubbling pot of disciplines and subspecialties that have compartmentalized human society. A generation ago, economist Joseph Schumpeter said, "Our time revolts against the inexorable necessity of specialization and therefore cries out for synthesis, nowhere so loudly as in the social sciences in which the non-professional element counts so much."[5] In a similar way, social theorist Abraham Kaplan was caustic in comments about the fragmented and competitive character of the social sciences. "The fragmentation of science into 'schools' is by no means unknown in as rigorous a discipline as mathematics; what is striking in behavioral science is how unsympathetic and even hostile to one another such schools are."[6] This bad habit of social science is reflected both in teaching and professional practice. The boundaries between fields and subfields are rooted in reward systems that measure achievement by one's ability to do independent work in a single discipline. The most common operational mode of modern social science is one computer terminal in one office operating in one discipline.

The third bad habit of social scientists is closely related to the first two, the tendency for social scientists to prefer quantitative research designs and techniques and downgrade qualitative research methods and data. Qualitative research can and does use numbers to interpret observations made by social scientists. It often involves presenting such data and in the form of generalizations with an empirical base, although typically (but not always) short of being able to bring to bear mathematical proofs based on samples of sufficient size to infer causality. Quantitative studies, on the other hand, may use data that are less than ideally precise, but in this case with the emphasis on having an adequate sample size to be able to attempt to identify causal relationships. Again, Abraham Kaplan was on target when he criticized what he called "the law of instrument" in referring to quantitative studies. Said Kaplan, "Give a small boy a hammer, and he will find that everything he encounters needs pounding."[7] The hammer of modern social science is the computer.

This book highlights three types of applied social science that can be useful and used in the governmental process: *(1) demonstration research* to test possible new policies and major programmatic departures; *(2) evaluation research* to assess the effects on ongoing public programs; and *(3) studies of conditions and trends*. The bulk of my attention is devoted to the first two of these categories, demonstration and evaluation research. Four key points are:

1. Demonstration studies to test new policies and program approaches and evaluation studies of ongoing policies and programs are different in ways that have not been sufficiently taken into account by the sponsors and funders of public policy research and by researchers.

2. Evaluation research to assess ongoing public programs is the frontier of applied social science. Social scientists interested in policy research have the most untapped potential and some of the hardest challenges in this area.

3. In designing and conducting both demonstration and evaluation research, greater attention should be given to the missing links of applied social science research. Two missing links highlighted in this book are those between disciplines within the social sciences and those between quantitative and qualitative research methods and data.

4. In selecting the subjects for both demonstration and evaluation studies, priority should be given to situations in which three conditions apply: first, that policymakers and government officials are genuinely interested in the questions being asked; second, that they are uncertain about the answers; and third, that they are willing to wait for them.

These ideas reflect a view about a role for public policy research that is both positive and limiting. I view applied social science as a supporting player on the political stage. We have come a long way in our thinking from the over-optimism of the post-World War II period, which is discussed in the next chapter. Implicit in the earlier view was the idea that social engineering could replace what many intellectuals viewed as an excessively competitive political process dominated by self-interest. Daniel Patrick Moynihan, whose extraordinary career bridges the worlds of social science and public policy, said in 1969, "the role of social science lies not in the formation of social policy, but in the measurement of its results."[8] This is the view of public policy research presented in this book.

In sum, I believe applied social science research should emphasize *how* to do things rather than *what* should be done. Combining social science disciplines in such studies involves much more than bringing additional data to bear. It adds variables to the research equation. When we leave out disciplines, we leave out dimensions of human behavior. Disciplinary compartmentalization rooted in a single intellectual paradigm distorts human experience. Economists, to their credit, have been the dominant players in applied social science, stressing quantitative methods and data. Other disciplines that place greater emphasis on qualitative methods and data need to be brought into the picture, especially political science and sociology.

Ultimately, the kind of applied social science that is conducted depends on the role played by government and foundation officials acting as the sponsors and funders of policy research; they are in the catbird seat. They occupy the critical territory between the producers and the consumers of public policy research. The last chapter of this book considers the role of the sponsors of policy research, how they behave, and how their behavior might change.

Endnotes

1 Michael M. Weinstein, "Economists Raise Questions About Educational Priorities," *The New York Times*, 15 October, 1998, p. 28-A.

2 Ibid.

3 Professor Irene Lurie of the State University of New York at Albany, recently wrote a paper on this methodology. See "Field Network Studies," forthcoming. See also Richard P. Nathan, "The Methodology for Field Network Evaluation Studies" (Chatham, NJ: Chatham House Publishers, 1982).

4 Beatrice Webb, *My Apprenticeship* (Cambridge, UK: Cambridge University Press, 1979), pp. 139-40.

5 Joseph Schumpeter, *Capitalism, Socialism, and Democracy* (London, UK: George Allen and Unwin, 1976), p. 45.

6 Abraham Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science* (Scranton, PA: Chandler, 1964), p. 39.

7 Ibid., p. 26.

8 Daniel Patrick Moynihan, *Maximum Feasible Misunderstanding: Community Action in the War on Poverty* (New York: Free Press, 1969), p. 193.

2

Optimism and Disillusionment

In the immediate post-World War II period in the United States, there were high expectations for applied social science. In the lead essay to a volume on what was called "the policy sciences," Harold Laswell wrote, "It is probable that the policy-science orientation in the United States will be directed towards providing the knowledge needed to improve the practice of democracy."[1] Laswell's co-editor for this volume, Daniel Lerner, wrote that he envisioned a future in which social science would deal with "the new human problems raised by the endlessly changing lifeways of modern society."[2] There is, remarked Lerner, "an integral connection between social science and social democracy."[3]

Education in the social sciences was infused with the idea of improving social and economic conditions. It was hoped that practitioners of different social science disciplines would work together. Robert S. Lynd, in a book entitled *Knowledge for What? The Place of Social Science in American Culture*, called for merging the work of the social sciences by focusing "on the concept of culture." By culture, Lynd referred to "all the things that a group of people inhabiting a common geographical area do, the ways they do things and the ways they think and feel about things, their material tools and their values and symbols." Lynd saw this as meeting "the need for an all inclusive frame of reference for all the social sciences."[4]

In this period, as efforts to apply social science in government came into vogue, one discipline, economics, increasingly came to have the upper hand. There are several ways economics is applied in government. Theories of macroeconomics support policies for producing stable, noninflationary economic growth. Microeconomics, in turn, is applied to government in two ways — in budgeting, for example, through the "planning-programming-budgeting" system, and through the demonstration and evaluation research movement involving large research projects to test new program ideas and to evaluate ongoing programs.

Applying Macroeconomics

Under President Kennedy in the 1960s, a feeling of ebullience emerged that the economy could be managed in a way that would repeal the business cycle. In an unprecedented action, Kennedy proposed a large reduction in federal personal and corporate income taxes in 1963 explicitly to apply Keynes's theories to national economic policy-making by deploying unused resources to create jobs and stimulate growth. After Kennedy was assassinated, Lyndon Johnson carried through on this tax-cut proposal; Congress enacted a $16 billion tax reduction in 1964. Walter W. Heller, then chairman of the Council of Economic Advisors, who was both the principal architect of the tax cut plan and an artful phrasemaker for it, described this economic policy as the nation "declaring a fiscal dividend" to combat "fiscal drag" in a faltering economy.

The following year, Heller delivered the Godkin Lectures at Harvard University on the subject of the new role of economics in government. The first sentence set the tone: "Economics has come of age in the 1960s." Heller went on, "The age of the economist arrived on the New Frontier, and is firmly entrenched in the Great Society." The Keynesian influence was clear. "What economists have wrought is not the creation of a 'new economics'," said Heller, "but the completion of the Keynesian revolution — thirty years after John Maynard Keynes fired the opening salvo."[5]

Unfortunately, the business cycle was not repealed. Stagflation (recession and inflation occurring simultaneously) in the 1970s generated major controversies among and between economists and politicians. Consensus on Keynesian economics fell away. This is not to suggest that economists lost their voice in macroeconomic policy-making. Quite the contrary. At the highest levels, they have gained influence as members of the Council of Economic Advisors and the Federal Reserve Board and as officials of the U.S. Treasury Department, the Office of Management and Budget, and other government agencies.

The Planning-Programming-Budgeting System

A major development of the 1960s for the application to government of microeconomics was Lyndon Johnson's effort in the mid-1960s to remake the budget process by establishing the Planning-Programming-Budgeting (PPB) system. This system, which involved a large number of economists working inside government, was initially applied in the Defense Department by Robert S. McNamara, formerly president of the Ford Motor Company, who was appointed by President Kennedy as secretary of defense. McNamara and his "whiz kid" policy analysts used systems analysis to compare alternative weapon systems in an effort to increase the leverage of the secretary of defense in relation to the individual services. Before the Vietnam War escalated, the McNamara system was riding high. As a result Johnson decided that this approach should be applied, not just in the defense sector, but to all of government. He embraced PPB with typical gusto. An executive order issued in August 1965 directed all agencies to apply the systems' analysis approach to the entire budget process. The government-wide approach announced by Johnson set forth extensive and detailed requirements that were supposed to take effect immediately. All federal agencies were required to prepare planning documents and issue-analysis papers to back up recommendations to the Budget Bureau (the name of the bureau before it was reorganized and renamed the Office of Management and Budget under President Nixon.)

According to the *Bulletin* issued by the Budget Bureau to set up this new system, the objective of PPB was "to improve the basis for major program decisions in the operating agencies and in the Executive Office of the President. Program objectives are to be identified and alternative methods to be subjected to systematic comparison." The system consisted of three main reports to be prepared by all government agencies: (1) *program memoranda*, comparing the cost and effectiveness of major alternative programs and describing the agency's strategy; (2) *special analytic studies* on current and longer-term issues; and (3) *program and financial plans*, multi-year summaries of agency programs in terms of their outputs, costs, and financing needs over a five-year period.[6]

The experience of PPB was, to say the least, disappointing. The paper just didn't flow. Or else it overflowed. Federal officials used bureaucratic stratagems to continue to manage the budget process in the way that they were used to doing. In some cases, they simply did not submit the required planning memoranda and analysis documents. Agency officials and the staff of the Bureau of the Budget operated in these cases as if nothing had changed. In other cases, agencies swamped the Bureau of the Budget with elaborate planning documents and issue papers that few if any high officials of the submitting agency had even seen. Three years after President Johnson's bold announcement of a government-wide PPB system, his successor, Richard M. Nixon, quietly issued a presidential memorandum abolishing the system. "Agencies are no longer required to . . . " it stated, and then it summarized the steps of the PPB system. Political scientist Allen Schick observed in an article on this non-event (or at least relatively unnoticed event) called "A Death in the Bureaucracy," that "No mention was made in the memo of the three initials which dazzled the world of budgeting" when the PPB system was announced.[7]

The unfortunate thing to me is that it seems as if government, at least the federal government, is incapable of learning when it comes to management reforms like PPB. The PPB system was followed by other systems with similarly inflated expectations under Presidents Nixon and Carter. President Nixon's management reform system was called MBO, standing for "management by objectives." Carter's plan was called ZBB, standing for "zero-based budgeting." Both had ambitious and elaborate requirements and no large or lasting effects. In 1993, Congress got into the act with the Government Performance and Results Act, which requires all federal agencies to produce annual reports on their progress in converting programs from being process-oriented to focusing on results. Because it is embedded in law with a long lead time for setting it up, and because both the legislative and executive branches are on record as supporting it, there is hope that this new law will have a more lasting influence in enhancing the role of policy analysis and research as inputs to government decision making.[8]

Demonstration and Evaluation Research

The second development in the application of microeconomics to policy-making is the one I am most interested in, the demonstration and evaluation research movement. As emphasized in the first chapter, I make a distinction between these two types of applied social science research — demonstrations to test new programs and evaluations of ongoing programs.

Most observers of government policy-making think the high point of the influence of economists coincided with the Democratic presidencies of the 1960s and fell off when Nixon and the Republicans came into power. It is true that the PPB system died a quiet death during Nixon's presidency, but other applications of economics discussed in this chapter lasted longer.

As it turned out, the timing of the demonstration and evaluation research movement was decidedly different from that of applied macroeconomics and the PPB system. Because of the long period of gestation required to design, conduct, and report on demonstration and evaluation studies, the bulk of the work done under this heading for applied social science was done in the 1970s and afterwards, not in the 1960s. Some of the major demonstration and evaluation studies, notably the negative income tax experiments described and discussed in the next chapter, were begun in the 1960s, but the results of these studies, which were planned and designed under President Johnson, did not emerge until well beyond his presidency. Most of the negative income tax research was carried out in the Nixon years, and in addition a number of large new demonstration and evaluation studies were launched under Nixon and his successors.

Cynics have depicted Nixon's interest and that of other Republicans in social experimentation as a strategy to delay the development of new programs or as an excuse for not enacting them. The same point is made — with more basis for doing so — about the Reagan administration. However, a look at the record of new program adoptions, policy changes, and domestic spending in the Nixon and Ford years casts doubt on this interpretation with respect to them. Large domestic initiatives were adopted under Nixon, and many existing programs were expanded. Nixon did not say so, but he was a big spender on the home front, and President Ford tended to follow suit. Total domestic spending by the federal government under Nixon rose from 10.3 percent of the gross national product at the outset of his presidency to 13.7 percent in 1974. Social Security accounted for less than half (1.4 percentage points) of this increase. Federal aid to state and local governments, which includes the revenue sharing program (mentioned in chapter 1) and block grants and other federal grant-in-aid programs that Nixon advanced, accounted for most of the rest of this 3.4 percentage point increase in the gross national product attributable to domestic public spending.

Although the PPB system is long departed from the scene, it left a legacy. Its legacy was the growth in the size and stature of the planning staffs in federal agencies throughout government, created to participate in PPB processes. In point of fact, PPB and its successor budgetary reforms have made the budget process at all levels of government (federal, state, and local) more analytical. They brought people trained in economics and policy analysis into government big time. Many of these people stayed, and more of them came, and their contributions have been important and lasting. Moreover, government planning and analysis offices often were, and in many cases still are, the sponsors of large demonstration and evaluation research projects.

Outside of government, the legacy of PPB is the army of contract researchers whose role is conducting demonstration and evaluation research — i.e., testing potential new programs and evaluating ongoing programs. Indeed, one reason for treating the demonstration and evaluation research movement separately has to do with precisely this point — its organization. Unlike the PPB system, the operation of the demonstration and evaluation research movement for the most part involves researchers working outside of government at universities, think tanks, and private consulting firms. Moreover, many universities across the country, seeing PPB and the growth of applied economics in government, public policy analysis, and demonstration and evaluation research as a new market, established and expanded public policy graduate schools in the 1970s. However, doubts rose in the late 1960s and 1970s about the efficacy of social science as a guide to policy-making.

Doubts Arise

The basic question as to whether human behavior can be studied the way it is in the natural sciences was the theme of a popular book in the 1980s on the limits of economics by Lester C. Thurow. Thurow wrote that economists "can't find hard empirical constants, such as the speed of light in physics, because economists are not studying the immutable rules of nature but the mutable laws that govern human behavior."[9] He added that mainstream economics reflected "more an academic need for an internal theoretical consistency and rigor than it reflects observable, measurable realities in the world we live in."[10]

Among the best-known statements associated with such self-criticism in economics was that by Wassily Leontief, a Nobel laureate as the inventor of input-output analysis. In his address in 1971 as president of the American Economic Association, Leontief criticized his colleagues for their overemphasis on theory building and their failure to establish "systematic cooperative relationships across the traditional frontiers now separating economics from these adjoining fields."[11] More than a decade later, Leontief identified his principal concern as the use of deductive models grounded in data inadequate to the task at hand. "Page after page of professional economic journals are filled with mathematical formulas leading the reader from sets of more or less plausible but entirely arbitrary assumptions to precisely stated but irrelevant theoretical conclusions." These data, he said, "fall short of what would have been required for concrete, more detailed understanding of the structure and function of a modern economic system."[12]

In a similar vein, a book published in this period by Andrew M. Karmarck complained that economists "insist on quantification but completely overlook the need to understand how much precision is actually attainable in the accuracy of the numbers used. . . . Very little attention is paid to the quality of data — data are dumped into a computer without close examination." Echoing Thurow, Karmarck said, "Forecasting future parameters or variables, is incomparably more difficult in economics than in the physical sciences. . . . In fields concerned with human behavior like economics, in contrast, constant or stable phenomena can rarely be relied on."[13]

A good suggestion made in response to Leontief's 1982 letter to *Science* magazine was contained in an article by Barbara R. Bergmann, who urged economists to broaden their scope of inquiry and rethink their position toward other, and what she called "softer" social science disciplines. "Economists might look with profit to the practice of social scientists in other disciplines, whose lower status and whose methods of research economists have been wont to scorn."[14] In a similar vein, Robert Kuttner, in the *Atlantic Monthly*, criticized economists for being "highly abstract, mathematical, and deductive rather than curious about institutions." Kuttner quoted Charles L. Schultze, like Leontief a former president of the American Economic Association, as saying, "When you dig deep down, economists are scared to death of being sociologists. The one great thing we have going for us is the premise that individuals act rationally in trying to satisfy their preferences. This is an incredibly powerful tool because you can model it."[15]

Other Views

At one time, there was an active institutional school of economics in the United States, led by John R. Commons, an economist at the University of Wisconsin. In the 1920s, Commons contended that institutions, as collectivities of people, behave differently than the sum of the individuals within them. Commons believed institutional economics should give greater emphasis to political and organizational behavior. This of course is the point I expressed earlier about the need for inter-disciplinary applied social science, although by no means gainsaying the fact that economics as a discipline deserve the most credit for the rise of applied social science in government in recent years. In my opinion, other social science disciplines, especially political science, sociology, and psychology now should be involved in a broadened paradigm for public policy research.

In one such effort to tie economics to psychology, Robert E. Lucas, Jr., of the University of Chicago, in the 1980s questioned the ability of economists to forecast the future on the ground that they do not give adequate attention to attitudes and shifts in opinion that influence behavior.[16] Lucas's critique was the key point of the "supply-side" challenge to macroeconomic theory advanced under President Reagan. Actually, Keynes made this same point about the importance of the psychological aspect of economic behavior in assessing economic systems.

Another economist whose work I admire, Albert O. Hirschman, has observed that economists neglect mood and attitudinal variables at their peril. Modern economics, he said, "oversimplifies." Specifically, he criticized "parsimony" in the discipline, asserting that the conventional approach presents "too simple minded an account of even such fundamental economic processes as consumption and production."[17]

Donald McCloskey, an economist who became something of a historian of his discipline, traced these and other ideas about the discipline. In the lead article in the June 1983 *Journal of Economic Literature,* McCloskey called the methodology of economics "modernism." Modernism for McCloskey refers to "the credo of Scientific Methods, known mockingly among its critics as the Received View," which "is an amalgam of logical positivism, behavioralism, operationalism, and the hypothetico-deductive model of science."[18] But then McCloskey added a more positive comment. He said that in practice economists do not follow their official rhetoric, and that in his view it is a good thing they don't. "If they did they would stand silent on human capital, the law of demand, random walks down Wall Street, the elasticity of demand for gasoline, and most other matters about which they commonly speak." Looking at the role economists play in government, in business, and in their classrooms, McCloskey argued, "economists in fact argue on far wider grounds. Their genuine workaday rhetoric, the way they argue inside their heads or their seminar rooms, diverges from the official rhetoric."[19]

For purposes of this chapter, my point is not that economists should lose standing as leaders of applied social science. They have the strongest body of theory and the most workable tools to aid governmental decision making. Walter Heller's vision in the 1960s of a Keynesian consensus assuring stable, noninflationary economic growth has been tempered, but not in a way that diminishes the influence of macroeconomics in government. In the case of microeconomics, the same situation applies in terms of lowered, but more realistic, expectations.

Demise of the PPB System

Charles L. Schultze, who is mentioned above as suggesting that economists might usefully act more like sociologists, is an important figure in the story of the Planning-Programming-Budgeting system. Schultze was a leader of the effort in the mid-1960s to apply microeconomics in government; he was director of the Bureau of the Budget when the PPB system was put in place.

After leaving the Johnson administration, Schultze in a thoughtful series of retrospective lectures delivered at the University of California in 1968 described his experience implementing PPB.[20] Schultze began by describing the aims and elements of the PPB system. But it is what he did next that is most interesting. He addressed the critique of PPB associated with the views of Charles E. Lindblom. Throughout his academic career, Lindblom has emphasized the inherently incremental nature of the American political process and the ways it is antithetical to the planning values embodied in government and management textbooks and in the PPB system.

After setting forth Lindblom's critique of PPB, Schultze proceeded to wrestle with Lindblom's incrementalist position. Some readers may find, as I did, that Lindblom seemed to come out the winner. Schultze indicated sympathy with Lindblom's argument about the difficulty of specifying the objectives of government programs.

Considerable precision, he said, is a necessary precondition for assessing programs rigorously and then comparing them according to the degree to which they achieve their objectives, as the PPB system required. Lindblom on the other hand argued that in most cases the goals of public programs cannot be precisely specified because politicians often are purposefully, and from their point of view, rationally vague about their objectives in the belief that this enhances their chances of putting together a sufficiently broad coalition to ensure that policies they favor are adopted.

Digging deeper into Lindblom's ideas about what he calls "the science of muddling through," Schultze argued that PPB must adapt to, and become part of, the political process. He noted that this is what actually happened under President Johnson. According to Schultze, "program planning and evaluation staffs in the agency head's immediate office, created by the PPB system, strengthen the role of the agency head in relations with the operating units." Schultze observed: "Cynics to the contrary notwithstanding, knowledge is power."[21] In effect, PPB made economists actors in the political process. This is a good thing, but it is surely a different thing from the exuberance of early advocates of applied social science that it could be the basis for scientific policy-making.

Like Lindblom, Bertram Gross argues that the major problem with PPB was that "microeconomists who have repeatedly used the term effectiveness have been chary about admitting, let alone explicitly stating, that they have been engaging in attempts at cause-effect analysis." Continuing, Gross said, "Once this is brought into the open, it becomes clear that estimates of presumed results must take into account many possible causative factors other than the program under analysis, and that many such factors, being social, psychological, and political in nature, are not readily understandable in terms of economics, or any other single discipline."[22] Again, the emphasis is on combining social science disciplines.

An important part of Charles L. Schultze's rethinking of the PPB system in his Gaither Lectures was what he said about the scope of the system. Looking back, Schultze said he saw wisdom in narrowing the scope of PPB: "I propose as a working hypothesis that analysis can operate with fewer constraints and can profit from a consideration of a wider range of alternatives in programs that produce pure public goods and do not directly affect the structure of institutional and political power than in programs that produce quasi-public goods, and fundamentally affect income distribution, or impinge on the power structure."[23] This is a substantial concession. It is hard to think of government programs, particularly domestic programs, that do not, in Schultze's terms, "fundamentally affect income distribution" and "impinge upon the power structure." In fact, in the earlier development of cost-benefit analysis, the focus was on relatively narrow-gauged studies of public works programs, such as water resource projects. Economist Roland McKean, a pioneer of cost-benefit analysis, contended that this analytical method is generally most useful for lower-level decisions involving "comparatively narrow problems of choice" where "the alternatives are usually rather close substitutes."[24]

Assessing Demonstration and Evaluation Research

What I call the demonstration and evaluation research movement can be seen as both a reaction to, and an outgrowth of, the PPB system. In effect it was a more modest effort to do part of what the PPB system was supposed to do. It concentrated on a limited number of problems and programs. Properly viewed, the demonstration and evaluation research movement reflects this same idea that some subjects should be selected for serious, rigorous study, and furthermore that this should be done without assuming, as under PPB, that experts can compare all major and related uses of governmental resources. Economist Richard R. Nelson in commenting on the relationship between demonstration and evaluation research and the PPB system sounded very much like Charles L. Schultze and Charles E. Lindblom. "It does seem fair to question," said Nelson, ". . . whether the new philosophy of experimentalism represents greater sophistication regarding the implications of models of optimization over time under conditions of uncertainty, or disguised acceptance of a strategy of muddling through."[25]

In reviewing issues raised about applying economics in government, I have relied mainly on the writings of economists, who in many cases were themselves participants in the events described. In considering the demonstration and evaluation research movement, I follow the same practice, beginning with a book published in 1985 by Henry J. Aaron, entitled *Politics and the Professors: The Great Society in Perspective.* Aaron was a high-level official on the policy research staff of the U.S. Department of Health, Education and Welfare in the 1960s and served as the assistant secretary for planning and evaluation of the successor department (the Department of Health and Human Services) in the Carter administration in the late 1970s.

The theme and tone of Aaron's book on applied social science were pessimistic. His conclusion was not only that demonstrations and evaluations in the field of social policy are very difficult to do, but that these studies often produce findings about programs that fall far short of what was promised by politicians. Aaron portrayed program evaluation as "a newly developed art" that "certified the ineffectuality of these programs," referring to social programs launched in the Great Society period under President Johnson. "Far from being an instrument for evenhanded, objective deliberation, evaluation was transmuted into 'forensic social science'."[26]

When I began collecting material for this book, I was aided in doing so by teaching seminars on applied social science at Princeton University. In one seminar session, a participant, a mid-career student from the U.S. Department of Defense, questioned why applied social science is so focused on social programs, particularly programs to aid the poor. "Liberals have shot themselves in the foot," he said, "by emphasizing studies that often show the limits of social programs and only rarely their successes." The image this suggests fits Henry Aaron's critique; applied social science in many instances has ended up undermining the case for social programs. According to Aaron, evaluation studies often revealed modest, if any, program results. Assuming that most researchers believe not only in the utility of applied social science, but also in the value of social programs, this is a gloomy situation from their point of view. Aaron summed up in the following terms: "The role that research and experimentation played in the demise of the simple faiths of the early 1960s was not accidental. The process by which R&E [research and experimentation] is created corrodes the kind of simple faiths on which political movements are built."[27]

Two points in Aaron's conclusion stand out. One is his accent on "faith" as the basis for political action. The other is his reference to research and experimentation as "corrosive" of the "simple faiths on which political movements are built." Several years after Aaron's book was published, he made a tongue-in-cheek comment about the role of demonstration evaluation research, this time directed at the efforts of Ronald Reagan's budget director, David A. Stockman, to cut spending on social research:

> Mr. Stockman is making a grave mistake in trying to put us all out of work. He has not realized that we are the *instrumentality for inaction.* By diverting us to teaching rather than research or even to still more reputable ways of earning a living, he will make easier the growth of ideas for activist social change undisturbed by critical analyses when the mood of the country shifts.[28]

Aaron was not alone in expressing doubts about demonstration and evaluation research. There were other critics in this period. Sar A. Levitan and Gregory Wurzburg said,

> It is not just a question of obtaining better data or spending a few million more dollars on evaluations. The problem lies in the basic assumptions of the methodologies employed by most evaluations, and in the choice of who is entrusted with the task. This entire field remains an art. The much-touted objective scientific conclusions of evaluations are too often found to be based on hidden political and social value judgments or personal interest.[29]

In a similar vein, Richard F. Elmore reviewed the studies of youth employment programs and concluded: "The fact that we find it easy to discredit interventions that merely deliver services, but difficult to find scientifically valid solutions to chronic social problems, may mean that we have gotten too sophisticated in using the rhetoric of social science to justify social interventions."[30]

On the conservative side, Charles Murray, in his influential book *Losing Ground: American Social Policy, 1950-1980,* made many of the same points as Aaron, Levitan and Wurzburg, and Elmore. He, too, served as an evaluator of social programs in the 1970s. Murray's book is best known for his conclusion, presented as a "thought experiment," that goes one step — a very big step — further than Aaron. He maintained not only that social programs enacted in the 1960s failed, but that in many cases they contributed to the problems they were supposed to solve. Murray said such programs should be abolished. He contended that "white condescension towards blacks" took a form that undermined work incentives, family structure, and self-esteem. His solution was "to repeal every bit of legislation and reverse every court decision that in any way requires, recommends, or awards differential treatment according to race, and thereby put us back on the track that we left in 1965." Murray was specific in naming programs that should be eliminated.

> The proposed program, our final and most ambitious thought experiment, consists of scrapping the entire federal welfare and income support structure for working-aged persons, including AFDC, Medicaid, Food Stamps, Unemployment Insurance, Worker's Compensation, subsidized housing, disability insurance, and the rest. It would leave the working-aged person with no recourse whatsoever except the job market, family members, friends, and public or private locally funded services. It is the Alexandrian solution: cut the knot, for there is no way to untie it.[31]

Besides lambasting social programs, Charles Murray devoted major attention in *Losing Ground* to applied social science research. One section of his book on the role of research is called "Hard Noses and Soft Data." In it, Murray said: "In the spirit of cost-effectiveness that McNamara has taken to the Pentagon, the early poverty warriors were prepared to be judged on the hardest of hard-nosed measures of success." Murray went on, "Social scientists who had been at the periphery of the policy process — sociologists, psychologists, political scientists

— had the answer: scientific evaluation. The merits of doing good would no longer rest on faith." The reference here to faith parallels Henry Aaron. Murray concluded his critique of applied social science research as follows: "Starting with the first evaluation reports in the mid-sixties and continuing to the present day, the results of these programs have been disappointing to their advocates and evidence of failure to their critics."[32]

If I had scrambled the various quotes about problems of applied social science, the reader would be hard put to sort them out on the basis of which statements had been made by liberals and which by conservatives. One is reminded of the riddle once posed by Aaron Wildavsky: What is the difference, he asked, between the New Deal and the Great Society? The answer, he said, evaluation research.

Of all the critical commentaries cited on the demonstration and evaluation research movement, the most despairing is from a technical paper by Gary Burtless and Robert H. Haveman, "Policy Lessons from Three Labor Market Experiments." The three experiments considered were the negative income tax experiments conducted in Seattle and Denver, the supported work demonstration undertaken by the Manpower Demonstration Research Corporation, and the Employment Opportunities Pilot Project carried out in the Carter years. Burtless and Haveman drew this conclusion: "Our experience in the last fifteen years has taught us that large-scale experiments can be relied on to teach us something of value about the policy in question, but what we are taught can seldom be relied on to aid the cause of reforming or improving policy."[33] They went on to say, "There is a moral here, and it is illustrated by the three experiments we have considered: if you advocate a particular policy reform or innovation, do not press to have it tested."[34] The motion picture scene this suggests (Woody Allen should direct it) would have social scientists sitting around a table trying to decide what program ideas they dislike most in order to test them and undermine their chances of adoption, if you will, "corroding the simple faiths" on which they are based. My view, as expressed in Parts II and III of this book is more upbeat, conditioned on realism and a broadened disciplinary perspective. Part II deals with demonstration studies; Part III with evaluations of ongoing programs.

Endnotes

1 Harold D. Laswell, "The Policy Orientation," in *The Policy Sciences*, eds. Daniel Lerner and Harold D. Laswell (Stanford, CA: Stanford University Press, 1951), p. 15.

2 Daniel Lerner, "Social Science: Whence and Whither?" in *The Meaning of the Social Sciences,* ed. Daniel Lerner (Gloucester, MA: Peter Smith, 1973), p. 14.

3 Ibid. Carol H. Weiss has written extensively about the enthusiasm of the early advocates of social research as an input to public policy-making. See, for example, her introductory essay in *Using Social Research in Public Policy-making,* ed. Carol H. Weiss (Lexington, MA: Lexington Books, 1977).

4 Robert S. Lynd, *Knowledge for What? The Place of Social Science in American Culture* (Princeton, NJ: Princeton University Press, 1948), p. 19.

5 Walter W. Heller, *New Dimensions of Political Economy* (Cambridge, MA: Harvard University Press, 1966), pp. 1-2.

6 U.S. Bureau of the Budget, Bulletin No. 68-9, April 12, 1968, in *Government Budgeting: Theory, Process, Politics*, eds. Albert C. Hyde and Jay M. Shafritz (Oak Park, IL: Moore, 1978), pp. 129, 130.

7 Allen Schick, "A Death in the Bureaucracy," in Hyde and Shafritz, *Government Budgeting*, p. 191.

8 Dall Forsythe, *Performance Management Comes to Washington: A Status Report on the Government Performance and Results Act* (Albany, NY: The Nelson A. Rockefeller Institute of Government, February 25, 2000).

9 Lester C. Thurow, *Dangerous Currents: The State of Economics* (New York: Random House, 1983), p. 8.

10 Ibid., p. xvi.

11 Wassily Leontief, "Theoretical Assumptions and Nonobserved Facts," *American Economic Review* 61 (March 1971): 4, 6.

12 Wassily Leontief, "Letters," *Science* 217 (July 9, 1982): 217.

13 Andrew M. Karmarck, *Economics and the Real World* (Oxford: Basil Blackwell, 1983), pp. 8, 9, 18.

14 Barbara R. Bergmann, "The Failures of Chair-Bound Science," *New York Times*, December 12, 1982.

15 Robert Kuttner, "The Poverty of Economics," *Atlantic Monthly*, February 1985, pp. 74, 76.

16 See Robert E. Lucas, Jr., "Expectations and the Neutrality of Money," *Journal of Economic Theory* (April 1972): 103-24.

17 Albert O. Hirschman, "Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse," *Bulletin of the American Academy of Arts and Sciences* 37 (May 1984): 11.

18 Donald N. McCloskey, "The Rhetoric of Economics," *Journal of Economic Literature* 21 (June 1983): 484.

19 Ibid., p. 482.

20 Charles L. Schultze, *The Politics and Economics of Public Spending* (Washington, DC: Brookings Institution, 1968).

21 Ibid., p. 94. See also Charles E. Lindblom, "The Science of Muddling Through," *Public Administration Review* 19 (spring 1959): 79-88.

22 Bertram M. Gross, "The New Systems Budgeting," in Hyde and Shafritz, *Government Budgeting*, p. 152.

23 Schultze, *Politics and Economics*, pp. 85-86.

24 Roland McKean, *Public Spending* (New York: McGraw-Hill, 1968), p. 141.

25 Richard R. Nelson, *The Moon and the Ghetto: An Essay on Public Policy Analysis* (New York: Norton, 1977), p. 32.

26 Henry J. Aaron, *Politics and the Professors: The Great Society in Perspective* (Washington, DC: Brookings Institution, 1978), pp. 4, 32. The phrase "forensic social science" is attributed by Aaron to Alice M. Rivlin.

27 Ibid., p. 159.

28 Henry J. Aaron, comment, in *Social Experimentation*, eds. Jerry A. Hausman and David A. Wise (Chicago, IL: University of Chicago Press, 1985), p. 277.

29 See Sar A. Levitan and Gregory Wurzburg, *Evaluating Federal Programs: An Uncertain Act* (Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, 1979). Statement provided by the authors.

30 Richard F. Elmore, "Knowledge Development under the Youth Employment Demonstration Projects Act, 1977-1981," paper prepared for the Committee of Youth Employment, National Research Council, National Academy of Science, Washington, DC: January 1985.

31 Charles Murray, *Losing Ground: American Social Policy, 1950-1980* (New York: Basic Books, 1984), pp. 222, 223, 227-28. AFDC stands for Aid for Families with Dependent Children.

32 Ibid., p. 36.

33 Gary Burtless and Robert H. Haveman, "Policy Lessons from Three Labor Market Experiments," in *Employment and Training: R & D Lessons*

Learned and Future Directions, ed. R. Thayne Robson (Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, 1984), p. 128.

34 Ibid. Burtless has softened his position on the value of social experiments. See Burtless and Larry L. Orr, "Are Classical Experiments Needed for Manpower Policy?" *Journal of Human Resources* 21 (fall 1986): 606-39.

PART II

DEMONSTRATION RESEARCH

3

The Nature of Demonstration Research

When the members of the House and Senate conference committee worked out a compromise on an emergency jobs bill to combat the 1981-82 recession, they added a provision at the behest of Representative Jamie Whitten, the powerful chairman of the House Appropriations Committee, to include $33 million for a highway project "demonstrating how a two-lane road can be widened to four lanes." A reporter for *The New York Times* covering the conference committee noted that "Mr. Whitten refused to say so, but all of the conferees expect that the money will wind up in Mr. Whitten's home district." It did. No mention was made in the *Times* story of a research design or a comparison road to be used as a basis for determining the efficacy of this road-widening demonstration.[1]

The use of the word "demonstration" in this way is not unusual. In some cases, the guise of research is more elaborate, but the underlying aim is the same — to use the cover of research to obtain funds for a pet project or program even though there is no or very little intention of studying its results systematically to decide whether a particular program idea should be replicated on a broader scale. This is not the meaning of the word "demonstration" used in this book. I am interested in demonstration research as a type of applied social science, conducted under conditions in which trained researchers apply their tools to produce results that can be used by policymakers to decide whether or not to adopt a particular course of action. Unfortunately, the line between real demonstrations and those with an ostensible, but not genuine, research purpose is often more subtle than in the case of Representative Whitten's highway project.

The next section provides a vocabulary lesson that defines what is tested in a demonstration study, the way social scientists think about program impacts, the "counterfactual" state, control groups, quasi-experiments, comparison sites, simulation techniques, dependent and independent variables, and the replication of the results of demonstration studies. The first and most significant demonstration research projects, which were central to the emergence of this kind of applied social science in the United States, were the negative income tax experiments. The chapter pays special attention to these demonstrations and their influence.

The Vocabulary of Demonstration Research

A potential new program being tested in a demonstration research project is referred to as a *treatment* or an *intervention*. Researchers seek to determine the impact of a particular policy by measuring selected characteristics of the members of the treatment group before, during, and after they have participated in a demonstration project. These characteristics are then compared to the characteristics for an untreated group of similar persons. The acid test is: Did the tested treatment make a difference? If so, what kind of a difference and how much of a difference did it make?

The hardest job is identifying an untreated group to be compared with the treatment group in order to establish the *counterfactual state*. The counterfactual state is what would have happened had there been no tested treatment. It is impossible to know the counterfactual for certain — that is, to have the same people both participate and not participate in a tested treatment. Instead researchers attempt to approximate this condition. One method for doing so is to have eligible participants randomly assigned, as in a lottery, either to a treatment group or to an untreated group called a *control group*. This experimental approach using random assignment in demonstration research was pioneered by

British statistician R. A. Fisher, and was used extensively in medicine before it was applied in the field of social policy.

Random assignment is strongly preferred by most social scientists over other research designs because it enables them to employ statistical techniques to establish causality and to assign a level of probability to the impact of a tested treatment. I am somewhat, but not fully, in agreement with this position. I believe random assignment is a good approach. Nevertheless, in my view many social scientists have gone overboard in advocating it. In other Western countries, the idea of conducting social experiments with random assignment has not taken hold as strongly as it has in the United States.

Random assignment is not always used in public policy demonstrations. Researchers also use what are called *quasi-experimental* methods. Thomas D. Cook and Donald T. Campbell define quasi-experimental studies in precisely these terms, as demonstration studies that do not use random assignment. Quasi-experimental studies, according to Cook and Campbell, "have treatments, outcome measures, and experimental units, but do not use random assignment to create the comparisons from which treatment-caused change is inferred."[2] The baseline group in a quasi-experiment is usually referred to as a comparison group rather than a control group. Researchers construct comparison groups in various ways. Comparison groups can be groups of people who are similar to treatment groups but are located in other places, for example, in a different city where the treatment being tested is not being administered. A popular alternative approach employs statistical techniques, using available data sets about people similar to the people in the treatment group for a demonstration study. This is the simulation or econometric method for establishing the counterfactual.

In some cases, demonstration studies are conducted without either a randomized control group or a statistically constructed comparison group, the object being to compare the treated group before and after they have participated in a tested program. Most public policy researchers do not like this approach because it is possible that the presumed effects (the before-and-after differences) might have occurred in any event, that is, in the absence of the tested treatment. In the field of job training programs, this is called "the aging-vat effect," referring to the idea that certain things happen to people (for instance, they work more and earn more) simply because they get older. This effect could confound a demonstration study that did not have either a randomly assigned control group or some kind of comparison group.

The bottom line is that random assignment is preferred by most public policy researchers because they believe it is the best way to predict whether a tested treatment will work if it is replicated, whereas the techniques used in quasi-experiments are regarded as less certain.

Two key terms need to be added — dependent variables and independent variables. Dependent variables refer to the outcomes of the tested treatment. They are also called "right-hand" variables. Independent variables, which are on the left-hand side of the equation, include two main types of variables, the treatment being tested and the characteristics of the people being treated.[3]

The Negative Income Tax Demonstrations

The most famous early demonstrations in the field of social policy in the United States that used random assignment were the negative income tax experiments. The negative income tax is an approach to welfare reform. Its aim is to provide financial incentives to working-age, able-bodied welfare recipients to encourage them to enter the labor force and ultimately become self-supporting. In order to provide these recipients with an incentive to work, those who work are allowed to retain some portion of their earnings. The rate at which their earnings are "taxed" (that is, reducing welfare benefits by some portion of each additional dollar earned) is called the "negative income tax rate," or more understandably the "welfare-reduction rate." This approach to welfare reform generated great enthusiasm, especially among liberals, in the 1960s and 1970s; some conservatives (notably economist Milton Friedman) embraced it too.

Negative income tax experiments were initiated under President Johnson. The main proponents of the negative income tax, who originally sought to have this idea adopted as part of President Johnson's "War on Poverty," were economists both inside and outside of government. This group included James Tobin, Robert Lampman, Joseph Pechman, Joseph Kershaw, Robert Levine (Levine was head of research in the Office of Economic Opportunity during this period), and Worth Bateman, a policy analyst in the U.S. Department of Health, Education, and Welfare. A negative income tax plan was proposed to the Johnson White House staff in September 1965, but, according to Robert Levine, "never taken seriously by the administration."[4] President Johnson and his secretary of Health, Education, and Welfare, Wilbur Cohen, were strongly opposed to it.

The most widely discussed and controversial issues about proposals for a negative income tax involved the effects of work incentives on able-bodied, working-age adults: Would a negative income tax increase or undermine work incentives, and by how much would it influence them? Proponents of the negative income tax argued that it would increase work incentives because recipients would be better off if they worked. Their total income (their earnings plus their welfare income) would rise as their earnings increased, so they would work more and work harder. But, as we shall see, this aspect of negative income tax plans also has its downside, and is more complicated than may appear on the surface.

The critical point about the labor-market effects of a negative income tax is that such plans also add people to the welfare rolls by virtue of introducing the work-incentive feature. While a negative income tax may increase work effort of people already on the welfare rolls, at the same time it may reduce the work incentive for people added to the rolls — i.e., because now these people do not need to work as hard to receive their current level of income.

People added to the welfare rolls were not getting welfare payments before, but are now eligible if their earnings from work are below what is known as the "break-even point," that is, the point at which welfare income phases out to zero as income rises. The ultimate question is what is the net effect of a negative income tax. Will the totality of people affected by a negative income tax — both those already on the rolls and those added to the rolls — work more or less because of the introduction of the negative income tax?

Politically, the negative income tax experiments inaugurated by the U.S. government in the 1960s represented a fallback. Rather than jettisoning this idea when it was advanced by senior officials of the Department of Health, Education, and Welfare (HEW) to Johnson's White House staff and rejected by them as the basis for a new national policy, supporters of the negative income tax approach saw a demonstration as a way to keep the idea on the policy agenda, although in a downgraded status. Officials of the Office of Economic Opportunity (OEO), which was the lead agency of Lyndon Johnson's "War on Poverty," advanced a plan developed by Heather Ross, then a doctoral student working for the White House Council of Economic Advisors, to conduct a demonstration with random assignment to test the negative income tax idea. OEO contracted with the Institute for Poverty Research at the University of Wisconsin in Madison and Mathematica Policy Research of Princeton, New Jersey, to design the experiment. Economists at Princeton University's Industrial Relations Section, working with the staff of Mathematica, had a major hand in the design of the experiment.

The role of social scientists in launching the negative income tax experiments raises an interesting political question. Some readers may feel that there is something wrong when social scientists conducting research are testing a policy they themselves favor, which was the case for most of the participants in the negative income tax experiments. Although I do not believe this is a problem, an important point needs to be added here: Social scientists can more legitimately conduct demonstration research to test new ideas they themselves favor if their research uses random assignment. Random assignment protects them. It reduces the problem of bias toward the tested plan on the part of both the researchers and the sponsors of a demonstration research project. But the catch is that randomized tests are expensive, time-consuming, and very hard to conduct.

A negative income tax is best viewed as having a number of movable parts. The design of a particular plan involves arranging these parts in a way that maximizes the objectives of the policymakers who support it. Invariably, they will want to achieve several purposes simultaneously, purposes that often are not compatible. Three major and frequent purposes are: (1) to establish an "adequate" level of income support; (2) to provide a work incentive for the people receiving this support; and (3) to avoid having a particular plan involve "excessive" costs. Typically, the two most movable parts in determining how these three objectives can be achieved simultaneously are the benefit level and the welfare-reduction rate. Holding the third purpose constant — i.e., setting the cost as a certain fixed amount — policymakers in the 1960s experimented with different combinations of benefit levels and welfare-reduction rates that in turn dictated varied break-even points. (The break-even point, to repeat, is the point at which a person's benefits under the terms of a particular negative income tax plan are reduced to zero. That is, as earnings rise and welfare benefits gradually decline, this is the point at which benefits end.)

Overall, the income-incidence pattern of the U.S. population is diamond-shaped. There are relatively smaller numbers of people in the bottom and top income groups. The largest numbers of people are in the middle ranges. Hence, coverage in population terms under a negative income tax plan increases materially as it adds people higher up the income ladder, into the middle ranges. The people added to welfare rolls in this way under a negative income tax plan receive a smaller benefit than people who are very poor, but the number of people involved can be large. The higher the break-even point under a negative income tax, the greater will be the cost for the people added. This can present problems to politicians who support the basic idea of a negative income tax, that is, if the costs of a given plan are regarded as too big.

The common and not surprising response of politicians in this situation is to compromise. They may, for example, decide to lower the welfare benefit or raise the welfare-reduction rate (say from a 30 percent welfare-reduction rate to 50 percent) in order to fit their plan within a cost parameter they regard as acceptable. The inexorable arithmetic involved here produces a conundrum for the designers of a negative income tax. The problem can be likened to squeezing a balloon: As you squeeze the air out of one area, it goes into another. Each time a compromise is made, such as lowering the basic benefit or raising the welfare-reduction rate, one of the arguments for the idea of having a negative income tax is weakened.

In the mid-1960s, when the original negative income tax experiments were being advanced, and as these aspects of the negative income tax idea became increasingly well understood, concerns about the impact of such a program in expanding the rolls and raising the costs of welfare gave political leaders cold feet about even testing the concept. When the time came to announce the start of the first demonstration, planned for New Jersey, officials of the Office of Economic Opportunity hesitated. The director of the OEO at the time was Sargent Shriver, John F. Kennedy's brother-in-law and a former head of the Peace Corps. Despite qualms, Shriver decided to proceed, but on a low-key basis. The contract for the first phase of the work on the New Jersey experiment was paid out of previously appropriated funds, so it was not necessary to seek an appropriation from the Congress. And the announcement of the demonstration was held up until after Congress had recessed for Labor Day in 1967. In recognition of potential political pitfalls, Shriver's aides also convinced him to change the name of the program from a "negative income tax" to a "work incentives" program.[5] The new name, however, did not stick.

The first payments in the New Jersey negative income tax demonstration were made in Trenton, New Jersey, in 1968. This demonstration, which also included other New Jersey cities, along with Scranton, Pennsylvania, focused on two-parent welfare families. Altogether, the demonstration had a total sample of 1,350 families in the treatment and control groups. The families in the treatment group were targeted for varying levels of payments combined with varying welfare-reduction rates applied to the earnings they received from work. Eight negative income tax plans were tested, with ranges of the basic income guarantee from 50 percent to 125 percent of the poverty line and with three welfare-reduction rates of 30 percent, 50 percent, and 70 percent.[6] The experiment lasted three years.

This capsule description does not begin to do justice to the great complexity of the task of designing and mounting this first big American social experiment. One year into the work, the State of New Jersey introduced a welfare plan statewide that aided two-parent welfare families on a more generous basis than several of the tested negative income tax plans. As a result, a new negative income tax experimental plan with a higher benefit level had to be quickly instituted. The dispute that arose between the University of Wisconsin's Institute for Poverty Research (which was responsible for designing the New Jersey study) and Mathematica Policy Research (which had the operations contract for this study) about how to do this became so intense that at the last minute it was assigned to an outside expert, James Tobin, a professor of economics at Yale, for arbitration.[7]

Owing to these and other complications of the New Jersey study, experts at the Department of Health, Education, and Welfare (which inherited the responsibility for the negative income tax experiments from the Office of Economic Opportunity in the Nixon years) took the position that results from the New Jersey demonstrations should not be used for policy purposes. Instead, they said, major policy reliance should be placed on the results from larger successor experiments designed during the execution phase of the New Jersey study. These successor negative income tax experiments were conducted in two cities, Seattle and Denver, as well as a number of smaller rural communities in three states, Indiana, Iowa, and North Carolina.[8] This expanded negative income tax research is known as SIME/DIME. (The "IME" portion of this acronym stands for "income-maintenance experiment," the "S" for Seattle, and the "D" for Denver.) Planning for this next series of negative income tax experiments began under President Nixon in 1970. Eventually, they included 4,800 families, both intact (two-parent) and single-parent families in both the urban and rural sites; this was over three times larger than the New Jersey sample.

As it turned out, the idea of a negative income tax as tested both in the New Jersey and the SIME/DIME (Seattle/Denver) studies seeped into the policy process long before the final results of the experiments were available. Welfare expert Gilbert Y. Steiner characterized this as "research following reality."[9] Nixon's welfare reform plan, called the Family Assistance Plan or FAP, was announced in 1969, just as the first payments were being made in the New Jersey and Scranton, Pennsylvania demonstrations. Nixon's plan embodied many of the ideas of a negative income tax, and it was heavily influenced by holdover policy analysts inside the government.

I was involved in the development of Nixon's proposals for welfare reform as a federal official during this period (assistant director of the Office of Management and Budget and later deputy undersecretary of the Department of Health, Education, and Welfare). Eventually, I came to view Nixon's FAP plan, grounded as it was in the concept of the negative income tax, as the wrong road to welfare reform. However that was later on. For purposes of this chapter, what is most interesting is the way in which the results of the negative income tax demonstrations affected the debates in Congress about Nixon's welfare reform plan.

Nixon's Family Assistance Plan, which he actively and strongly advocated and which received wide publicity at the time, passed the House twice under the strong leadership of Ways and Means Committee Chairman Wilbur Mills.[10] However, it came under fire in the Senate. Several senators, notably Finance Committee Chairman Russell Long (D-LA) and John Williams of Delaware, then the ranking Republican member of the committee, attacked the plan on the grounds that it would undermine, rather than enhance, work incentives. Both senators became astute analysts of the negative income tax idea. Under the pressure of their attacks and those of others, officials in the Office of Economic Opportunity, which still existed at the time (but soon afterwards was abolished), released "preliminary" results of the New Jersey demonstration in February 1970. These results indicated no adverse effects on work effort under the New Jersey negative income tax experiments. The OEO report went so far as to state that there was "in fact, a slight indication that participants' overall work effort actually increased during the initial test period."[11] Senator Williams, by then a vociferous opponent of Nixon's Family Assistance Plan, was indignant. He questioned both the veracity and the timing of the OEO report and called on the U.S. General Accounting Office (an investigative arm of the Congress) to review the OEO findings. The GAO responded saying that the report on the New Jersey study was "premature."[12] It is unlikely that this controversy over the early release of the New Jersey findings had an effect on the chances for passage of Nixon's Family Assistance Plan. It was already in deep trouble; however, it certainly did not help the advocates of social experimentation.

A similar political backfire occurred later over the results of the Seattle/Denver income maintenance demonstration. By now, a coterie of people had experience with the negative income tax experiments. In particular, Daniel Patrick Moynihan's role — always central on these issues — exemplifies the rise and fall of the negative income tax idea. Moynihan had been a senior advisor in the White House during Nixon's first term. Despite the fact that in his prior government service in the Kennedy and Johnson administrations and also as a professor at Harvard University he had been a proponent of universal children's allowances (rather than a negative income tax) as the best road to welfare reform, Moynihan became a leading and influential advocate of the negative income tax approach to welfare reform as embodied in Nixon's Family Assistance plan.[13] He teamed up with the secretary of Health, Education, and Welfare and long-time Nixon aide Robert Finch to convince Nixon that this plan, largely drafted by Worth Bateman and other holdover HEW policy analysts, was the best and most dramatic approach for overhauling welfare.

In this period, concern was widespread about the problems of welfare — rapidly rising costs and caseloads and large disparities in benefit levels among the states, with very low benefits in some states. Many observers believed that welfare (then called the Aid for Families with Dependent Children program, AFDC) encouraged families to break up or never form, and that it discouraged work effort. Vincent J. and Vee Burke, in their book on the history of Nixon's Family Assistance Plan, emphasized the part played by fast-rising welfare rolls in getting the issue of welfare reform, "typically shunned by the White House," on Nixon's agenda. In the decade of the 1960s, the Burkes said, the proportion of children on relief had doubled from 3.5 percent of those under eighteen in 1960 to 6.8 percent in 1969 and 8.7 percent in 1970. According to the Burkes, "The welfare explosion angered taxpayers and put severe pressure on state treasuries, especially in states with very big cities, such as Illinois, California, Pennsylvania, and New York."[14]

In this setting, Nixon saw welfare reform as an opportunity to surprise and outmaneuver liberals on social policy. He was especially influenced by Moynihan, then a White House aide, whose knowledge of the subject was extensive and whose engaging personality and knack for an elegant turn of phrase appealed greatly to Nixon. Nixon delighted in Moynihan's sprightly memos and conversations on this subject, compared with the buttoned-down style of most of his advisors.* This is not to say that Nixon lacked a personal commitment to his welfare reform plan. I saw him often in this period, and am convinced he was genuinely excited about this plan.

Nixon's Family Assistance Plan was announced in August 1969. As already mentioned, the New Jersey negative income tax experiments were barely under way. Full results of the New Jersey experiments would not appear for another four years. In 1978, a decade after Nixon's welfare reform plan was announced, the results of the Seattle/Denver negative income tax demonstrations became available. By then, the roles of many of the players had changed. Moynihan, now a Democratic U.S. Senator from New York, chaired the welfare subcommittee of the Senate Finance Committee. He used his subcommittee as a forum to examine the results of the Seattle/Denver research. The hearing record makes interesting reading.

Moynihan was the only senator in attendance. His exchanges with witnesses (most of whom were social scientists in fields closely related to Moynihan) resembled a graduate seminar in social science more than a congressional hearing. The main idea that emerged from this postmortem on the Seattle/Denver experiments was that the results undercut the idea of a negative income tax. At the hearing, researchers disagreed about the seriousness of the problems involved in ways that generally reflected their political orientation. Liberals among researchers acknowledged that the results undercut the case for a negative income tax,

---

* I still have a notebook of Moynihan's memos to Nixon. The first line of the first memo, dated January 13, 1969, is pure Moynihan: "Like the girl in the book about the crocodiles, I fear that I may end up telling you more about welfare in New York City than you want to know." The reference to crocodiles eludes me; nevertheless, this reflects the style Moynihan used in establishing an engaging personal relationship with Nixon.

but they tended to downplay the magnitude of these effects. Conservatives on the other hand were almost gleeful in their use of findings from the research to show the futility of the idea of a negative income tax.

Moynihan, in an interesting way, was in the middle. He had supported the idea of the negative income tax in the Nixon years, but now he sided with the conservative analysts in assessing its implications for social policy. Writing to William F. Buckley in 1978, he said, "We were wrong about a guaranteed income. Seemingly, it is calamitous."[15]

The results of the Seattle/Denver experiments undercut arguments of the proponents of the negative income tax in two ways. The payment schemes tested resulted in reduced net earnings and hours of work for recipients. They also appeared to have an adverse effect on families, encouraging family break-up rather than enhancing family stability, as was claimed would be the case.

In November 1978, when Moynihan held his second hearing on the experiments, Robert Spiegelman, director of the Seattle/Denver research for the Stanford Research Institute, presented what was treated as dramatic testimony on the adverse effects of these experiments on family stability. The report on the Seattle/Denver study also showed that the tested negative income tax plans caused substantial reductions in labor activity for persons enrolled. Gary Burtless and Robert Haveman, in summarizing these results, wrote that, "prime-aged men reduced their annual hours of work by 9 or 10 percent; . . . their spouses reduced annual hours by 17 to 20 percent; and . . . women heading single-parent families reduced annual hours by more than 20 percent — perhaps as much as 28 to 32 percent."[16]Again, the policy pot was boiling. These findings were issued just when officials of the Carter administration were putting together Carter's welfare reform plan, which like Nixon's had basic features of a negative income tax. The immediate result was a decision to trim back the Carter plan, because the research results increased the cost estimates for its benefit schedule.

Moynihan's view of these events is contained in a book on social policy he published in 1986, eight years after the hearings. In this book, he was especially critical of Carter's secretary of Health, Education and

Welfare, Joseph Califano, for his failure to present an assessment on the adverse effects of the Seattle/Denver demonstration on family stability, calling his behavior "inexcusable."[17] However, the most interesting element of Moynihan's postmortem in 1986 is his comments on the testimony presented by economist John Cogan of the Hoover Institute, which is located at, but not officially part of, Stanford University. Cogan, politically a conservative, testified at the Moynihan hearings in 1978 about his reanalysis of the New Jersey results. He showed much larger reductions in employment and earnings than those reported by the researchers who conducted the experiment. Although Cogan's methodology was debated among researchers, his reanalysis indicated a work-withdrawal effect as much as four times greater than that reported by Spiegelman of the Stanford Research Institute.

Cogan divided the treatment group according to whether people did or did not participate in the demonstration. In the case of the controls, he divided them in a similar way according to whether the members of the control group did or did not receive welfare benefits. Although many social scientists objected to Cogan's methodology and still do, it is notable that Moynihan, both in 1978 and in 1986, did not. At the hearing in 1978, Moynihan expressed indignation at Cogan's findings, remarking that the earlier reports on the New Jersey study were "bordering on malpractice" in light of Cogan's testimony.[18] He asked for, and received, agency comments on Cogan's work. Agency officials believed (at least this is what they told me) that Moynihan later accepted their reasoning as to why Cogan's reanalysis was flawed. However, the written record differs from what agency officials claimed. In his 1986 book, Moynihan described Cogan's testimony as follows: "The subcommittee, which is to say the general public, learned nothing until one afternoon in November 1978 when John Cogan, a young economist from Stanford, came to testify and told us, 'They won't tell you this, but it hasn't worked.'"[19]

The little drama was complete. The high hopes of supporters of demonstration research failed to materialize. Henry Aaron was right. The effects of the negative income tax demonstrations were perverse from the point of view of supporters of this approach to comprehensive welfare reform as advanced by Presidents Nixon and Carter. In retrospect, I believe the negative income tax demonstrations were moderately successful as research projects, although much less successful as an aid to policy-making. Their results came very late in the policy process and were at best ambiguous from the point of view of advocates of the negative income tax concept.

However, the experiments should not be judged solely on policy grounds. The purpose of research is to answer questions. Moynihan put it well in an exchange with Spiegelman of the Stanford Research Institute at the September 1978 hearings. He observed: "The bringing of systematic inquiry to bear on social issues is not an easy thing. There is no guarantee of pleasant and simple answers, but if you make a commitment to an experimental mode it seems to me — I am not enjoying this hearing one damn bit, but if you make a commitment to an experimental mode, something larger is at stake when you begin to have to deal with the results."[20] At its roots, the reason for the essentially negative findings of the negative income tax experiments involves the inexorable political arithmetic of this approach to welfare reform.

Policy researchers, although they learned a lot from these experiments, did a great deal of soul searching in the aftermath. Having participated in the appraisal (but not the launching) of the negative income tax experiments, I came away a skeptic. As I see it, the value of demonstrations in the field of social policy is very much a function of the type of programs being studied.

For me, the key distinction is between testing income-maintenance and service-type programs. I have reservations about the usefulness of demonstrations to test programs such as the negative income tax, where the establishment of such a program would be universal* and highly visible. After a national debate on such a policy change where it becomes widely known that the rules of the game have changed because a universal program has been adopted, there is every reason to expect that people "exposed" to the new program would change their behavior in ways that could not be known in a research environment in which such a debate and event had not taken place. You cannot test for such

* By "universal," I mean that if adopted a program would apply to everyone in the country who is eligible.

big-picture attitude and behavioral changes. On the other hand, service-type interventions are not as intrusive. A new service program (involving, say, an intensive job training or a special child care program) is not as likely to change attitudes and behavior in the society, because fewer people will be aware that such a new policy has been adopted.

Demonstration studies are expensive and time-consuming. They should be undertaken, as stated previously, in situations in which three conditions apply: (1) Politicians and program administrators are genuinely interested in the new policies or major new program departures to be tested; (2) they are uncertain as to how they will work; and (3) they are willing to wait for the results of a demonstration study. The negative income tax experiments did not satisfy the first of these conditions. The impetus of the demonstration came from the research community. To the credit of many of the researchers involved, the experiments did show that it is possible in the United States to conduct large-scale demonstration research projects with random assignment, and they provided valuable insights about program design and operations, but in my opinion that was not enough.

An additional important lesson of the negative income tax experiments relates to their disciplinary auspices. At a 1974 postmortem conference at the Brookings Institution on the New Jersey negative income tax experiments, sociologist Peter Rossi said he thought it was "paradoxical" that despite the heavy reliance of sociologists on primary data collection and the extensive use by psychologists of experimental designs, it was economists who "played the major role in designing and fielding the income maintenance experiments."[21] In a similar vein, sociologist Lee Rainwater complained that what was learned about social behavior in the negative income tax experiments was "remarkably skimpy." In a paper for a conference on the lessons of this research, Rainwater argued that more qualitative data were needed about the people in the experiment and its effect on their lives.[22] For such critics, the challenge involved is to devise ways to incorporate other variables — psychological, social, and political — into demonstration research. Among the kinds of questions one would like to have asked in the negative income tax demonstrations were: What were the effects on people's feelings of self-worth and achievement? What were its effects on children, on families, on communities with a concentration of poor families, on state and local governments, on public agencies, and on the society? I realize that not all such variables can be taken into account. Yet, trying to get at these kinds of questions is basic to the argument of this book about the missing links between social science disciplines in the conduct of applied research. If we leave out disciplines in demonstration research, we are likely to be neglecting important outcomes that are of importance to policymakers.

Other Income Maintenance Demonstrations

The negative income tax studies were the foremost income maintenance demonstration studies, but not the only ones. Other demonstrations have been conducted since the New Jersey flagship study was launched. Such research projects (some of them very large) were initiated, and much of this research was conducted, under the Nixon and Ford administrations. Besides the negative income tax demonstrations, income maintenance demonstrations were conducted to test housing allowances (vouchers to poor people to change their demand for housing and the supply of housing), health insurance for low-income families and individuals, and the use of education vouchers. By far the largest of these demonstrations were the housing voucher demonstrations.

The Housing and Urban Development Act of 1970 provided funding to test alternative approaches to converting existing housing programs of the federal government, which focused on producing new units, into housing allowances that subsidized low-income residents. The idea of a housing allowance is to influence the demand for housing by providing low-income people with an allowance that increases their purchasing power for housing. Most housing programs at the time supported the construction of low-income housing, thereby stimulating the supply side of the housing market.

Three housing-allowance demonstrations were conducted. The first was a demand-side demonstration in two sites (the Pittsburgh and Phoenix metropolitan areas), which studied how families would respond to housing allowances; the second was conducted in two smaller metropolitan areas to assess the effects of housing allowances on housing markets (their supply-side effects). The third was an administrative demonstration to collect information about the management of housing allowances: Could they be implemented effectively?[23] These demonstrations were carried out in the mid-1970s. The largest demonstration in terms of the number of participants was the housing-demand demonstration conducted by the Rand Corporation. Altogether $160 million was spent on this study. Half of the federal funds were used to pay for the tested programs and half for data collection and research. Some features of the housing-allowance approach have influenced housing policy; however, a number of the main questions raised by the demonstrations were never answered, and controversy still exists within the housing policy community about the design and usefulness of these demonstrations.

Another set of demonstrations, which began in 1974, was linked to proposals that Presidents Nixon and Ford advanced to establish a national health insurance system. (Ironically, Nixon's plan was similar to the plan unsuccessfully advanced by the Clinton administration two decades later.) The aim of the health insurance demonstrations was to answer questions that could not be "reliably resolved through analysis of non-experimental data."[24] The demonstrations, conducted in six sites over an eight-year period (November 1974 to January 1982), enrolled more than 7,000 people. The research, conducted by the Rand Corporation, tested a range of health insurance alternatives to determine the effects of different benefit structures and financial features on the utilization of services, the health status of the participants, and the type and quality of the care they received.[25] All things considered, these health insurance demonstrations were of questionable value and were very costly. As far as I know, the uses of this research were limited.

Another case in which a demonstration was used to study an income transfer was education vouchers. This was the most flawed of all the studies conducted of income maintenance programs. This demonstration emerged not under Lyndon Johnson's Great Society, but in the Nixon years, to test a favorite idea of University of Chicago economist Milton

Friedman. He argued then (and still does) that governments should promote competition in elementary and secondary education by providing families with vouchers to purchase educational services for their children. The demonstration study of this idea, funded initially by the Office of Economic Opportunity and later by the National Institute for Education, was developed by the Harvard University Center for the Study of Public Policy and was conducted by the Rand Corporation.[26] Although six school districts initially came forward as candidates to participate in this demonstration, three of them dropped out when they learned more about the terms. As it turned out, only one school district (Alum Rock in San Jose, California) received research funds. The Alum Rock school voucher demonstration lasted five years and cost $9 million. The story is long and complicated, but the plot is clear. Just about everything that could go wrong did go wrong. Not only did the project narrow down to one school district, but the State of California failed to pass the necessary enabling legislation, teachers and parents resisted essential features of the voucher plan, and parents were confused by what in the final analysis turned out to be a program with limited variation between the conventional and tested approaches. Although the idea of school vouchers later caught on, it was not because this research showed the way.

Demonstrations of Service-Type Programs

The other major category of demonstration research that has been conducted with random assignment is the study of service-type programs. Although a larger number of demonstration studies have been conducted of service-type programs than in the case of income maintenance programs, in the aggregate the amount of money involved is much less. Some service-type demonstrations predated the New Jersey negative income tax demonstration. The best known project was the Perry preschool demonstration in Ypsilanti, Michigan, initiated in 1962.[27] Despite the fact that the sample for the Perry preschool demonstration was very small (123 black children at risk of failing in school), the positive benefits of this program over a long period of time (through age nineteen in the fourth phase of this study as reported in 1984) were widely cited in the media and in the literature on education.

In considering service-type demonstrations in this book, which I argue is the most appropriate area for the application of randomized demonstration research, I rely heavily on the experience of the demonstrations I know best, those conducted by the Manpower Demonstration Research Corporation (MDRC). The Manpower Demonstration Research Corporation is a nonprofit intermediary corporation based in New York City.* It conducts applied social science research, using funds from governments and foundations. The studies conducted by MDRC focus on the most disadvantaged groups in society. They emphasize welfare policy, job training, and related social services. MDRC came into existence in 1974 to conduct the national supported work demonstration. The chapter that follows immediately examines the challenges involved in conducting demonstration studies. Chapter 5 then describes several demonstration research projects conducted by the Manpower Demonstration Research Corporation, including the supported work demonstration.

Endnotes

1 Steven V. Roberts, "House and Senate Conferees Settle on $4.6 Billion Employment Plan," *New York Times*, March 22, 1983, pp. 1, 9.

2 Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Chicago, IL: Rand McNally, 1979), p. 6.

3 Dependent variables are often also characteristics of the people treated.

4 Robert A. Levine, "How and Why the Experiments Came About," in *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment*, eds. Joseph A. Pechman and P. Michael Timpane, (Washington, DC: Brookings Institution, 1975), p. 16.

5 Ibid., p. 21. The delay of this announcement due to Congress's Labor Day Recess was especially ironic in this case.

6 Ibid., p. 25.

7 Felicity Skidmore, "Operational Design of the Experiments," in Pechman and Timpane, *Work Incentives*, p. 45.

* The term "intermediary" used in this way refers to organizations that have a linking function between research and public policy. The Manpower Demonstration Research Corporation was one of the first intermediary research organizations; there are now similar organizations in other areas of domestic public affairs.

8 Initially, the main urban portion of this next round of negative income tax experiments was planned for only one city, Seattle, but a decline in Washington State's volatile economy led to the decision to add Denver as a major urban site.

9 Gilbert Y. Steiner, "Reform Follows Reality: The Growth of Welfare," in *The Great Society: Lessons for the Future*, eds. Eli Ginzberg and Robert M. Solow (New York: Basic Books, 1974), pp. 47-65.

10 Julian E. Zelizer, *Taxing America: Wilbur D. Mills, Congress, and the States, 1945-1975* (Cambridge, MA: Cambridge University Press, 1998), p. 316. Mills at that time was considering running for president and, according to Zelizer, "did not want to end up on the losing side" on welfare reform.

11 Office of Economic Opportunity, "Preliminary Results from the New Jersey Graduated Work Incentive Experiment" (Washington, DC: OEO, February 18, 1970), pp. 10-11.

12 Daniel P. Moynihan, *The Politics of a Guaranteed Income: The Nixon Administration and the Family Assistance Plan* (New York: Random House, 1973) pp. 509-12.

13 Children's allowances are paid to all families with children according to the number of children and without regard to income. Many European countries have such schemes, with recoupment from higher-income families achieved through the tax system.

14 Vincent J. Burke and Vee Burke, *Nixon's Good Deed: Welfare Reform* (New York: Columbia University Press, 1974), p. 41. The Burkes' book is the best on this subject, despite the fact that it does not treat my role in a kindly way.

15 As quoted in Morton Hunt, *Profiles of Social Research: The Scientific Study of Human Interactions* (New York: Russell Sage Foundation, 1985), p. 292.

16 Gary Burtless and Robert Haveman, "Policy Lessons from Three Labor Market Experiments," in *Employment and Training R&D: Lessons Learned and Future Directions*, ed. R. Thayne Robson (Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1984), p. 108.

17 Daniel Patrick Moynihan, *Family and Nation* (New York: Harcourt Brace Jovanovich, 1986), p. 151.

18 Subcommittee on Public Assistance, Committee on Finance, U.S. Senate, Hearing: *Welfare Research and Experimentation*, November 15-17, 1978 (Washington, DC: U.S. Government Printing Office, 1978), pp. 96-97.

19 Moynihan, *Family and Nation*, p. 152.

20 Subcommittee on Public Assistance, *Welfare Research and Experimentation*, p. 82.

21 Peter H. Rossi, "A Critical View of the Analysis of the Nonlabor Force Responses," in Pechman and Timpane, *Work Incentives*, p. 151

22 Lee Rainwater, "Sociological Lessons from the Negative Income Tax Experiment: A Sociological View," in *Lessons from the Income Maintenance Experiments,* ed. Alicia H. Munnell, proceedings of a conference held in September 1986 (Boston, MA: Federal Reserve Bank of Boston, 1987), p. 194.

23 See Garland E. Allen, Jerry J. Fitts, and Evelyn S. Glatt, "The Experimental Housing Allowance Program," in *Do Housing Allowances Work?* eds. Katherine Bradbury and Anthony Downs (Washington, DC: Brookings Institution, 1981), p. 110.

24 Larry O. Orr, "The Health Insurance Study: Experimentation and Health Financing Policy," *Inquiry* 11 (March 1974): 28.

25 Robert H. Brook, John E. Ware, Jr., William Rogers, Emmett B. Keller, Allyson R. Davis, George A. Goldberg, Kathleen N. Lohr, Patricia Camp, and Joseph P. Newhouse, *The Effect of Coinsurance on the Health of Adults: Results from the Rand Health Insurance Experiment* (Santa Monica, CA: Rand Corporation, 1985).

26 U.S. National Institute of Education, Rand Corporation, *A Study of Alternatives in American Education*, 7 vols. (Santa Monica, CA: Rand Corporation, 1976-81).

27 See John R. Berrueta-Clement et al., *Changed Lives: The Effects of the Perry Preschool Program on Youths through Age 19* (Ypsilanti, MI: High Scope Press, 1984).

4

Hurdles of Demonstration Research

This chapter considers eight hurdles that must be cleared in conducting demonstration research to test possible new approaches to public problems. They involve: (1) selection bias; (2) the null hypothesis; (3) contamination; (4) relations with program operators; (5) the quality and consistency of the treatment being tested; (6) the cost and quality of the data used in demonstration studies; (7) the treatment of human subjects; and (8) the uncertainty of cost-benefit analysis as the final step in the demonstration research process. Taken together, these hurdles indicate the wide range of scientific and operational decisions involved in selecting the subjects for demonstration research, developing the design to be used, and executing such studies.

Selection Bias

Selection bias is the most important challenge in thinking about demonstration research. The aim of such studies is to compare the effects of a tested new program to the counterfactual state — that is, the situation that would have obtained without the treatment. It is necessary to make such a comparison in order to answer the bottom-line question: Did the tested treatment make a difference? And, furthermore: What kind of a difference did it make, and what was its magnitude? An example helps to show what is involved here.

Suppose state education officials are considering a new method to improve the reading proficiency of junior high school students who have reading problems. The new method is a computer-assisted reading-remediation program, which was tried on a pilot basis and seemed to be successful. Further assume that the program was sponsored by the state government, administered in a number of junior high schools, and that the decision to be made now is whether it should be replicated statewide. School administrators seek to compare the reading scores of students in the special program with the scores of students who were not in the program.

One way to do this is to compare changes in the scores of the students in the computer-assisted reading program with the average change in the reading scores for all junior high school students in the school districts in which the pilot program was conducted. But this may not be a good basis of comparison from the point of view of people who like the program. Assume, for example, that the average increase in reading scores for all students was 105 percent compared with 78 percent for the students in the pilot program. School officials may reject this "normal" achievement increase of 105 percent as the standard for comparison, arguing that the proper approach is to compare the results of the pilot program with the scores of other students who have reading problems. Still, would this be a satisfactory baseline for comparison?

Assume the pilot program was offered to all students with reading scores below a certain level in the schools in which it was conducted. Some students applied for the program, some didn't. Maybe what really mattered was the motivation of the students who came forward and the motivation of parents who wanted special help for their children. If we compare the reading scores of students in the program with the scores of apparently similar students who were not in the program, we may find a positive impact, but actually it may reflect the impact of the "M factor" of motivation. This would be unfair in the opposite direction; it would give too much credit to the pilot program. The point is not that motivation may have made the difference, but that we do not know what made the difference. This is what is involved in the idea of selection bias, the possibility that the results of a tested program are biased — consciously or unconsciously, deliberately or accidentally — because the people selected for the program were special in some way.

This is where random assignment has its great strength. If we had tested the reading program on a basis in which eligible students were randomly assigned (like a lottery) to either a treatment group or a control group, we would be in a better position (not perfect, but better) to say what difference the pilot program made. Random assignment allows school officials to control for a bias that might be introduced either by teachers or parents in selecting the students to participate in the new program from among those who met the program's eligibility standards. This applies either to a deliberate bias, for example, a "T factor," because teachers selected their best or worst students, or the "M factor" whereby students came forward because they were motivated to do so, or because they came from a home environment in which family support and encouragement had more of an effect than the special program being tested. Random assignment solves the problems of T factors, M factors, and X and Y factors because it creates conditions under which, if the sample is large enough and properly drawn, there is just as much chance that a T- or M-factor or other factor student will be in the control group as in the treatment group.

There is little dispute among researchers that random assignment is the best way to deal with the problem of selection bias; however, it is an expensive and often difficult procedure to use. A key question therefore is whether we can find an acceptable alternative when for some reason relating to cost, feasibility, the time frame involved, or the ethics of a given research setting, random assignment is deemed not to be possible. Researchers often use statistical modeling procedures to attempt to control for selection bias when random assignment is not used, but in the opinion of most experts this alternative is not as good as random assignment.

The history of the Manpower Demonstration Research Corporation's work is helpful in making this point. The most striking finding from the MDRC supported work demonstration mentioned earlier was that this program showed positive results for two of four tested participant groups.* Although the participants in each of the four groups were randomly assigned to a treatment group or a control group, the results were by far the strongest for one group — long-term female family heads on welfare. While it gets a little ahead of the story (this research project is described in more detail in the next chapter), it works well here to examine how one expert used this demonstration study to devise an ingenious approach to test the efficacy of random assignment.

Using data from MDRC's supported work demonstration, labor economist Robert J. LaLonde studied whether a research approach other than random assignment could have replicated the supported work control group for female welfare family heads. If LaLonde could have used a statistical modeling technique to identify a comparison group that was just like the supported work control group for long-term welfare family heads, or even very similar to this group, then researchers could have saved themselves a lot of time and expense. They could have used statistical modeling to create a comparison group and measured the impact of the supported work program by comparing the outcomes for the people in the tested program with the people in the statistically simulated comparison group.

According to LaLonde, when researchers do not have a randomly selected control group, "an econometrician must first select a group of individuals from the population to serve as a comparison group and then specify an econometric model that accounts both for the difference in earnings between the treatment and the comparison groups and for the treatment groups' decision to participate in training. . . . MDRC's experimental data offer labor economists an opportunity to test the non-experimental methods of program evaluation."[1] LaLonde used econometric techniques and data from three sources (the University of Michigan's Panel Study on Income Dynamics, the U.S. Bureau of the Census, and the Social Security Administration) to compute the

* The supported work approach tested a way of helping four groups of people with limited employment experience and skills enter the regular labor market. It consisted of a series of gradual steps — called "the graduated-stress approach" — that would get them acclimated to, and prepared for, regular employment. The four treatment groups were long-term female welfare family heads, ex-addicts, ex-offenders, and problem youth.

earnings of a group of people like the female welfare family heads in the MDRC supported work program.

In short, LaLonde's experiment on experimentation did not work. He concluded: "The econometric models used to evaluate training programs generate imprecise estimates of training effects. This imprecision underscores the importance of a classical experimental design both to the evaluation of the national supported work program and perhaps other programs as well. Without random assignment an econometrician faces a considerable range of training effects; it is unlikely he will choose the correct one."[2] The MDRC supported work demonstration showed net additional earnings of $851 per year for female family heads on welfare, the group most aided by the supported work program being tested. LaLonde compared this outcome with the earnings of four simulated comparison groups of poor women. The result for one of his four groups was close to the findings from the MDRC study; it showed a net gain for program participants of $1,090. A second simulation was also positive, but exaggerated the benefits for supported work, showing an earnings gain of over $3,000. Two other simulations showed negative results. The women in these simulated comparison groups earned less than participants in the demonstration. In one case the negative earnings difference was $2,822; in the other, $3,357.

The clinching argument for LaLonde on the desirability of using random assignment in demonstration research was that he could find no basis on which to know how to select the right comparison group from among these possible simulated comparison groups. Labor economist Orley Ashenfelter, head of the Princeton University Industrial Relations Section, of which LaLonde was a member, commented as follows on his study: "The evaluation of the economic benefits of training programs will be greatly enhanced by the use of classical experimental methods. . . . Much of the non-experimental estimation of the effects of training programs seems dependent on elements of model specification that cannot be subjected to powerful statistical tests. . . . In sum, it appears that in the area of analysis of training programs the econometric methods available may not be able to deliver the benefits that randomized trials offer."[3]

LaLonde's work stimulated other researchers to examine alternatives to random assignment. Labor economist Rebecca Maynard, who had a long-standing interest in this subject, co-authored an article with LaLonde supporting his skeptical view of alternative methodologies. Later, a special issue of the journal *Evaluation Review* (August 1987) was devoted to this subject, including the LaLonde-Maynard article and a contrary position advanced by James J. Heckman. Heckman's article, written with associates at the University of Chicago, expressed his belief that "reliable non-experimental evaluation methods can and will be developed in the future for all subsidized employment and training programs."[4]

As previously stated, I believe random assignment is the best methodology for demonstration research. It was already observed that researchers who conduct demonstration research, especially in the field of social policy, often test programs they favor. One can think of this as a potential problem of selection bias — selecting programs they like — on the part of the researchers. The best way to deal with this problem is to use random assignment. Once researchers have designed a demonstration research project with random assignment, it greatly limits their opportunities (consciously or unconsciously) to manipulate their data.

Nevertheless, random assignment, because of its high cost and the difficulties involved in using it, is not the only way demonstration research is conducted. Researchers sometimes compare people in other places with those in the places in which a demonstration is carried out. This is called the "comparison-site" approach. Also, researchers sometimes use a before-and-after design to assess a program. This approach is especially flawed because we don't know if an observed effect would have occurred anyway over time. In still other studies, people who fit the eligibility standards of a particular program, but for some reason did not participate in it, are used as the comparison group. They may be applicants who initially came forward but eventually did not apply, or they may be applicants for whom space was not available. In other studies (and this, as already mentioned, is the most common alternative approach used), econometric techniques are applied to data sets as LaLonde did statistically to simulate a control group.

My view is that some of these "less-good" approaches (that is, less good than random assignment) are better than others. The research hurdle of selection bias is high and crucial. However, I do not believe it justifies the rigid position, which unfortunately is widely held by public policy researchers, that there are no acceptable alternatives to random assignment for demonstration research. For now, suffice to say that the problem of selection bias is important enough to make random assignment the best methodology for demonstration research.

The Null Hypothesis

The second hurdle for demonstration research is the "null hypothesis." This has to do with the choice of programs to test, not the choice of the method to test them. The key point here is that in real world settings in which many factors and forces impinge in rapid fire fashion on people's lives, the only kind of program one can test is one that is big enough to make a detectable difference.

Again, the history of the Manpower Demonstration Research Corporation's supported work demonstration helps to explain what is involved here. When MDRC was designing the supported work demonstration, Robert Lampman, a welfare economist at the University of Wisconsin and an MDRC board member, stressed that this intervention must be large enough and last long enough so that it could reasonably be expected to have a significant impact. Lampman wanted to be sure we did not load the deck in favor of the null — or no-effect — hypothesis. Lampman said: "When persons with severe employment handicaps and disabilities are singled out for remediation, positive and lasting effects are not likely. In the case of supported work, the odds in favor of the null hypothesis were even greater . . . since the four groups chosen were from among those least likely to succeed in the labor market."[5] Lampman was right. If funders are going to the expense of conducting a demonstration research project with random assignment, such studies should be of policy interventions that can be expected to make a discernible difference in the lives of the people treated.

Contamination

Contamination is the third hurdle of demonstration research. In high school chemistry classes, students are often told about the need to avoid contamination by conducting experiments in an airtight chamber to avoid having contaminants in the air impair its results. In demonstration research, we use the word "contamination" in a way that some people may find objectionable. If, for example, we are testing a program to provide health care for infants, on scientific grounds we would like the treatment and control groups to be "pure" in the sense that one group gets the treatment and the other does not. But what if we are studying such a program in a particular community, and while the study is under way a local group decides that it should open a clinic to provide similar health care services for infants? If we are using a research design with random assignment, the result might be that the treatment would have no or a small impact because the infants in the control group are receiving services similar or identical to those provided to the infants in the treatment group. Occurrences like this are not unusual. Under such conditions, we are comparing apples with apples. We are comparing the treatment that is the subject of the demonstration with other treatments provided in the community in which the demonstration is being conducted.

In another of MDRC's studies, this issue came into play in an important way. The case involved Project Redirection, a program to provide integrated, intensive social services to young welfare mothers. (The treatment group can be thought of as "children with children.") The research design in this case was a quasi-experimental approach using comparison sites. What MDRC found in this case was that the impacts of the program were "mixed but disappointing." Overall, its report said, "the early benefits proved to be largely transitory: By twenty-four months after baseline, most had disappeared."[6] The apparent reason for this was that the people in the comparison sites received services like parenting classes, medical care for the baby, birth control counseling, educational counseling, and employment counseling — similar to those provided to the participants in Project Redirection. The MDRC report said: ". . . the comparison teens received many more services than had been anticipated. For example, 43 percent of these teens were enrolled in a special teen parent program after their entry into the research sample and during Redirection's first operational period."[7]

This situation is referred to as contamination, although this use of the term should not imply that there is something wrong. It should be noted that this problem was not a function of the decision to use comparison sites. Random assignment was found to be infeasible for political reasons. The participating organizations in the demonstration were small. MDRC staff members faced resistance from program administrators on the grounds that they did not have a large enough pool of applicants to select randomly for treatment one out of every two (or some other proportion) of applicants. Actually, program operators often have a different point of view from researchers about research methods. In any event, the contamination that occurred in the case of Project Redirection could just as easily have occurred with a research design that used random assignment.

Relations with Program Operators

The hurdles discussed in this chapter indicate the kinds of research compromises that often are necessary in demonstration research because of the complexity of the real world. In the case of MDRC's Project Redirection, the reason for using a comparison-site design, as opposed to random assignment, was that program operators objected to random assignment. In discussions with MDRC staff, they maintained that the limited size of the pool of eligible young welfare mothers who could be expected to come forward would not allow them to run a program that required them to turn away some applicants. Besides, they did not want to do this. It is not unusual for program operators to resist having to assign needy people who are otherwise eligible for a social service they provide to a no-treatment group. The challenge to researchers under these conditions is to convince the operators of a program to be tested that they should participate in research that employs random assignment. This can require extensive and complicated negotiations. MDRC staff members are very good at this, but they don't always succeed. Program operators in this situation are being asked systematically to reject eligible applicants who, they argue, are *people*, not research objects.

Researchers have two main types of leverage under these conditions. One is money. Typically, the sponsors of a demonstration pay for the tested program. This may influence program operators to join a demonstration project despite their misgivings. Their reasoning in this case could be that they are getting something extra for people they want to serve in a situation in which they could not otherwise afford this service. Indeed, the rationing of this service on a random-assignment basis can be viewed as justifiable. One can argue that random assignment, like a lottery, is the fairest method of rationing a service under conditions of limited resources. The second argument that the managers of a demonstration can bring to bear in this situation involves the case for research. If the people who provide a service can help to prove that it works, then that service eventually may be provided on a broader basis to more people who need it because of the effect the research findings have on policymakers and on the political process.

Quality and Consistency of Treatment

Once a research purpose and design are decided upon and a sponsor agrees to pay for a demonstration study, other critical issues come into play. Consider, for example, the case in which a preschool education program being tested in a demonstration project is not being provided in the way the researchers intended. It may be because the staff of the organization providing the service do not like the research specifications, or it may be because they simply do not perform well. What is a researcher to do under such conditions?

We have what might be called, "the repair dilemma." Should researchers see to it that the quality of the program is maintained at a high level? In this case, researchers would be taking the position that what they are testing is a treatment that is well administered. This issue arose in one of MDRC's demonstrations, the Youth Incentive Entitlement Pilot Project. This demonstration was mandated by Congress to test a "saturation" program in selected communities designed to provide employment to all disadvantaged youth, both those in school and school dropouts, on the condition that the participating youth remained in or returned to school. Over a period of two-and-a-half years, a huge sum ($240 million) was spent on this demonstration study in seventeen communities in which 76,000 persons were program participants. The U.S. General Accounting Office reported that this demonstration was well managed by MDRC, but they said the demonstration was artificial.[8] GAO said MDRC's management oversight was too intrusive. The argument GAO made was that in the conduct of this program, were it to be replicated on a broad basis, the responsible authorities (mainly states and localities) could not be expected to be as demanding and rigorous in maintaining the quality of the program as was MDRC. My view is that this is one research hurdle we can't clear, namely to improve the standards of ongoing governmental program management. The main lesson of this experience is that it points up the challenge of maintaining acceptable standards with respect to the character and consistency of the treatment in a demonstration study.

Cost and Quality of Data

Almost any large-scale demonstration project is bound to confront mundane but crucial problems involving the data needed to conduct the research. Data must be collected in three time periods for both the treatment group and a control or comparison group in a demonstration: (1) a baseline period before the program starts up; (2) the in-program period (that is, the period during the demonstration); and (3) a post-program period. This third period, after the tested treatment is administered, can last a long time, sometimes five to seven years, sometimes longer, as in the case of the Perry preschool demonstration described in chapter 3. The data needed for both program participants and controls often must be obtained from a number of sources, for example, program operators, surveys, and governmental records and statistics.

Data collection tasks associated with a demonstration study are easiest to carry out for the people in the treatment group while they are participating in the tested program. But even this can be difficult. Researchers and program operators often do not see eye to eye. Conflicts can arise about what program sponsors regard as intrusive data requirements. Other problems can also arise. An example is the requirement in research protocols for program sponsors to collect data for the people in the research sample who are selected as participants but do not show up and participate in the program being tested.

However, compared to the in-program period, it is more difficult to obtain data for participants *after* they leave a program. Program operators usually cannot be enlisted to administer questionnaires to this group, so it is often necessary to employ a survey firm to obtain information from participants in the post-program period. Surveys are expensive, and if former participants cannot be located, the success of a demonstration may be threatened.

Still, all of the problems of collecting data on the participants engaged in a demonstration project pale in significance in comparison to those involved in collecting data on the people in control or comparison groups, people who were not in the tested program at all. Data for them are needed for all three time periods — baseline, program, and post-program. Often, the members of control groups are paid when they are interviewed, but this is not always or necessarily enough to overcome the problems of finding them and winning their cooperation. The people in control or comparison groups are likely to be highly mobile and hard to locate; they may also have limited language skills, which can add to the difficulty of obtaining information about them. Compromises are often made. Data elements that are considered desirable are dropped in order to reduce the time needed to conduct interviews and the costs involved in doing so. In other cases, the frequency of data collection is sacrificed to cut costs or to limit the intrusiveness of the research process. In still other cases, public data files are used instead of interviews in the follow-up period, in which case the data may be less complete and accurate than in situations where a survey is administered. Wherever compromises are made, questions arise about the effect such decisions have on the integrity of the research. These nitty-gritty issues are not the kinds of challenging issues that fascinate social scientists. They are nonetheless crucial in the conduct of demonstration research.

Treatment of Human Subjects

As demonstration research gained ground in the U.S. in the late 1960s and 1970s, so did concerns about the protection of human subjects in such studies. In 1974, the U.S. Department of Health, Education, and Welfare issued regulations that required researchers receiving HEW research funds to establish "institutional review boards" to ensure that demonstration research projects protect the rights of human subjects.[9] At first, the President's Office of Management and Budget required that these regulations be applied on a government-wide basis. Later, this policy was changed, and the regulations were applied only to demonstrations (and even then not all demonstrations) paid for by HEW or its successor department, the Department of Health and Human Services.

Two main premises are reflected in the HEW rules for dealing with human subjects and in the similar standards of other groups. Both premises are derived from medical research. One premise is that social program operators, like medical practitioners, should *do no harm*. They should make certain that no one is disadvantaged by virtue of being a member of a treatment or control group for a demonstration study. The second premise is positive, requiring that researchers obtain *informed consent*, that is, that participants in a demonstration study (both those treated and the members of a control group) should be given an explanation about the research and should be asked if they are willing to agree to participate in it. Writing in the mid-1970s, P. Michael Timpane and Alice Rivlin said that "informed consent is by now an entrenched canon of medical experimentation and has been adopted implicitly by most social experimenters."[10] Although few policy researchers quarrel with these premises, their implications for both the substance and process of demonstration studies are considerable.

The Brookings Institution held a conference in 1975 on the ethical and legal issues of social experimentation. The conference considered a number of questions: Can children be the subject of a social experiment? Who can give informed consent for them? Their parents? School officials? Some participants at the conference ruled out all social experiments involving children, along with other groups for whom it was argued informed consent cannot be obtained effectively, such as prisoners and mental patients. This view was challenged by other participants at the conference who saw little risk, only potential gains, for the subjects of social experiments. They argued that all decisions involving social programs involve risks similar to those of a social experiment, yet we do not require informed consent in every case in which society intervenes in a life situation.

Despite the fact that the requirement to obtain informed consent is now widely accepted by researchers, questions of execution (how researchers tell potential participants about a research project and how much they tell them) can introduce problems. For example, people may react to a social experiment in a way that reflects what they perceive to be the expectations of its sponsors as expressed in what they are told in obtaining their informed consent. They may try to "prove" that a program works regardless of whether they are in the treatment group or the control group, thus confounding the research. Or members of the treatment group may decide to undermine the program for reasons having to do with their attitude toward what researchers tell them, toward program operators, or toward social programs generally. This is called the "Hawthorne effect," people reacting simply to being the subjects of an experiment.*

Although issues having to do with obtaining informed consent are important, the most serious ethical issues concerning the treatment of human subjects come after informed consent has been obtained. Suppose we are testing a home health care program for the elderly with a

* The Hawthorne effect occurs when behavior is altered because people know they are being observed. Its name derives from studies at the Hawthorne Works of the Western Electric Company during the 1920s and 1930s. The studies showed productivity increased when lights were darkened and rest periods shortened, which researchers attributed to the extra effort made by workers who knew they were part of an experiment. See Theodore H. Poister, *Public Program Analysis: Applied Research Methods* (Baltimore, MD: University Park Press, 1970), pp. 266-67.

demonstration using random assignment. Assume the service is expensive and that the key dependent variable of interest to the funders and the researchers is whether the service over the long run reduces the costs of institutionalization in a nursing home or hospital. An older person comes into a senior citizen center and finds that she is eligible for "Home-Help." She is urged to sign up but she is told that she may or may not be one of the participants selected. She decides to apply and fills out the necessary application and certification forms and also an informed consent agreement. The program director then checks by phone with the research staff and is told that this particular participant is assigned to the control group. The director is face to face with the applicant and must tell her that she was not selected for the program. Assume the applicant asks, "Is there another program I can get into?" This raises a troublesome, yet not unusual, issue. The program director knows that if the applicant is referred to another program and is aided, this could reduce the likelihood that the program that is the subject of this demonstration will have an impact, because the difference between the treatment and control group is likely to be less than what it might otherwise be. Moreover, showing that there is a big difference between the two groups often is precisely the reason the program director agreed in the first place to join the research project.

This is the "program director's dilemma." The director is between a rock and a hard place. We can make this dilemma even more difficult if we assume that the program director is sitting across the desk from an applicant who does not ask about another program, yet the director knows about one that exists and has openings. Should she make a referral even though it wasn't solicited? There are no easy answers to these questions.

The Uncertainty of Cost-Benefit Analysis

The eighth hurdle of demonstration research concerns cost-benefit analysis, appropriately treated here as the final hurdle because it is the last step in the demonstration research process. In this final step, information about all of the costs of a program being tested and all of the benefits are combined in an effort to come to a bottom-line number, which identifies the full and long-term impact of the program. Researchers typically present their cost-benefit findings according to three perspectives — to participants (were they better off?), to taxpayers (that is, nonparticipants as a group), and to the society as a whole. Often, big assumptions are needed to do cost-benefit analysis.

The aim of a cost-benefit analysis is appealing. In an ideal world, one would want policymakers to be in a position to compare the cost-to-benefit ratios for all, or anyway a number of, program alternatives designed to achieve a particular policy objective, and then select the most effective one. Cut down to size, this was the aim of Lyndon Johnson's Planning-Programming-Budgeting (PPB) system described in chapter 2. But, as we have seen, the task of conducting demonstration research is so complicated, time-consuming, and expensive that we can only study the effects of those few programs where we decide circumstances warrant the conduct of rigorous research. No matter how good our intentions may be to use social science in making social policy, decision makers will never be able to choose among all possible government programs to achieve a given purpose on the basis of definitive cost-benefit findings.

Where demonstration studies are carried out, the immediate purpose is to determine whether the tested program had a sufficiently strong measurable impact when it was tested to justify its replication on a broader basis. Unfortunately, this is not enough for policymakers who want to know the long-term costs and benefits of the program.

My view is that this step often involves problems that go beyond what researchers can do in a demonstration study. Frequently they do not have sufficient data for the far out years — that is, projected way beyond the study period. Attempts to ascertain the cost-benefit ratio of a tested program often require going beyond the time period of observed variables in a demonstration project.

It is wise to be cautious about cost-benefit analysis, and indeed many governmental and foundation research sponsors take this position. Cost-benefit analysis is an uncertain art form. Policy researchers and the sponsors of policy research are best advised to stop at the point of ascertaining the observed impacts of the tested program and let the political process take it from there.[11] There are two reasons for this conclusion: (1) that there are likely to be costs and benefits of a social demonstration that are left out of the cost-benefit equation, or are included with such rough approximations as to be of questionable value; and (2) that the methods for estimating the effects of a tested program way beyond the study period are often highly uncertain.

On the first point, the problem is that outcome variables that are omitted from a demonstration often are left out precisely because they are difficult to measure. In the case of the "Home-Help" program for the elderly mentioned earlier, we may decide that the fact that participants are happier in their own home is a very important benefit. Yet, how would we monetize this variable for a cost-benefit analysis, which some readers of research results are likely to think should be or is part of the analysis? This point can be shown using MDRC's supported work demonstration as an illustration. Policymakers may be interested in the effects of supported work on distressed communities. They may believe that in the long run a positive and cumulative community effect is desirable for the society, and they may also believe that over time this will result in lowering welfare costs. However, we would be hard put to measure these community-improvement and long-term welfare cost-reduction outcomes of supported work. The same holds for the effects of supported work on the children of program participants: Are they better off, happier, more or less successful in school? These dependent variables (effects on communities, the society, and children) were not included in the design for the MDRC supported work demonstration. But, the idea of a comprehensive cost-benefit analysis that purports to show the whole picture may suggest to some readers that factors such as these were included.

Other types of benefits are often included in a cost-benefit analysis that are very hard to measure. A good example is the cost-benefit study that was conducted of the Job Corps. Researchers reported a favorable cost-benefit ratio for the society as a whole of $1:1.46 — that is, $1.46 in benefits for every $1.00 spent on the Job Corps. This was regarded as a very good result. However, 40 percent of these benefits came from reductions in criminal behavior; accordingly, the values assigned to reduced injuries and loss of life due to reduced criminal behavior had a very large effect on the cost-benefit findings. A murder was estimated to cost society $125,305.[12] The cost-benefit ratio of the Job Corps would have been negative (.8:1) if crime reduction had not been considered.[13]

Long-term impacts also can be hard to assess. As already noted, the follow-up data collected in a demonstration study often do not cover as long a time period as is believed to be needed for a cost-benefit analysis. Because of this, researchers sometimes make assumptions about what are known as the "decay rates" of program benefits over time. They measure the benefits for a given period and then estimate or assume how they will be sustained or taper off in future years. The literature on this subject is complicated, but the point is that decay rates are extremely difficult to measure. This is another soft ingredient that goes into some cost-benefit analyses.

In sum, the task of assigning a single overall cost-benefit ratio to programs tested in demonstration studies often gets researchers into difficult terrain. This is not to say that costs and benefits should be ignored, but only that public policy researchers should focus their attention in demonstration studies on measured impacts. They should tell policymakers what they measured, how and for how long they did so, and what they found. This knowledge can be combined by policy analysts with other types of information, including what is known or believed about longer-term program effects. But this should be done in a more modest way than is the practice in cost-benefit analyses, often performed with high levels of precision, as the final step in the conduct of social experiments.[14] Some readers may acknowledge the validity of these admonitions, but take the position that politicians want — some even demand — one clear, simple, bottom-line number. This is true. But I would argue that scientific considerations should outweigh these political exigencies. Candor is required on the part of researchers in their discussions with policymakers about the strengths and weaknesses of the findings from demonstration research.

Since I rely heavily on the research conducted by MDRC, I feel obliged to discuss MDRC's practice in the use of cost-benefit analysis.

In most of its studies, the MDRC has presented cost-benefit findings, but with considerable care and soul searching. When the corporation issued its summary report on the national supported work demonstration, its first demonstration study, a carefully crafted statement about the limitations of cost-benefit analysis was included:

> Although the approach is useful in providing an overall assessment of supported work's effectiveness, it has limitations and risks. Certain important benefits and costs simply cannot be accurately measured and are therefore not included in the summary estimates. Moreover, this type of analysis calls for assumptions about the value of specific items and for judgments on the longer term extrapolation of benefits and costs that were directly measured only for up to 27 or 36 months.[15]

Perhaps MDRC should have stopped there and not presented an overall cost-benefit analysis. Labor economist Lloyd Ullman argues that the uncertainty of cost-benefit analysis produces an "expert witness" mentality whereby the proponents and opponents of a given program can trot out experts to argue both sides on a basis that often undermines the value of a well conducted demonstration.[16] At the very least, researchers should express cost-benefit findings using ranges of numbers for their findings, and also should present tables and text that clearly and prominently indicate qualifications.

Conclusion

These eight hurdles to demonstration research point to the same conclusion: Rigorous research on what works can be a useful and important input to the policy process, but it is not easy to do. Understandably, politicians and policy makers are bound to ask: Does a given new program idea work? Demonstration studies are most effective when policymakers care about a particular issue, are genuinely uncertain about how to handle it, and are willing to wait for the findings of a research project.

It needs to be added that the demonstration research has an important byproduct: *It teaches.* The history of MDRC is a success story in these terms. It has deepened understanding of social programs — what they are, how they work, and how they should be assessed. This is an important outcome in and of itself. The next chapter carries the MDRC story further and shows how the results of demonstration studies can be and are used in the policy process.

Endnotes

1 Robert J. LaLonde, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," Working Paper No. 183, Industrial Relations Section, Princeton University, October 1984, p. 183.

2 Ibid., p. 67.

3 Orley Ashenfelter, "The Case for Evaluating Training Programs with Randomized Trials," Working Paper No. 203, Industrial Relations Section, Princeton University, January 1986, pp. 1, 8.

4 James J. Heckman, V. Joseph Hotz, and Marcelo Dabos, "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?" *Evaluation Review* 11 (August 1987): 395-427.

5 Robert Lampman, as quoted in Eli Ginzberg, Richard Nathan, and Robert Solow, "The Lesson of the Supported Work Demonstration," in *The National Supported Work Demonstration*, eds. Robinson G. Hollister, Jr., Peter Kemper, and Rebecca A. Maynard (Madison, WI: University of Wisconsin Press, 1984), p. 308.

6 Janet C. Quint and James A. Riccio, *The Challenge of Serving Pregnant and Parenting Teens: Lessons from Project Redirection* (New York: Manpower Demonstration Research Corporation, April 1985), p. 15. New MDRC follow-up data for five years indicate positive longer-term impacts of Project Redirection.

7 Ibid., p. 16.

8 U.S. General Accounting Office, "CETA Demonstration Provides Lessons on Implementing Youth Programs," *Report to the Congress by the Comptroller General of the United States*, HRD-81-1, 8 December 1980, p. i.

9 See Marshall J. Breger, "Randomized Social Experiments and the Law," in *Solutions to Ethical and Legal Problems in Social Research*, eds. Robert F. Boruch and Joe S. Cecil (New York: Academic Press, 1983), pp. 106-7.

10 See Alice M. Rivlin and P. Michael Timpane, "Introduction and Summary," in *Ethical and Legal Issues of Social Experimentation*, eds. Alice M. Rivlin and P. Michael Timpane (Washington, DC: Brookings Institution, 1975), p. 7, 15.

11 I am indebted to my MDRC board colleague Eli Ginzberg for having patiently planted the seeds of doubt expressed in this analysis, although many readers will quarrel with the position taken here. Lloyd Ullman also helped me in working through these issues. Neither Ginzberg nor Ullman are responsible, however, for the statements made in this section.

12 Office of Research and Evaluation, Employment and Training Administration, U.S. Department of Labor, "Evaluation of the Economic Impact of the Job Corps Program," September 1982, p. 230.

13 Michael E. Borus, "Why Do We Keep Inventing Square Wheels? What We Know and Don't Know about Remedial Employment and Training Programs for High School Drop Outs," paper prepared for the Manpower Demonstration Research Corporation, New York City, June 1984, p. 9.

14 While it is not my area of expertise, it would seem that for physical (public works-type) policy studies, cost-benefit analysis is more feasible.

15 Board of Directors, Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported Work Demonstration* (Cambridge, MA: Ballinger, 1980), p. 135.

16 I am indebted to Lloyd Ullman for suggesting the expert witness metaphor.

5

Welfare Demonstration Studies

This chapter uses studies conducted by the Manpower Demonstration Research Corporation to show how demonstration research affects public policy. The way the corporation developed was very much a function of decisions made about the conduct of the original supported work demonstration. The supported work program was developed by the Vera Institute of Justice in New York City to provide employment for people accused of minor offenses in the criminal justice system. The aim was to provide participants with work experience in a group situation, under conditions of gradually increasing stress, as an alternative to incarceration. An operating subsidiary of the Vera Institution, the Wildcat Service Corporation, was established in 1969 to run the supported work program; its roster grew from six participants in the fall of 1969 to nearly 1,400 in mid-1974. Participants worked in jobs such as cleaning, plastering, and painting buildings being restored; clearing refuse from vacant lots; serving as messengers; providing building- and park-maintenance services; clearing construction sites; interpreting for Spanish-speaking hospital patients; and driving for the elderly.

Early studies of the Vera Institute's supported work program showed promising results.[1] As a consequence, Mitchell Sviridoff, vice-president for national affairs of the Ford Foundation, decided to set up a national test of the supported work idea. Sviridoff sought assistance from a number of federal agencies and established a research committee, on which I served, to advise the Ford Foundation. The advisory committee, headed by economist Eli Ginzberg of Columbia University, recommended a multi-site national demonstration with participants randomly assigned either to a supported work program or to a control group. Six federal agencies, with the Department of Labor as the lead agency, committed funds to this research, which began in March 1975. Over the full three-year period of the supported work demonstration, 6,616 people were in the research sample — 3,214 as participants and 3,402 in the control group.

As planning for this national demonstration moved forward, it became apparent that the management tasks involved were formidable. The list of participating organizations was growing — the Ford Foundation, six federal agencies, the fifteen sites operating the program, the research contractors, plus a central staff to coordinate and manage the demonstration. By mid-1974, the total cost of the demonstration, including funds from local sources, was estimated at $80 million, most of which was to be spent on participants.

The assumption up until this time had been that an interagency committee of federal officials would administer the supported work demonstration. However as the planning proceeded, it became increasingly clear that an interagency committee would have great difficulty managing such a multi-pronged, dispersed project. Hence, it was decided in mid-1974 to convert the advisory committee for the study into a nonprofit corporation, chartered in Delaware, which would receive funds from federal agencies and the Ford Foundation to manage the supported work demonstration. Because the intermediary corporation formed for the supported work demonstration proved to be an effective device, it has continued in existence (now in its 26th year) to conduct many other large demonstration studies.

I was an original member of the corporate board, as were Eli Ginzberg, Robert Lampman, Robert Solow, Gilbert Steiner, and Phyllis Wallace (all academic social scientists). Ginzberg chaired the board from 1974 to 1981, when I succeeded him. I served as chairman until 1997. The first president of MDRC was William G. Grinker. He was followed in 1982 by Barbara B. Blum, who served until 1986, when she was succeeded by Judith M. Gueron, who has been an officer of the corporation from its inception. MDRC has had an increasingly diverse research portfolio and a staff in some years as large as 200 people.

In addition to the supported work demonstration, other demonstration studies conducted by MDRC include: (1) a study of the tenant management approach for administering public housing projects; (2) a saturation guaranteed-job program known as the Youth Incentive Entitlement Pilot Project, which was established by federal law as a demonstration project; (3) Project Redirection to provide services to very young women receiving welfare benefits; (4) an employment and training program (called the WIN-Laboratory or WIN-Lab project[2]) conducted in Denver; St. Louis; Madison and Racine, Wisconsin; and Louisville, Kentucky; (5) the Structured Training and Employment Transition Services (STETS) demonstration (a program similar to supported work) for young mentally retarded workers; (6) an eight-state test of work/welfare as an approach to welfare reform in the early 1980s; (7) an intensive remediation and training program for school dropouts called Jobstart; (8) New Chance, a comprehensive service program for very young welfare mothers; (9) the Parents' Fair Share Demonstration project for noncustodial welfare family heads (mostly fathers); (10) a study of a large employment subsidy program in Canada; and (11) a study of the effects of welfare reform in distressed neighborhoods in four large urban counties after the enactment of the 1996 national welfare reform law.

As the Manpower Demonstration Research Corporation gained experience, it moved from relatively small and focused demonstration projects to broad-scale evaluations of large ongoing programs. Examples of the latter are the corporation's evaluation of the statewide work/welfare program in California and its "Urban Change" study of the effects of the 1996 national welfare reform law on distressed neighborhoods in large urban counties. The discussion of MDRC's experience in this chapter highlights the corporation's demonstration research. Later, in chapter 7, the discussion considers MDRC's evaluations of large ongoing social programs.

Supported Work

Since the early 1970s, welfare reform for able-bodied, working-age adults and their children has been the Mount Everest of U.S. social policy — climb it if you can! Two different approaches to welfare reform have been championed. One is the *income strategy* as embodied in the negative income tax idea discussed in chapter 3. (The aim of this approach is to structure welfare benefits so they provide an incentive for working-age, able-bodied parents to enter and stay in the labor force.) The other major approach is the *service strategy.* It emphasizes services, such as job placement and counseling, child care, transportation, and training and education to enable poor family heads to move off welfare and enter and stay in the regular labor force. The two approaches can go together, but the emphasis of different plans advanced or tested tend to be very different.

In 1962 under President Kennedy, the federal government first enacted legislation embodying the service strategy — to provide job counseling, job training, and related rehabilitative services to welfare family heads. Then, toward the end of the 1960s, the emphasis of national welfare policy-making shifted to the income strategy. The negative income tax demonstrations were initiated in 1968; and in 1969 President Nixon proposed the Family Assistance Plan (FAP) for welfare reform. Although Nixon's proposal was a hybrid of the income and service strategies, the newest and most controversial features of his plan incorporated the negative income tax approach. Nixon's plan died in the Congress in 1972 just before his re-election as President. Because of Watergate, it could not be revived afterwards, despite efforts to do so. President Carter later went down this same road, also unsuccessfully.

MDRC's supported work demonstration, although not initiated with the main purpose of studying welfare policies, reflects the service strategy as an approach to welfare reform, which is the approach that has come to dominate nearly three decades later under President Clinton. The supported work demonstration began operations in 1974, not long after last rites were administered for Nixon's Family Assistance Plan. The idea of supported work, as its name implies, is that disadvantaged people of working age should be "supported" as they enter the workforce. As they increasingly become adjusted to the workplace, this support is gradually withdrawn, the goal being to help participants acquire work experience and job skills so eventually they can become regular members of the labor force. Presumably, experience interacting and working with peers, meeting deadlines, and relating successfully to supervisors are skills that later on will help people with limited work experience obtain and stay in *un*supported jobs.

The supported work program as originally developed by the Vera Institute of Justice did not apply to welfare family heads; it was focused on criminal justice offenders as an alternative to incarceration. The aim of MDRC's national demonstration was to test the supported work approach on a broader basis, both geographically and in terms of the groups aided. Question number one was: What groups should be aided? A key factor was money. The funders of the demonstration, particularly the federal agencies involved, had different interests and constituencies. The Labor Department was interested in youth. The Department of Justice was interested in previous offenders. The drug-abuse prevention agency was interested in former addicts. The Department of Health, Education, and Welfare (now the Department of Health and Human Services) initially was reluctant to participate in the funding consortium for the national supported work demonstration. Caspar Weinberger, at that time secretary of HEW, had reservations about the administrative arrangement whereby a nonprofit intermediary would coordinate and manage primarily federal funds for the demonstration. Although HEW eventually provided money for the supported work demonstration, a decision was made by MDRC prior to HEW's decision to participate to include long-term female welfare family heads as one of the target groups in the demonstration. This was done on the assumption that MDRC could use Ford Foundation and some federal funds to serve this group. As it turned out, this was a fortunate decision.

Altogether, there were four target groups in the supported work demonstration — problem youth; former offenders (that is, persons who had previously been incarcerated as a result of a conviction for a criminal offense); former narcotic addicts; and long-term female welfare family heads. (See Table 5.1.) The definition of the target groups

Table 5.1
Supported Work Eligibility Criteria, by Target Group

Target Group	*Eligibility Criteria**
AFDC	Women on AFDC both currently and for 30 out of the preceding 36 months; youngest child 6 years old or older
Ex-addicts	Age 18 years or older; enrolled in a drug treatment program currently or within the preceding 6 months
Ex-offenders	Age 18 years or older; incarcerated within the last 6 months as a result of a conviction
Youths	Age 17 to 20 years; no high school or equivalency degree; not in school in the last 6 months; delinquency record, conviction, court appearance, or similar (for at least 50% of the youth)
All groups	Currently unemployed†; spent no more than 3 months in a job during the past 6 months

Source: Board of Directors, Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported Work Demonstration* (Cambridge, MA: Ballinger, 1980), p. 23.
* Supported work eligibility criteria refer to conditions prevailing at the time of application to the supported work program. If a person in supported work voluntarily or involuntarily leaves the program and subsequently reapplies for a supported work job, he or she is not reviewed again for acceptance under the eligibility criteria.
† Worked no more than 10 hours a week for the last 4 weeks.

was only the beginning. Many other decisions had to be made. Since this was the first demonstration conducted by MDRC, the learning process was more challenging and time-consuming than for later demonstrations.

As the supported work demonstration got under way, discussions within the corporation came to reflect an increasingly cautious view about the challenges to be faced. Eli Ginzberg, the first chairman of the board of MDRC and a master at getting to the heart of issues, produced the following wise admonitions based on the discussion at one of the early meetings in this planning process:

- The best-designed and controlled social experiment can never take account of exogenous factors.
- The odds are very strong that any social intervention will be too weak to show clear-cut positive effects.

- Many types of effects cannot be caught in even the most sophisticated evaluation.
- Because of hidden bias in clients and selectors, the matching of experimentals and controls is likely to be flawed.

As implied by Ginzberg, all the hurdles discussed in the previous chapter had to be dealt with in planning this research. At the outset, concern about selection bias and the null hypothesis led to a decision to test what was regarded as a relatively long (nine months to one year) treatment, and to compare the experience of the treatment groups with that of randomly assigned control groups. (Actually, the decision to adopt random assignment was an integral part of the research design for this study from the very start.)

Despite the fact that HEW funds were not initially included in the demonstration, the corporation decided to set up its own human subjects institutional review board, headed by MDRC board member Gilbert Y. Steiner. (None of the other members of the institutional review board were members of the MDRC board.) In turn, the MDRC institutional review board decided to adopt HEW's rules for the treatment of human subjects. This meant that an informed consent agreement had to be obtained from all participants in the demonstration.

The sponsoring organizations of supported work projects at the local level were mostly small nonprofit organizations. The possibility always existed that the project would suffer from what was described in chapter 4 as service contamination; however, the fact that most other employment and training programs for the eligible population in the communities studied tended to be shorter in duration and less intensive was seen as a factor that would mitigate this problem.

At the outset of the research, considerable staff and board time were devoted to the selection of research contractors. Later, the relationship with contractors was the focus of a major controversy over prospective cost overruns that resulted in protracted bargaining to prevent this from happening. Robert Solow (then vice-chairman of the MDRC board) and I (then treasurer) participated in intense negotiations with the research contractor, Mathematica Policy Research of Princeton, New Jersey. We won. Mathematica researchers conducted the baseline and follow-up surveys of supported work participants and controls and were responsible for the impact and cost-benefit studies, with assistance provided by researchers at the Institute for Research on Poverty at the University of Wisconsin.

Results Focus on Welfare

The most important findings from the supported work demonstration involved differences among the four treatment groups. The impact of supported work was largest for the welfare group. The report on the demonstration stated, "The program has proved most effective in preparing for employment a substantial number of women who have been on welfare (AFDC) for many years."[3] Table 5.2 shows the results by time period for this group. The most important period is the nineteen to twenty-seven month post-enrollment period. The differences between the experimentals and controls in the earlier periods (one to eighteen months) in hours worked and earnings were likely to have been a result of the fact that the welfare family heads in the treatment groups were enrolled in the supported work program during this period.

The 8.5 percent difference (bold and underlined in Table 5-2) in employment between the AFDC treatment and control groups in the period nineteen to twenty-seven months after enrollment is statistically significant at the 5 percent level. The treatment group worked nearly sixteen hours more per month in this period than the controls and earned an average of $77 per month more than the controls. Their AFDC benefits were reduced by an average of 14 percent, and a like proportion for food stamps. In short, supported work was successful for the welfare group, although the gains made were not large or dramatic. For the former-addict groups (see Table 5.3) the results were also positive.

A close reader of these two tables will notice an important point. The impact of the supported work program on the welfare (AFDC) group is in large part a result of the fact that the controls fared relatively poorly in the labor market. (Compare the labor market experience of the AFDC and former-addict control groups in Tables 5.2 and 5.3.) Similar

Table 5.2
Experimental-Control Differences During 27 Months Following Enrollment in Supported Work Demonstration; AFDC Group

Outcome Measure	*Experimentals*	*Controls*	*Difference*
Percentage employed during period			
Months 1-9	96.3	36.5	59.8*
10-18	76.5	39.4	37.1*
19-27	49.1	40.6	8.5*
Average monthly hours worked			
Months 1-9	135.3	26.6	108.7*
10-18	79.4	40.3	39.1*
19-27	60.9	45.2	15.7*
Average monthly earnings ($)			
Months 1-9	400.44	78.28	322.16*
10-18	274.06	131.08	142.98*
19-27	242.89	165.88	77.01*
Cash welfare payments † (Percentage receiving)			
Months 1-9	93.8	97.7	-3.9*
10-18	82.4	80.1	-7.7*
19-27	71.4	85.1	-13.7*
Average monthly amount ($)			
Months 1-9	169.82	277.90	-108.09*
10-18	164.28	246.60	-82.32*
19-27	172.06	224.00	-51.94*
Food stamps: average monthly bonus value ($)			
Months 1-9	44.83	63.46	-18.63*
10-18	42.15	58.02	-15.87*
19-27	47.14	60.25	-13.11*
Average monthly total income ($) ††			
Months 1-9	628.06	435.10	192.96*
10-18	524.47	454.44	70.03*
19-27	497.50	470.14	27.36

Source: Board of Directors, Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported Work Demonstration* (Cambridge, MA: Ballinger, 1980), p. 153.
Note: Averages are calculated for all members of the sample, including those with no employment or transfer payment receipt in the covered period.
* Statistically significant at the 5 percent level.
† Welfare includes AFDC, General Assistance, Supplemental Security Income, and other unspecified cash welfare.
†† Total income includes earnings, unemployment compensation, welfare, food stamp bonus value, and other unearned income (Social Security, pensions, alimony, and child support).

Table 5.3
Experimental-Control Differences During 36 Months Following Enrollment in Supported Work Demonstration; Ex-Addict Group

Outcome Measure		*Experimentals*	*Controls*	*Difference*
Percentage employed during period				
Months	1-9	95.0	50.2	44.8*
	10-18	63.9	53.1	10.8*
	19-27	56.5	53.0	3.5
	28-36	64.0	53.9	10.1†
Average monthly hours worked				
Months	1-9	118.7	40.5	78.2*
	10-18	66.4	50.0	16.4*
	19-27	60.1	58.6	1.5
	28-36	70.9	52.6	18.3*
Average monthly earnings ($)				
Months	1-9	361.23	159.79	201.44*
	10-18	259.62	220.42	39.20†
	19-27	277.75	261.33	16.42
	28-36	326.09	224.36	101.73
Average monthly welfare and food stamp benefits ($)††				
Months	1-9	57.97	115.17	-57.20*
	10-18	92.42	110.89	-18.47*
	19-27	89.90	93.94	-4.04
	28-36	94.34	103.79	-9.45
Percentage using any drug other than marijuana or alcohol				
Months	19	36.1	38.2	-2.1
	10-18	34.1	32.7	1.4
	19-27	28.0	27.5	0.5
	28-36	23.4	20.7	2.7
Percentage using heroin				
Months	1-9	20.2	21.5	-1.3
	10-18	16.8	17.8	-1.0
	19-27	13.4	11.7	1.7
	28-36	10.1	8.8	1.3
Percentage arrested				
Months	1-18	25.3	33.5	-8.2*
	1-36	35.0	53.1	-18.1*
Percentage arrested for robbery				
Months	1-18	2.3	7.5	-5.2*
	1-36	0.2	13.4	-13.2*

Continued

Table 5.3 (Continued)
Experimental-Control Differences During 36 Months Following Enrollment in Supported Work Demonstration; Ex-Addict Group

Outcome Measure	*Experimentals*	*Controls*	*Difference*
Percentage arrested on drug charges			
Months 1-18	4.1	7.9	-3.8*
1-36	6.8	14.0	-7.2
Percentage convicted			
Months 1-18	13.5	17.8	-4.3†
1-36	19.3	32.9	-13.6†

Source: Board of Directors, Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported Work Demonstration* (Cambridge, MA: Ballinger, 1980), p. 155.
Note: Averages are calculated for all members of the sample, including those with no employment or transfer payment receipt in the covered period.
* Statistically significant at the 5 percent level.
† Statistically significant at the 10 percent level.
†† Welfare includes AFDC, General Assistance, Supplemental Security Income, and other unspecified cash welfare.

findings have been made in other studies. As it turns out, the problems of female welfare family heads with children are so serious in terms of their ability to enter and participate in the labor market that an intervention for this group is likely to have a bigger impact than for other groups. Society, in effect, gains more from investing in programs for this hard-to-serve group.

This finding that the most disadvantaged groups gain the most from job training programs has to do with what is known in employment and training parlance as "creaming." Many employers and also the managers of training programs intuitively see themselves as benefiting the most from a focus on the most job-ready people — that is, the people most likely to make it in the labor market. Helping them, it is felt, can provide employers with reliable workers at minimal cost and put program sponsors in a strong position to claim success for their efforts. But the point that is often missed is that the most job-ready participants are likely to make it anyway. The tested program is not providing added value for the society. Thus, targeting, as in this case on disadvantaged long-term welfare family heads, should have been expected to produce a positive program impact — and it did.

For two groups in the demonstration (problem youth and previous offenders) the results showed no or very little impact of supported work. According to MDRC's first report on the demonstration, the program did not yield long-term positive results for the youth group. It had a marginal positive impact on the offenders group.

Implications for Welfare Reform

The supported work demonstration was initiated in the mid-1970s, a period in which generally liberal attitudes on social policy predominated. But, by the time MDRC had completed the demonstration and published its results, it was 1980. The tide had turned. Supported work had been successful for disadvantaged welfare family heads. However, it was seen by many politicians and public officials in 1980 (the year Ronald Reagan was elected president) as too expensive to be replicated on a broad basis.

This initial reaction to the findings of the supported work demonstration vis-à-vis welfare family heads proved to be short-lived however. As the results became more widely known, government officials at every level (national, state, and local) began to notice the finding about the employability gains for long-term welfare recipients and sought to replicate supported work and derivative concepts. In 1981 and following, as President Reagan's plans for welfare were advanced, the lessons of the supported work demonstration came to play an increasingly important role in policy processes.

Actually, welfare reform had long been a priority policy area for Reagan. In his second term as governor of California, Reagan made welfare reform the central issue of his administration. Experts debate the degree to which he succeeded, but for Reagan there were no doubts: "When I took office, California was the welfare capital of the nation. The caseload was increasing 40,000 a month. We turned that 40,000 a month increase into an 8,000 a month decrease. We returned to the taxpayers $2 billion and we increased grants to the truly needy by forty-three percent."[4] According to Fred C. Doolittle, the California

Welfare Reform Act, passed in 1971, "marked a turning point in the political career of Ronald Reagan."[5] Reagan biographer Lou Cannon agrees. He viewed the battles about welfare reform in California as "Reagan's transformation from communicator to governor."[6] Reagan's initial involvement in welfare reform at the state level coincided with the decidedly more liberal welfare reform efforts of the Nixon administration at the national level beginning in 1969.

Throughout the Nixon period, Reagan challenged Nixon's views on social policies, and was a leading proponent of a conservative state-centered position. His advocacy of "workfare" and state authority over welfare met with strong resistance generally from liberals and from officials in the Nixon administration. Later on, in discussing the national welfare reform act signed by President Clinton in 1996, we shall see that Reagan's views regarding both his focus on work and workfare and decentralizing social programs largely prevailed in America. Nixon's and Carter's centralized and more liberal proposals to replace the Aid to Families with Dependent Children (AFDC) program with a comprehensive national payment system, heavily embodying the income-maintenance approach, were stymied, although important liberalizing changes were made. The most notable liberalizing and centralizing change involved the nationalization under Nixon of cash assistance for the disabled and the elderly poor.* However, in the case of the most controversial welfare program — AFDC — major changes were resisted up until the Clinton presidency. Financial responsibility for the AFDC program continued to be shared by the federal government and the states on a matching basis.**

---

* Under Nixon, federal grants-in-aid to the states for welfare programs to aid the aged poor, blind, and disabled were converted into centralized programs with uniform eligibility requirements, administered, not by the states, but by the U.S. Social Security Administration. This is the Supplemental Security Income (SSI) program. The food stamp program, begun as a pilot program under President Kennedy, also was expanded greatly under Nixon. It was made automatic and universal, in effect, becoming a mini-negative income tax operating on a uniform basis throughout the country.

** In 1996, welfare became a block grant to the states, rather than a matching grant-in-aid, as explained in chapter 8.

As governor of California, Reagan along with other conservatives, strongly opposed the overhaul of the AFDC program along the lines of a negative income tax as embodied in Nixon's Family Assistance Plan (FAP). In fact, although both Nixon and Reagan were Republicans, they had a history of frosty relations.* Reagan not only was a leader among conservatives in opposing Nixon's welfare reform plans, he offered his own counterproposals. In testimony before the Senate Finance Committee in 1972, Reagan presented his state-based California approach to welfare reform as a better model for national legislation. He boasted that these proposals "are the product of our experience with an actual reform program that is succeeding in California, they are not a theory."[7] Reagan's California program consisted of increased benefits for the "truly needy," reduced benefits or no benefits for the "working poor," and "workfare" (the mandatory-work-for-you-benefits approach). Rather than centralizing the AFDC program, Reagan favored turning it over to the states in the form of a block grant (shades of 1996!) whereby states would receive a lump-sum amount of money to use on a flexible basis to aid poor families.

There is a paradox in Reagan's role in the debate on welfare policy. Despite his strong support for decentralizing welfare, Reagan was much more successful as president than he was at the state level as governor in advancing his welfare reform ideas. Controversy exists over whether Reagan's California reforms reduced welfare costs and caseloads; his "workfare" plan resulted in minuscule participation in mandatory public jobs. On the other hand, Reagan made major gains advancing his welfare goals at the national level when he was elected president in 1980. As part of Reagan's 1981 changes to the budget President Carter had submitted in his final month in office, Reagan won approval for welfare policies that tightened eligibility and benefit rules in ways that reduced the number of family heads on AFDC and Medicaid, and cut benefits for many other families.[8] For purposes of this book, the most important element of the Reagan welfare reform strategy was workfare.

* I attended meetings between the two men on welfare reform in the early 1970s. You could cut the tension between them with a knife.

The word "workfare" has an interesting history. It was used by Nixon in his national television address in 1969 presenting his "New Federalism" domestic program. "What America needs now," said Nixon, "is not more welfare, but more 'workfare.'"[9] Nixon intended that his welfare reform plan should be known by this term.* But "workfare" as a label did not catch on in the way Nixon intended. The word in 1970s usage came to have a harsher meaning than Nixon called for in his 1969 speech on the "New Federalism." I have never been able to pin down why this happened. The most plausible explanation is that the press attached the term "workfare" to what was viewed as a conservative alternative to Nixon's welfare plan advanced by Senator Russell Long (D-LA), then chair of the Senate Finance Committee. Senator Long proposed to limit cash assistance payments just to employable needy family heads to compensate for work performed. This kind of workfare was an anathema to liberals in the 1970s; they lambasted it as "slavefare," although later it became the central concept of Wisconsin's far-reaching welfare reform of the mid-1990s.

In California in the 1970s, Reagan had strongly advocated this mandatory work-for-your-welfare approach. Again, a decade later as president he made this concept a key point, proposing that workfare be made compulsory and nationwide. Although many of Reagan's 1981 welfare reform proposals were incorporated in the budget act passed that year, Congress was reluctant to go all the way with his version of workfare.[10] Congress would only agree to give the states authority to implement a watered down version of Reagan's workfare proposal, and only on a trial basis. But even though in 1981 workfare under Reagan was permissive to the states, it did not get off the ground rapidly in any state.

Up to now, I have referred to the workfare part of Reagan's 1981 welfare reform efforts in the singular. Actually, there were several such provisions included in the budget act passed in 1981. One provision authorized the Community Work Experience Program (CWEP), the strict work-for-your-welfare alternative to cash payments. The law in this case authorized the states to use AFDC funds to pay eligible family

* William Safire, then a Nixon speechwriter, wrote this speech and boasted about this effort at wordsmithing.

heads in exchange for mandatory public employment. Another employment and training provision of the 1981 budget act was sponsored by Senator Long, then ranking minority member of the Senate Finance Committee (he was no longer chair, because the Republicans had taken over the Senate). Long's plan gave the states authority to subsidize on-the-job experience for AFDC recipients by "diverting" welfare grants to wage subsidies to private employers. This came to be known as "grant diversion" in the 1980s.* The third pertinent provision in the 1981 budget act, inserted by Senators Moynihan (D-NY) and David Boren (D-OK), turned out to be much more important than was originally expected. Its purpose was to overhaul the Work Incentive Program (WIN), enacted in 1967 under President Johnson to provide funding and authority for the states to run job placement, training, and related service programs for welfare family heads.[11] Up until 1981, WIN programs conducted by the states had been required to be jointly administered by a state's employment service (the labor-exchange agency) and its welfare agency. Moynihan argued that this requirement for joint management, which he said often resulted in "byzantine administrative arrangements," should be changed to allow states to apply for waivers to administer the WIN program just under the state's welfare agency. This provision, called "WIN-Demo" authority ("Demo" standing for demonstration), turned out to be very important.

MDRC's Work/Welfare Demonstrations

In the changed and more conservative social policy environment of the early 1980s, the Manpower Demonstration Research Corporation had to change too. It shifted its focus from Washington to the states. The corporation began work with states on a series of work/welfare demonstrations. MDRC used a challenge grant from the Ford Foundation to pay part of the costs of demonstrations, and the states (or, in some cases, other foundations) provided the remaining funds needed. MDRC demonstrations in eight states, using the "WIN-Demo" authority described above, tested different ways states reformed their welfare systems. This

* The term "diversion" has a different meaning now in the welfare field, referring to "diverting" people from welfare to jobs. See chapter 10 for a discussion of this point.

relationship between social science and social policy was very different from the case of the negative income tax experiments a decade earlier. Then, the research agenda was set at the federal level mostly by social scientists to test ideas that in the main originated with them. This time, state political leaders were setting the agenda.

Not surprisingly, the first state to have an MDRC work/welfare demonstration was California — in San Diego County. Beginning in August 1982, San Diego County assigned 5,000 AFDC family heads to two work/welfare programs. Another 2,000 people were randomly assigned to a control group. The two programs tested were group job search (called a "job club")* and a combination of a "job club" followed by assignment to community work experience in cases where participants in the "job club" did not find employment.[12] The community work experience component of the San Diego demonstration lasted thirteen weeks and had a time limit of thirty-two hours per week. Actually, officials of San Diego County had a long-standing interest in programs of this type. In the case of this new work/welfare demonstration, the state Employment Development Department (California's employment service agency) contracted with MDRC to conduct the research. The seven other states that later joined with MDRC in work/welfare demonstrations were Arkansas, Illinois, Maine, Maryland, New Jersey, Virginia, and West Virginia. Altogether, 35,000 people participated on a random assignment basis in these eight state-based work/welfare demonstrations.

Although the nature and scope of these state demonstrations were varied, there were common elements. In all of them, with the exception of New Jersey and Maine where the demonstration was of private-sector grant diversion, there was some degree of obligation. That is, welfare family heads in the treatment group were required to do something in order to receive their AFDC benefits. "Doing something" could mean engaging in a job search or participating in a "job club" or in a training or education program. It could also include a period of

* A job club is a group activity with a trainer-instructor in which welfare recipients develop skills in preparing applications, locating job openings, and being interviewed for employment. Often banks of telephones are used in the job search, and in some cases training sessions for job interviews are videotaped as part of an instruction-feedback process.

community work experience or some sequenced combination of these program elements.

There was considerable variation among the states in the MDRC work/welfare demonstrations, particularly in the character and degree of obligation in the tested programs. The San Diego project and the West Virginia project, both for male welfare family heads, had a strong mandatory character; whereas in Maryland, there was an obligatory feature of the program, but it was not strong. These state demonstrations did not represent an altogether new idea. Efforts had been made in the past at both the national and state levels to require welfare family heads (usually those with children above preschool age) to search for employment, accept a "suitable" job if one was offered, and, if not, to participate in a job training program. What was distinctive about these demonstrations was that the obligation part was supposed to be taken seriously. People who did not cooperate were supposed to be sanctioned.* One can think of these demonstrations as ratcheting up efforts to convert the Aid to Families with Dependent Children program from an entitlement to a conditional program, which is precisely what was done in the national welfare reform law enacted in 1996.

In most cases, these MDRC work/welfare demonstrations showed positive impacts, although the results generally were modest. (In fact, this point about modest results is very important in the policy process. My reading of this history is that the research worked. It was useful and used. Politicians concluded, not incorrectly, that the basic policy — focused as it was on the human-capital (training and education) approach to welfare reform — was not big enough or bold enough to produce what to political leaders regarded as satisfactorily positive outcomes. The result in the 1990s was the adoption of a stronger work, and specifically a "work-first," approach to welfare reform, as opposed to the softer human-capital approach. The research lessons learned had big reverberations, first at the state level in California and other states and

* Sanctioning procedure varied in these demonstrations. Usually, they involved a partial reduction of the family's welfare benefit for a period of time. Welfare recipients could also be required to have social workers manage their finances as a sanction for noncompliance. For a discussion emphasizing the importance of sanctioning in welfare programs, see Lawrence M. Mead, *Beyond Entitlement: The Social Obligations of Citizenship* (New York: Free Press, 1986).

ultimately nationally in 1988 and 1996. In the near term, state leaders in California, in large part on the basis of the results in San Diego, decided that the obligatory employment approach to welfare should be adopted statewide. Eventually, this led to a new role for MDRC, conducting a statewide evaluation of this ongoing program, called "California's Greater Avenues to Independence" or GAIN program. Part III of this book, which deals with evaluation studies, includes a chapter on the experience of MDRC in evaluating the GAIN program and also chapters on evaluations of the federal welfare reform laws enacted in 1988 and 1996.

Endnotes

1 Lee S. Friedman, "An Interim Evaluation of the Supported-work Experiment," *Policy Analysis* 3 (spring 1977): 147-70; and "The Use of Ex-Addicts to Deliver Local Services: The Supported-work Experiments," in *Urban Problems and Public Policy Choices*, eds. J. Bergman and H. Wiener (New York: Praeger, 1975), pp. 58-71.

2 WIN is an acronym for the Work Incentive Program, a federal grant-in-aid program to the states to provide employment and training services to welfare family heads.

3 Board of Directors, Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported-work Demonstration* (Cambridge, MA: Ballinger, 1980), p. 2.

4 Fred C. Doolittle, "Ronald Reagan and Conservative Welfare Reform," Princeton Urban and Regional Research Center, July 1986, p. 1-3.

5 Ibid., p. 1-3.

6 Lou Cannon, *Reagan* (New York: G. P. Putnam's Sons, 1982), p. 183.

7 Doolittle, "Ronald Reagan," pp. 3-20.

8 A number of studies have been conducted of Reagan's changes in domestic policy. Studies funded by the Ford Foundation conducted at Princeton University and the Urban Institute produced reports and articles describing the Reagan changes and evaluating their effects in the terms discussed in this chapter. See Richard P. Nathan et al., *Reagan and the States* (Princeton, NJ: Princeton University Press, 1987), and George E. Peterson and Carol W. Lewis, eds., *Reagan and the Cities* (Washington, DC: Urban Institute Press, 1986).

9 The President's Address to the Nation on Domestic Programs, August 8, 1969, *Weekly Compilation of Presidential Documents*. 5, no. 32, Monday, August 11, 1969 (Washington, DC: U.S. Government Printing Office), p. 1111.

10 Wonder of wonders. It was President Clinton's welfare reform in 1996 that most closely approximates Reagan's ideas. See chapter 9.

11 When it was first enacted, the Work Incentive Program was known by the acronym WIP (as in whip) but this identifier was soon changed to WIN.

12 The Manpower Demonstration Research Corporation (MDRC) has issued several reports on the California demonstration. See Barbara Goldman, Judith Gueron, et al., *California: The Demonstration of State Work/Welfare Initiatives; Preliminary Findings from the San Diego Job Search and Work Experience Demonstration* (New York: Manpower Demonstration Research Corporation, February 1984); Barbara Goldman, Daniel Friedlander, et al., *California: The Demonstration of State Work/Welfare Initiatives; Findings from the San Diego Job Search and Work Experience Demonstration* (New York: Manpower Demonstration Research Corporation, May 1985); and Barbara Goldman, Daniel Friedlander, and David Long, *California: The Demonstration of Work/Welfare Initiatives; Final Report* (New York: Manpower Demonstration Research Corporation, February 1986).

PART III

EVALUATION RESEARCH

6

The Nature of Evaluation Research

Eugene Bardach, in a book on what he calls "the implementation game," made a point that is recurrent in the political science literature. It is hard enough, he said, to design public policies that look good on paper. It is harder still to sell them to the public. "And it is excruciatingly hard to implement them in a way that pleases anyone at all, including the supposed beneficiaries and clients."[1] In a similar vein, political scientist Clinton Rossiter said many U.S. presidents found their hardest job is "not to persuade Congress to support a policy dear to his political heart, but to persuade the pertinent bureau or agency — even when headed by men of his own choosing — to follow his direction faithfully and transform the shadow of the policy into the substance of the program."[2] According to Angela Browne and Aaron Wildavsky, "Policy implementation is hypothesis testing: It is *exploration*."[3] An explorer cannot predict what will be discovered. Every voyage is different.

As compared with a demonstration study of a potential new policy, researchers have less control over the conditions under which an evaluation study of an ongoing program is carried out. This is true, despite the fact, as we have seen, that the ability of researchers to control the environment in a demonstration study is much more limited and constrained than might at first blush appear to be the case. However, the control problems of demonstration research pale in comparison with the environment of an evaluation study in which the purpose of the policy or program being studied is not to learn something but to do something.

Under these kinds of real world conditions policymakers and administrators often have other, and for them much more critical, needs to take into account than those of researchers. They are unlikely to be moved by arguments that the policy being implemented must be carefully specified, closely monitored, uniform everywhere, and sufficiently distinguishable from other policies in order to clear all the research hurdles described in chapter 4. There are sure to be players in the governmental process who regard the evaluation of a program as an intrusion that will delay, complicate, or even undercut the achievement of their policy goals. They may be antagonistic to researchers as a species. Or they may not want research to be conducted for political reasons — because they fear it will show a policy they favor to be ineffective, or if it works, to have results that fall far short of what they had promised. Overpromising is endemic in American government. In fact, some participants in governmental processes argue that it is essential to overpromise in order to get anything agreed to.

Even when policymakers are sympathetic to the purposes of evaluation research, differences in perspective between policymakers, program operators, and researchers are likely to be substantial. The essential point is that, because of these and other factors, social scientists doing evaluation research do not have the same kinds of opportunities they have under the more controlled conditions of a demonstration study to test new program ideas.

The most important consequence of the more limited control researchers have in an evaluation study as compared to a demonstration study is that it is much more difficult to know the counterfactual. Demonstration and evaluation researchers ask the same bottom-line question: What happened as a result of this pilot or ongoing program that would not have happened if the program being studied had not existed? We can never have a perfect situation, as we have seen — one in which the same person or group is both treated and untreated by a given program. The next best solution is randomly to assign a number of eligible persons to treatment and non-treatment groups and compare the differences between the two groups in terms of the outcomes a given program is supposed to produce. However, in the case of an evaluation study of an ongoing program where research is a much lower-order objective, the suggestion that a research design should be adopted that will evaluate the program by randomly excluding some participants from it is frequently and strongly resisted. Moreover, on ethical grounds, if the policy being evaluated is universal (that is, if it applies to all eligible persons), one is hard put to argue that the policy should be suspended for some otherwise eligible participants in the interest of research.

As has been discussed, less good alternatives compared to random assignment exist for establishing the counterfactual. If the policy we are evaluating is selective on a geographic basis, one could argue that the government should select some places as comparison sites. Although such an approach may make sense to researchers, it, too, can involve formidable problems for public officials. They may feel that it is politically unwise to have comparison sites because this would be perceived as unfair and be a source of controversy.

The previous discussion in Part II of this book on demonstration research considered statistical simulation techniques to create a baseline for comparison in studies to test new policy approaches. Despite the problems involved, this is often the approach used in evaluation studies, especially to study impacts on people. This approach involves predicting a particular set of dependent variables (that is, dependent on the policy being studied) for a population eligible to be treated by the policy. A variant of the statistical analysis approach is to use benchmarks based on past research, or on the views of experts, as the standards against which to evaluate the impact of an ongoing program on individual participants. None of the previous evaluation studies of Lyndon Johnson's Great Society programs that I know of had a randomly assigned control group.[4] Some studies of Great Society programs, as in the case of the Job Corps mentioned in chapter 4, used comparison sites, and a new evaluation of the Job Corps now in progress uses random assignment. However, most evaluations of Great Society programs used econometric or other statistical and comparison-group techniques to evaluate program effects on people. This is not to argue that there never can be value added through rigorous randomized studies of the effects on individuals of ongoing programs, and especially of planned variations in their implementation. But I maintain that such studies are extremely hard to design and carry out in the constantly shifting policy terrain of most U.S. domestic programs.

Moreover, a big difference between demonstration and evaluation research is that the latter is more likely to focus on *institutional* as opposed to *individual* effects. The reasons for this and the importance of this point can be shown by again using the example of a computer-assisted remedial reading program. I stipulated in chapter 4 that this was a state government demonstration that aided school districts and local schools. Assume now, as is not unusual in American government, that the federal government is providing grant-in-aid money to the states for this type of special reading remediation. Assume further that earlier randomized tests (demonstration studies) were conducted to show that this approach worked, and now the tested program is being generalized (or what social scientists call "replicated") — but that the policy specifications in the federal legislation enacted to accomplish this purpose are very general. This is often the case. And add one more complication: In most states, this federal aid is incremental — that is, in addition to money provided by state and local governments for the same or similar purposes.

Policymakers in Washington who provide grant-in-aid funds to the states for reading remediation understandably have an interest in the response to "their" program. However, the questions they want answered are likely to center on the behavior, not of individual students, but of different levels of government and types of organizations, especially schools and school districts, which are the recipients of these federal grants-in-aid for reading remediation: What did states, school districts, and schools do with these funds in terms of the types of services provided and the number of schools aided? Were the funds additive? Were computer companies, consultants, experts, etc., important actors in using these grant funds? These are important policy effects, but they are different from those we focused on in our consideration of demonstration studies. They are, to repeat, *institutional.*

The Federalism Barrier Reef

This discussion dramatizes what I believe is the crucial intellectual difference between demonstration and evaluation research. Demonstration research focuses on *individual* outcomes: Will a computer-assisted reading remediation program help children learn to read? If it is shown in a demonstration that a particular approach has significant positive effects, it may well be decided to adopt a national program in the form of a grant-in-aid from the federal government to the states to advance this purpose. However, once such an intergovernmental aid program exists, political leaders are likely to be especially, and for practical reasons, most interested in whether and how it changes the behavior of the institutions that provide this service. As these and other funds trickle down in the American intergovernmental system, it would be extremely difficult (even if higher level officials wanted to do so) to learn how the policies involved could be shown to have affected the reading scores of individual students.

Most of the business of providing public services in the domestic public sector in the United States is conducted in this way — *indirectly* — that is, *intergovernmentally* through grants-in-aid to states and localities. Localities, in turn, frequently contract with nonprofit and for-profit organizations for the services provided. In essence, politicians are seeking to affect the behavior of institutions through these grants-in-aid in the belief that this will influence the activities of the recipient institutions in ways that ultimately affect individuals. This is the federalism barrier reef of American public policy. It is highly underappreciated. In fact, it is often just plain ignored in studying the effects of U.S. domestic policies and programs.

Researchers must watch this reef carefully, for it can ruin their voyages if they do not take it into account. The federalism barrier reef affects decisions about public policy research in many ways. An intervention in the form of a grant-in-aid from one level of government to another may be small. It may in fact be so small that, picking up on the reasoning stated in chapter 4 in regard to the null hypothesis, we could not possibly expect to find detectable effects on individuals of one such stream of money as opposed to other streams of money from a different level of government or from foundations or private citizens or from fees and charges.

Scientific Implications

These observations about the political and federalism terrain of the evaluation of ongoing programs bring into play other major themes of this book about the need for linkages between social science disciplines and between quantitative and qualitative research designs and methods. Many of the kinds of observations that researchers are called upon to make about institutional behavior in response to a given policy involve the way different types of organizations behave.

In theory, we could study differences in organizational behavior by randomly selecting a sample of a given type of organization (for example, school districts or schools in the previous illustration of a computer-assisted reading remediation program), treating them in different ways, and then analyzing their responses to the policy being evaluated in order to generalize about their institutional effects. We don't do this. One of the reasons we don't is that the study of the effects of public policies and organizations does not lend itself to such experimental research. We lack sufficiently strong and widely agreed upon theories about organizational behavior on which to base rigorous comparative research. We also lack agreed upon, available, and measurable statistics on which basis we could assess and compare organizational behavior. Other reasons, too, come into play.

Organizations are harder to manipulate in a research environment than individuals, although the latter is hard enough to do. Consider the difficulties that would be involved in seeking informed consent from a local government to subject itself to a random selection procedure that would determine whether it would receive, or not receive, a particular grant-in-aid. Local governments are simply too complex, and most are too politically assertive, to give informed consent to participate in this way in a tightly structured evaluation study. The same point can be made about the nonprofit organizations that operate most social programs.

We need next to look at the way researchers deal with causality in studying institutional as opposed to individual behavior. The most common method for studying the behavior of institutions is to observe them closely in order to make informed judgments about how a given stimulus or event (e.g., a policy change or a new program) appears to have affected their behavior. What researchers do in such a situation is model the counterfactual on the basis of their understanding of the behavior of the types of organization being studied. They observe the behavior of a particular organization or groups of organizations under conditions in which a new policy or program is operating and compare that behavior to what they expect would have been their behavior without the new policy. The catch — and it is a significant one — is that this reasoning process cannot be either specified or replicated.

Consider next the perspectives of different social science disciplines. Since the mid-1960s, as social scientists increased their role in public policy research, economists have had the upper hand. This is very much to their credit as I stated earlier. Political scientists and sociologists on the other hand have not played anywhere near as big a role. In fact, and I think this is regrettable, the study of public administration as a sub-field of political science has relatively low status in the discipline. Moreover, management studies in general tend not to have an extensive or well developed intellectual base, although recently there has been new interest in institutional studies in political science.*

Sociologists, as suggested earlier, also have a claim on the field of organizational behavior. Robert A. Scott and Arnold R. Shore, in their book *Why Sociology Does Not Apply*, concluded that sociologists, insofar as they have been involved in policy-related research, have done studies to analyze policy problems and prescribe solutions, but that the results of these studies have been disappointing. They said they are "politically unrealistic, administratively unworkable, or simply impractical." They also complained that policy-relevant sociological research is often theoretical and removed, "stimulated by a desire to advance disciplinary knowledge."[5] Social psychologists, too, have a claim on organizational behavior. Thomas F. Pettigrew, a leader in social psychology, while not optimistic about what could be achieved, reached similar conclusions to Scott and Shore about the need for greater attention to the development

* A seminal article by James G. March and Johan P. Olsen in 1984 called for "a new institutionalism" in political science. The authors maintained that because institutions are not the sum of the actions of their members, their behavior is not easily predicted. (James G. March and Johan P. Olsen, "The New Institutionalism: Organizational Factors in Political Life," *American Political Science Review*, 1978.) March and Olsen held that organizational behavior is *ahistorical*, hard to measure, and characterized by symbolic action. They called for greater emphasis in political science research on studies that take cognizance of these characteristics of organizational behavior. This position is similar to that of institutional economist John R. Commons described in chapter 2.

and application of research techniques for organizational studies. Pettigrew advocated "more interdisciplinary work within social science" and efforts to merge inductive and deductive methods.[6]

The essential point of this chapter is that many of the critical questions asked in evaluation research on the effects of ongoing public programs do not lend themselves to research approaches as rigorous as those used in studying the impacts of pilot programs on individuals in the more controlled and more controllable conditions of a demonstration study. My further contention is that a careful view of the way causation is inferred and proof is built up over time in other scientific disciplines (physics and cosmology are good examples) shows that the methods used in evaluation research, which is often inductive, are legitimately scientific. I turn next to specific studies to amplify this point in examining the challenges involved and techniques used in assessing the effects of ongoing social programs in the real world of American federalism.

Endnotes

1 Eugene Bardach, *The Implementation Game: What Happens When a Bill Becomes a Law* (Cambridge, MA: MIT Press, 1977), p. 3.

2 Clinton Rossiter, *The American Presidency* (New York: New Amsterdam Library, 1956), p. 42.

3 Jeffrey L. Pressman and Aaron Wildavsky, eds., *Implementation*, 3rd ed. (Berkeley, CA: University of California Press, 1984), p. 254. Emphasis added. Angela Browne co-authored chapters 9, 10 and 11.

4 The two MDRC research projects considered in Part III on evaluation research, the evaluations of the California work/welfare program and the Job Training Partnership Act (the latter is conducted jointly with Abt Associates), are the only efforts I know of that used random assignment in evaluating a large-scale ongoing social program.

5 Robert A. Scott and Arnold R. Shore, *Why Sociology Does Not Apply: A Study of the Use of Sociology in Social Policy* (New York: Elsevier, 1979), pp. 12, 28.

6 Thomas F. Pettigrew, "Can Social Scientists Be Effective Actors in the Policy Arena?" in *Social Science and Social Policy*, eds. R. Lance Shotland and Melvin M. Mark (Beverly Hills, CA: Sage Publications, 1985), p. 132.

7

Evaluating the California GAIN Program

California's 1985 welfare reform law setting up the Greater Avenues for Independence program (GAIN) reflected an intricate political balancing of legislative interests and concerns rooted in San Diego County's experience, the subject of an earlier demonstration study conducted by MDRC. The essential purpose of this 1985 statewide legislation was to convert the administrative system for the Aid to Families with Dependent Children (AFDC) category of welfare from a cash payment process into a service system for job preparation and work facilitation in order to reduce dependency. Family heads whose youngest child was six years of age or older were required under the GAIN law to participate in this program if child care was available. (Welfare family heads with children under six years of age could volunteer to participate in the program, but were not required to do so.) The authors of the 1985 California legislation specified each component of the GAIN process in excruciating detail in order to win and hold the political support of a fragile coalition of liberals and conservatives.

The two leading players in crafting the legislation were David B. Swoap, secretary of California's Health and Welfare Agency, and State Assemblyman Arthur Agnos. Swoap was the principal negotiator for Governor George Deukmejian, a conservative Republican who served as California's Governor for eight years. Previously, Swoap held a similar job as director of welfare in California under Ronald Reagan; he also served in Washington in the Reagan administration as undersecretary of the Department of Health and Human Services. Swoap's counterpart in the negotiations, Arthur Agnos, was a Democratic member of the state legislature who represented San Francisco when the GAIN program was enacted. Agnos later was elected mayor of San Francisco in 1987, serving until 1992.

Shortly after the California GAIN law was enacted, Swoap described the delicate political balancing act involved in its enactment by saying the program "incorporates a unique blend of what have traditionally been considered 'liberal' and 'conservative' attitudes towards caring for the poor." Swoap credited the San Diego County work/welfare demonstration conducted by the Manpower Demonstration Research Corporation with showing the efficacy of the GAIN approach. "The seeds of California's reform were planted over the last three years, by the success of San Diego County's Experimental Work Project."[1]

The debate on the GAIN program in the California legislature focused on participation requirements as the key to the political compromise of the new program. Again according to Swoap, "the opposition of many legislators to the mandatory component began to fade as results from San Diego showed that the overwhelming majority of participants themselves felt the program should be mandatory, because, had it not been, they never would have participated, and acquired valuable training and experience."[2]

Although Swoap, on behalf of the Deukmejian administration, underscored the obligational character of the GAIN program as embodied in the San Diego work/welfare demonstration, Agnos and other members of the legislature had a different set of values and preferences. They stressed the services provided, and drew heavily on Massachusetts's E.T. (Employment and Training) Choices program sponsored by Governor Michael Dukakis and initiated two years prior to the enactment of the GAIN program in 1983. Dukakis publicized this program widely, stressing that it involved *voluntary* participation. Recipients had a "choice" of what "E.T." services they wanted to select.[3] A delegation of California officials, including Swoap and Agnos, visited Massachusetts while the California legislature was working on the GAIN legislation. They also visited other states with more conservative welfare programs.

The GAIN Process

The legislation establishing the GAIN program had an elaborate step-wise structure. The description of the steps dramatizes the research task MDRC faced in mid-1986 when the corporation entered into a contract with the state to evaluate the GAIN program, a contract negotiation in which this author played a role.[4] In effect, this was an interesting two-step dance by social policy researchers. First, they conducted a demonstration in San Diego based on randomized assignment to test a new program focusing on the welfare system providing work. Now they were being asked to evaluate a similar new program after it was enacted in a statewide law to do the same thing.

The first requirements of the GAIN process were registration and orientation. All welfare family heads with children six years of age and older were required to register. The second step was applicant appraisal.[5] Applicants who passed literacy and skills tests were expected to engage in a job search either on an individual basis or in a group setting in a so-called "job club." (A "job club" is group work-preparation and job search often in a classroom setting using phone banks and sometimes video teaching techniques.) For eligible applicants who failed literacy or skills tests as step one, the law required that remedial education or special language training be provided.

For those GAIN participants who did not secure employment after a job search or after they had gone through a remedial education or an English as a Second Language (ESL) program, a service plan was required to be drawn up and signed by the participant and a caseworker. These contracts embodied the so-called "mutual-obligation concept" central to the GAIN program. This concept, in effect, involved a bargain between the client and the state, whereby recipients of aid had to agree to participate in services tailored to their needs and in turn the welfare agency had to agree to provide these services. Determination of the types of services appropriate in a given case in California was to be based on a needs assessment, also spelled out in detail in the GAIN legislation. The law prescribed a specific time period for job search. If participants were not employed after this time period, they then could be assigned to a community work-experience job.

Work-experience jobs, again as specifically described in the GAIN law, were called PREP, standing for pre-employment preparation. There were two types of PREP. One was "basic PREP," which provided work experience that was supposed to help participants obtain references that could assist them in obtaining unsubsidized employment. The other was "advanced PREP," which was supposed to enhance their job skills.[6] The law stipulated that PREP assignments could be short term (lasting up to three months) or long term (lasting up to a year) and that work assignments could not exceed thirty-two hours per week. Publications describing the GAIN program included the flowchart on page 115 showing the steps and branches in the GAIN process.

The MDRC Evaluation

MDRC began its evaluation of the GAIN program in March 1986, six months after the enactment of the legislation. The corporation's first report on the planning and early implementation of the GAIN program was issued in April 1987. It characterized the new program as "one of the broadest arrays of services and support ever offered in a welfare employment program."[7] The implementation process involved literally thousands of public agencies and nonprofit organizations at the state and county levels. According to MDRC, the lead role assigned to the state's Department of Social Services was "to harness the resources and expertise available from the community colleges, adult schools, regional occupational centers and programs, child care agencies, Job Training and Partnership Act (JTPA) programs, and the local offices of the employment services (administered by the Employment Development Department in California)."[8]

The first MDRC implementation report on the California GAIN program identified four features that distinguished the program: (1) its wide array of services; (2) the mandate to participate, projected to involve more than 200,000 welfare family heads by 1990; (3) the

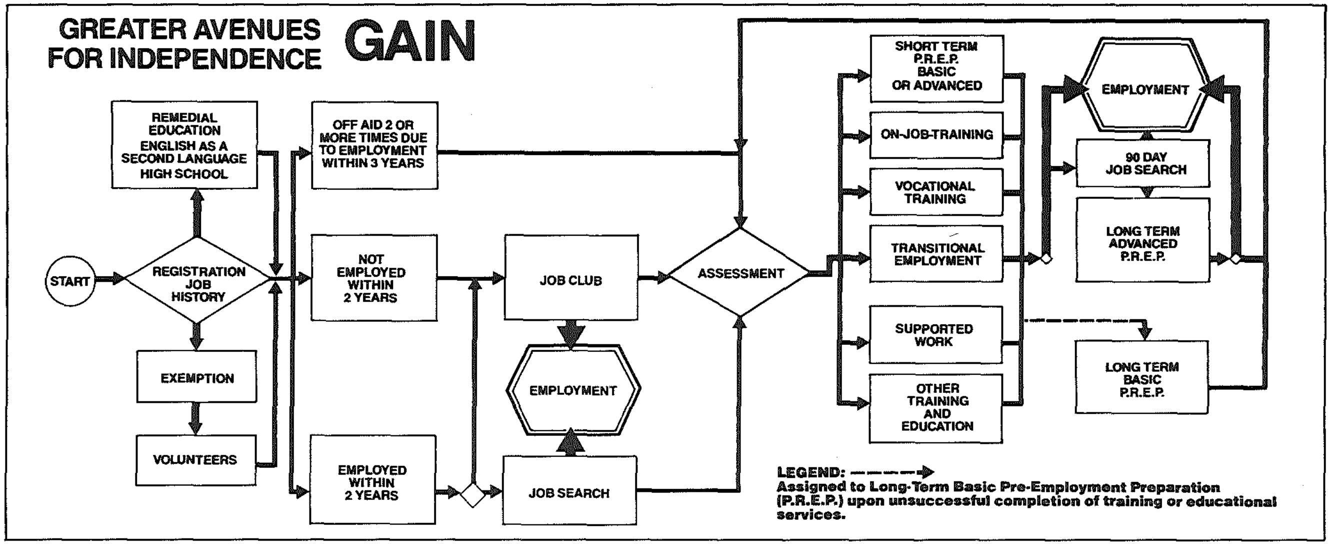

FIGURE 7.1

GAIN Flow Chart

Source: State of California, *GAIN Participant's Manual,* sec. 4, p. 5.

individualized character of participation; and (4) the anticipated continuous nature of the participation process.

The fourth feature of the GAIN program, the requirement that it be operated as a continuous process with no unassigned pool of registrants (as had been the case of many past welfare employment programs), has to be viewed as a major policy departure on a statewide basis. There almost always was a large "holding" category in most states under the federally aided Work Incentive Program (WIN) enacted in 1967. (This aspiration of continuous and obligational GAIN participation had an important effect on the federal laws passed in 1988 and 1996.)

Referring to this commitment to provide continuous services on a mutual-obligation basis to all eligible participants until they exited from welfare, the MDRC early-implementation report noted that "welfare payments to this portion of the caseload would no longer be an unconditional entitlement, but become, instead, a reciprocal obligation."[9] When the GAIN program was enacted, it was projected that when it was fully operational employment and training services would cost $335 million per year. This represented almost as much spending for employment and training services associated with AFDC in one state as there had been for the whole country under the federally aided WIN program at the peak level of its spending.

Adding to the challenge of implementing the goals embodied in the GAIN legislation is an important point about governmental structures going back to the discussion in chapter 6 of the federalism barrier reef. Welfare is administered by *county* governments in California. There are fifty-eight counties in California, ranging in population from Los Angeles with over eight million people (the largest county in population size in the nation), to rural Alpine County in northern California with 1,113 people in 1990. The law establishing the GAIN program gave counties two years to develop their plans, which were subject to review and approval by the state. The goal was to have the full statewide caseload phased into GAIN over five years, from 1985 to 1990.

The initial MDRC report on GAIN was completed as the first nine counties were beginning to participate in the program. The authors of the report had to walk a fine line. They found the state Department of Social Services to be taking a "strong leadership role." They also found "broad support" for the program and enthusiasm on the part of county welfare directors "about playing a leadership role that could reduce the stigma attached to both welfare agencies and recipients." But despite these positive findings, the report indicated that the participating counties had discovered the planning process to be "far more complex, demanding, and time consuming than originally envisioned."[10]

The Research Challenge

When the earlier edition of this book was published, the implementation of the GAIN program had been under way for two years. Even then, in mid-1987, the lessons from this experience reflected the wisdom of Aaron Wildavsky's observations about implementation being a discovery process. The statute establishing the GAIN program indicated the way the program was supposed to work. But is that what actually happened? Would the prescribed services be provided on an interconnected, sequential, and continuous basis that eventually would reach all of the intended participants?

In all of its demonstration studies, MDRC researchers collect what they refer to as "process data" to understand the "black box" of program operations. But for the GAIN study, this part of the task was a bigger deal than usual. The early phase of the research was exclusively a process study. It was not possible, as in the smaller confines of a controlled setting in a demonstration study, to immediately initiate research on impacts on individuals. Later on as described below, a study of the impacts of the GAIN program on individual participants was conducted as a component of MDRC's full GAIN research plan.

The MDRC California GAIN research plan stands out for the way it included three components — an implementation or process study (I prefer the term "implementation study") an individual-impact study, and a cost-benefit analysis. Earlier, in describing evaluation studies, I said I did not know of previous studies of the social programs adopted as part of Lyndon Johnson's Great Society program (the "War on

Poverty") that involved randomized tests to assess the impact on individuals of an ongoing social program. The GAIN project stands out as an example of a broad-gauged evaluation that looked both at the institutional implementation and the impact of an ongoing program, in the latter case using a random-assignment design. This is the "five-star" version of applied social science. But a caution is needed. Such comprehensive, full-service applied social science is expensive. It takes a long time. It is very hard to do. Final results are likely to come out late in the political cycle. Once a program is enacted and it is ongoing, research on its effects usually involves faster, cheaper, less elaborate research designs than that adopted for the California GAIN program. As a general rule, I do not advocate this kind of full-service, applied social science in the fast-paced, complex world of ongoing programs in which there are so many and constantly changing variants of every public policy under the American sun! The federalism barrier reef discussed earlier is my main reason for this conclusion. Another reason is the service-contamination hurdle described in chapter 4, which is very high — often too high to surmount — in complex program settings like that of the California GAIN program.

Implementation as studied in the GAIN program had several dimensions. One was political and organizational: Did the agencies that were supposed to carry out the requirements of the new program do so in a way that resembled what was envisioned in the legislation? There is also a second dimension, what one might call the psychological or attitudinal dimension: Did the attitudes of the people in the agency and of the recipients of welfare benefits change in the ways they were supposed to change? Did county governments and key groups in the state make it clear (I like to call this *signaling*) what was expected of welfare recipients, and did recipients understand and accept these new signals?

Actually, the full MDRC research plan (studying participant progress, gauging staff and participant attitudes, measuring individual program impacts, and conducting a cost-benefit analysis) was not carried out on a statewide basis. Rather it was conducted in selected counties. The implementation research was conducted in eight of the first ten counties to start up the GAIN program. These were designated as "Tier I" counties. Another group of six "Tier II" counties (surprisingly from a design perspective, with only one overlap county) was the subject of the individual-impact part of the study. Aggregate data were collected for the rest of the counties in the state. The individual-impact study with randomized assignment involved more than 33,000 people in the Tier II counties.

The implementation portion of the GAIN study was based on interviews and on program and survey data. The bulk of the interviews were of administrators in county welfare departments. A smaller number of interviews were conducted of the staffs of education, training, and child care (mostly nonprofit) organizations, and other service providers. This research, to MDRC's credit, was multidisciplinary. MDRC staff worked with sociologists, social psychologists, political scientists, and specialists in organizational behavior and program management in the design and execution of this evaluation. The study of the participant flow through the GAIN system focused on data about the types and length of the services provided.

Initially, it was hoped that there would be a uniform statewide data system on the status of GAIN participants and the content of the services provided. This, as we shall see later on, is one of the most crucial aspects of social programs, both for social policy research, and even moreso for program operations. However, efforts to set up a new uniform statewide data system historically have come to naught in California. In the mid-1980s, they came up against not just a federalism barrier reef, but a brick wall — the long-standing dug-in position of county officials in California about local control. The result for the MDRC research on the GAIN program was the need to work out individualized data systems in each of the research counties.

The components of the MDRC evaluation plan for the GAIN program were designed to reinforce each other. The survey research on the attitudes of welfare personnel, for example, included questions about the tasks performed (how they were conducted and viewed), which in this multifaceted design was also used in the analysis of administrative processes.

Discoveries in the Implementation Process

The findings from MDRC's GAIN evaluation, as stated earlier, confirm Aaron Wildavsky's point about implementation being a discovery process. Education as a component of the GAIN program is a good example of this point. In the planning process for GAIN, state officials estimated that one-fourth of the participants would need remedial education. As the first counties began to operate under the GAIN program, however, it became apparent that these estimates were way off the mark. This discovery was made in the development of the testing instruments for the GAIN program.

Literacy and skills testing was an early prescribed step in the GAIN process. Examinations were to be administered to all participants. Test instruments were developed by the California departments of Education and Social Services, and pilot tests were administered in five counties. Over 6,000 people participated in the pilot tests between July and December 1986.

A report on these tests was issued in April 1987. It indicated that 57 percent of the people who took the pilot tests required some form of remediation in literacy or arithmetic skills. The purpose of these tests was to validate the exams used. According to Carl Williams, director of the GAIN program, the results of the pilot tests showed that the "testing instruments selected will perform successfully."[11] Neither this statement nor the report released to the press at the time commented on the large size of the group that required remediation. However, newspaper reporters used these results, not primarily as a validation of the tests, but as an indication of the seriousness of the problem of inadequate literacy and arithmetic skills on the part of welfare family heads. In one newspaper account of the results, Carl Williams was quoted commenting on the broader significance of these results.[12] The GAIN program, Williams observed, needed some remediation of its own. "The message is pretty clear that we have one heck of a population out there that's in need of remediation." The welfare system, said Williams, "is a holding area for people who did not get a good enough education."[13] This was not a surprise to experts in the field. Gordon Berlin and Andrew Sum had earlier described "the basic skills crisis" as the single most important national social problem. "Inadequate basic skills — the ability to read, write, compute, and communicate — is a common thread running throughout the web of social problems."[14]

Publication of these test scores had a big impact. At the highest levels of state government, it caused concern in the legislature, particularly among conservative members who had supported the GAIN program and were worried about the cost implications of the test results. Public programs in sensitive areas like this are constantly subject to scrutiny, debate, *and change*. This is true at both the state and local levels. As it turned out, the ultimate effect of the controversy about test scores was a good outcome for supporters of the GAIN program. An additional appropriation of $41 million was provided, with support from Governor Deukmejian, for remedial education.

This incident also cast light on the crucial role schools played (or didn't play) in the GAIN program. The linkage between welfare and education was found to be problematic at precisely the time remedial education was discovered to be even more important than had originally been anticipated. The GAIN law required that in counties in which there were unused funds that had been allocated by the state to local school districts under the state's education aid program, these funds should be used for remediation under GAIN. On the surface this sounds logical. In effect, the law said that rather than having the state appropriate new funds for remedial education, counties should use available unused funds already allocated to school districts. The "Catch-22" was that school districts didn't like this idea at all. Furthermore, they were not used to serving adult welfare recipients.

Timing was a glitch too. Schools operate on a September to June calendar. GAIN participants who were referred to local schools toward the end of the school year (for example, in March or April) for remedial education were told that class space would be available in September. This defeated the whole idea of continuous participation with no unassigned pool of participants. A thirty-year-old welfare mother was unlikely to believe (and understandably so) that GAIN was a new regime signaling and entailing a serious and active commitment to her future if she was told that there would be a five-month delay between her first experience under the program and the availability of its services.

This problem constituted a special challenge in urban areas. In Santa Clara County, one of the first urbanized counties to enter the GAIN program, the press reported that the failure rate for the literacy and skills test was 76 percent. The challenge involved in integrating the school and welfare systems was brought to light in a press account of one recipient's experience.

> Nancy, a 37-year-old mother of three boys who has been on welfare for five years, said she has refused to join the workfare program because of her job experience. The San Jose woman, who said she has been a waitress and a candy store clerk, asked that her last name not be used because she feared that her welfare check would be cut off.
>
> "I went down one day, and they had an (orientation) class for me," she said. "Then I was supposed to see a social worker a few days later. But no one called me for about three weeks. When they finally called they said I had to learn how to read and write better to get a job."
>
> "Listen," she said. "I've worked before. I have three sons. I haven't been to school since I was 15. It's too late for me to learn, but no one down there will listen to me. I can work; I know I can. But they say they have their rules and that's that."[15]

Other discoveries made in the GAIN implementation process provided useful (though not always appreciated) lessons. An MDRC report on implementation, for example, uncovered a problem of "no-shows" when welfare family heads were referred to the GAIN program. The report described the lower-than-anticipated enrollment levels as follows: "In some cases it was 40 to 80 percent below projections early on, and as high as 25 to 50 percent after several months of operation."[16] Some explanations for this were obvious. In the past, requirements that welfare family heads show up for counseling and service programs (and there had been many such requirements in the law for a long time) were not enforced, and both local welfare workers and recipients knew this to be the case. Under the GAIN program, welfare workers were told to refer recipients to the program, and to stress that their attendance was required. The workers had to be convinced that this was serious, which of course is a familiar bureaucratic challenge that is easier to state than effectuate.

In short, it is extremely difficult to change the culture and behavior of large public bureaucracies. Furthermore, when it happens, and it does happen, it is often uneven and varied in character and intensity. Attention to, and the acceptance of, changed signals (people must work or be trained) meant that ingrained attitudes had to be changed both for the people who administered the program and for the people who received aid under it. It was early evident in the GAIN implementation process that big efforts (really big!) would be needed to change deep-seated bureaucratic and personal attitudes and procedures in this way. Passing a law was only the beginning.

MDRC's evaluation of the California GAIN program lasted for over a decade and, as stated earlier, covered the research waterfront. It included a randomized impact study and a cost-benefit analysis. Three years out, randomized follow-up data for 33,000 people showed a 22 percent gain in average earnings and a corresponding 6 percent reduction in payments.[17] The most impressive results were found for Riverside County, California (one of MDRC's six "Tier II" counties), which had long been, and continued afterwards to be, a showplace county in the nation for the "culture change" of welfare reform. Many government leaders, Presidents Reagan and Carter included, and journalists, traveled to Riverside to hear about this county's "success story" in creating a new job-focused regime for social programs. Correspondingly and in a reassuring way, MDRC's cost-benefit results for Riverside County were described as "exceptionally large" — a $2.84 return for every $1 invested. Three of the six counties in which the impact research was conducted showed positive results in these terms. *Three did not.* In describing these results, the MDRC report said, somewhat defensively: "It is worth mentioning that return per net dollar invested is a standard of success by which few social programs are assessed."[18] It is important to note that MDRC's impact findings from the GAIN research were not published until well after the Family Support Act, the national welfare reform law passed in 1988, had been enacted, which largely mirrored the California GAIN law. Again, as Gilbert Steiner observed, "research followed reality."

From the vantage point of applied social science, several observations should be made. One is that we needed to know the implementation story. Second, the nice model of MDRC's full-service research design for the GAIN study sets an extremely high standard. In part owing to the costs involved, and in part owing to research-timing considerations and the political hurdles discussed earlier, it is not a standard that can be easily applied. In the case of the 1988 and 1996 national welfare reform laws described in the next chapter, we shall see that evaluation studies of many types and under many auspices have been conducted to assess both the institutional and individual effects of these two national welfare reforms. Their results are important, but one would have to say, limited — that is, limited in terms of how much actually can be known about what governments do and achieve in the complex lives of their citizens.

Endnotes

1 Statement made by David B. Swoap, Secretary, Health and Welfare Agency, State of California, Sacramento, September 24, 1985.

2 Ibid. The San Diego work/welfare study conducted by MDRC included a survey asking participants about their attitude toward a work requirement. Recipients interviewed six months after entering the work experience component of San Diego's work/welfare demonstration were asked how fair it was that "you must accept a job assignment in order to receive or keep receiving [welfare] benefits?" Of the 49 respondents, 84 percent said, "It's perfectly fair." A larger sample in a related survey produced similar results. Of 311 persons interviewed on the general question of the fairness of a work requirement, 69 percent of those aware of the requirement said it was fair. See Goldman, Friedlander, et al., *California: The Demonstration of State Work/Welfare Initiatives, Findings from the San Diego Job Search and Work Experience Demonstration* (New York: Manpower Demonstra-

tion Research Corporation, May 1985) chapter 4, "Program Requirements and Noncompliance."

3 See Richard P. Nathan, *Turning Promises Into Performance: The Management Challenge of Implementing Workfare*, (New York: A Twentieth Century Fund Book, Columbia University Press, 1993), pp. 39-43, 97-106. The Massachusetts welfare reform program, along with Massachusetts's booming economy in the mid-1980s were keys to Dukakis's nomination in 1988 as the Democratic Party's candidate for President.

4 At the time I chaired MDRC's board and was on a sabbatical leave from Princeton University at the University of California at Berkeley working on the original version of this book.

5 I treat GAIN in the past tense because it was later replaced with a new program under the 1996 national welfare reform law enacted under President Clinton.

6 John Wallace and David Long, *GAIN: Planning and Early Implementation* (New York: Manpower Demonstration Research Corporation, 1987), p. 29.

7 Ibid., p. v.

8 Ibid., p. vi.

9 Ibid., p. ix.

10 Ibid., pp. xiii, xvi.

11 "GAIN News," Sacramento, California, April 24, 1987, p. 1.

12 Carl Williams was especially cooperative with and supportive of MDRC's research. In fact, it was Williams who originally broached the idea of an MDRC evaluation of the GAIN program in 1985.

13 Richard C. Paddock, "State Finds 57% on Welfare Lack Basic Job Skills," *Los Angeles Times*, April 27, 1987, p. 1.

14 Gordon Berlin and Andrew Sum, "American Standards of Living, Family Welfare and the Basic Skills Crisis," New York City, Ford Foundation, December 1986.

15 Gary Richards, "Three-Fourths of Those in Workfare in Santa Clara County Lack Skills," *San Jose Mercury News*, July 26, 1987, p. 1-A.

16 Wallace and Long, *GAIN*, p. 185.

17 James Riccio, Daniel Friedlander, and Stephen Freedman, *GAIN: Benefits, Costs, and Three Year Impacts of a Welfare-to-Work Program* (New York: Manpower Demonstration Research Corporation, September 1994), p. xxi.

18 Ibid., p. xxiii.

8

The 1988 and 1996 National Welfare Reform Laws

The Family Support Act of 1988

Having tasted victory enacting welfare legislation as part of the budget act in Reagan's first year as president, four years later, when he had just been re-elected and his popularity was at its highest point, administration officials decided to go to the well again for work-oriented welfare reforms. In his 1986 State of the Union message, Reagan called upon Congress to cooperate with his administration to enact welfare legislation "to escape the spider's web of dependency." Quoting Franklin Roosevelt, Reagan described welfare as "a narcotic, a subtle destroyer of the human spirit." He directed the White House Domestic Policy Council to conduct "an evaluation of programs and a strategy for immediate action to meet the financial, educational, social, and safety concerns of poor families." He said: "I'm talking about a real and lasting emancipation because the success of welfare should be judged by how many of its recipients become independent of welfare."[1]

This time Reagan intended more far-reaching changes. Soon, however, prospects dimmed for this. The administration's standing was tarnished by accounts of the President's role in the Iran-Contra controversy.* Also, the House and the Senate were sharply divided.

---

* The Iran-Contra affair involved covert operations coordinated by National Security Council staff to obtain the release of American-held hostages in the Middle East by selling U.S. weapons to Iran, despite the embargo on such sales. This operation, from 1984 through most of 1986, involved secret U.S.

Both were controlled by the Democratic party. In this setting, Reagan's White House staff decided that at the outset of the 100th Congress it would be best to seek modest instead of more sweeping changes, essentially expanding already-existing authority for state demonstrations of work-oriented approaches to welfare reform. But this was not the road taken. Instead, the Family Support Act of 1988 consisted of an amalgam of new substantive authority and associated funding. The key power broker in working out the ultimate political compromise was Daniel Patrick Moynihan, now chair of the welfare subcommittee of the Senate Finance Committee.

With the administration officially supporting welfare reform and Moynihan working on a compromise, the spoiler role at the outset was played by the House of Representatives. Democratic leaders in the House refused to work with, or even talk to, Republicans. They adopted a bill containing major liberalizations of welfare benefits along with a generous new program of grants to the states for training, education, child care, and other services. The employment and training provisions of the House bill were tightly prescribed. This engendered opposition by governors who wanted flexibility for the states.

At this point, Moynihan stepped up his efforts working with the governors. The Senate enacted a bill that took a middle-of-the-road position between the president and the House. It provided flexible grants to the states for employment and training programs and also for other services to welfare recipients, proposed limited benefit liberalizations as compared with those proposed by the House, and called for strict requirements for child support payments by absent (noncustodial) parents, mainly the fathers of children on welfare.

In the negotiations, the bill that emerged most closely resembled the Senate bill. Seemingly, the key to the agreement was a compromise on the Aid for Families with Dependent Children (AFDC) program for two-parent families, called AFDC-UP (the "UP" standing for

governmental support of the Contra military and paramilitary forces in Nicaragua fighting the Sandinista regime. This was after the Congress had ended aid to the Contras (in October 1984). Funds generated from the sale of weapons to Iran were diverted to support the Contras in Nicaragua.

unemployed parent). Democrats in the House and Senate wanted to make the provision of welfare cash benefits to two-parent families mandatory on the same basis as aid to single-parent welfare families But officials of the Reagan administration (especially White House aide Charles Hobbs) would not go along. The compromise extended the AFDC-UP program to all states, with a requirement that one parent in each covered family engage in community work experience for at least sixteen hours per week. Although this work obligation for the heads of two-parent families was ostensibly the reason the administration went along on the final bill, its impact was limited; it affected only a small portion of the caseload. The more important basis for compromise was the concept of "mutual obligation" in much the same way it was embodied in the California GAIN program.

Under the Job Opportunities and Basic Skills Training (JOBS) program, the most important programmatic component of the 1988 Family Support Act, states were required to achieve a 20 percent level of program participation for all eligible AFDC family heads on a phased basis over a period of five years. The JOBS program provided $1 billion annually to the states in the first four years (fiscal years 1991-1994) for training, education, job search, assessment, and other supportive services, and $1.3 billion in the fifth year. The 20 percent participation requirement applied to all welfare family heads who had children three years of age or older. The compromise on this age level was reached after considerable debate similar to the way California legislators had brokered this issue four years earlier.

The 1988 Family Support Act contained authority for Community Work Experience (called CWEP) as in Reagan's 1981 budget.[2] However, unlike the way this was done in the case of two-parent welfare families for which the law mandated up to sixteen hours of participation in CWEP by one parent, the law left the decision as to the required amount of participation (if any) in CWEP by female welfare family heads up to the states. The inclusion of education and training as eligible activities under the JOBS program and the law's emphasis on young welfare mothers going back to school were major themes of the 1988 law, which can be described as adopting the human-capital approach to welfare reform, stressing education and training. It is ironic that this was happening at precisely the time the GAIN program in California was revealing the pitfalls of the relationship between schools and welfare agencies. The Family Support Act required states to devote at least 55 percent of their federally aided JOBS expenditures to the "hard-to-serve" part of their caseload, i.e., unwed teen mothers and long-time welfare recipients.

A close reading of the 1988 Family Support Act suggests that both sides achieved something. Liberals won additional funding for education and training for welfare family heads. They also won on another important point, the extension of child care and Medicaid benefits for one year after a welfare family head entered the labor force and was no longer eligible for cash assistance under AFDC. These were called "transitional benefits." Conservatives on the other hand put their stamp on the bill in the way work requirements and mandatory participation were highlighted.

President Reagan signed the Family Support Act with a flourish at a ceremony in October 1988 in the White House Rose Garden attended by governors and other guests. "This bill, H.R. 1720," the President said, "represents the culmination of more than 2 years of effort and responds to the call in my 1986 State of the Union Message for real welfare reform — reform that will lead to lasting emancipation from welfare dependency."[3]

Before the ink was dry on Reagan's signature, however, ideological differences emerged. The law, as indicated, provided one year of transitional benefits for Medicaid and child care after a welfare family head went to work and earned enough money so that she was no longer eligible for welfare benefits. This provision, which was not a matter of contention in the Congress, quickly emerged as controversial. In an article in the *Wall Street Journal*, which appeared on the very day Reagan signed the bill, conservative welfare expert Charles Murray took aim at this entitlement, which he claimed would cause more people to enter the welfare system. According to Murray, "Once again, we have adopted a policy on the basis of people who already exhibit the problem we want to solve, while being blind to the effects of the policy on people who do not yet exhibit the problem."[4]

Although the Achilles heel of the 1988 law did not turn out to be the one identified by Murray, the act disappointed its proponents for two reasons. One reason, as stressed in the next chapter, was that its execution was pretty half-hearted; it fell far short of the law's spirit and main purposes. A second reason was that the idea of emphasizing the human-capital approach (featuring education and training) did not live up to expectations. Eight years later, out of frustration with past efforts, Congress enacted a stronger, bolder, substantially different national welfare reform law.

The Personal Responsibility Act of 1996

In this book, the 1996 welfare reform act (officially the Personal Responsibility and Work Opportunities Reconciliation Act) is referred to as the "Personal Responsibility Act." These first two words are key. Policymakers were making an effort to change the way society defined personal responsibility for poor family heads. The focus was on *work first.*

However, in federalism terms, the new law is schizophrenic. Did it liberate the states or tie them down? The law is 250 pages long. It contains detailed behavioral requirements with respect to work and school responsibilities for children in welfare families. At the same time, it provides block grants to the states to give them increased fiscal and program flexibility. I view the law as essentially devolutionary, passing power down to the state and local levels, but the answer to the question about its effects on U.S. federalism is by no means cut-and-dried. In this setting in which the buck is being passed down, important questions are raised about the roles of the federal government, the states, localities, and service providers (both nonprofit and for-profit) in implementing the 1996 law.

The Personal Responsibility Act limits to five years the time that a family head can receive federal funds for cash assistance, with states being permitted to exempt 20 percent of their case load from this requirement. Under the new law, welfare is renamed the Temporary Assistance to Needy Families (TANF) program. The old program, AFDC (Aid for Families with Dependent Children), is no more.

The history of the new law is interesting. President Clinton in his 1992 campaign for the presidency promised to "end welfare as we know it." He said this over and over again. Clinton featured this phrase in TV spots, in his Inaugural address, in his first Sate of the Union message, and on many other occasions. As president, he did end "welfare" semantically in the 1996 Personal Responsibility Act. The word "welfare" doesn't appear in the law, and as far as I know, no states use this word in the name of the agency that administers the TANF program for temporary cash assistance and the related job-facilitation programs and services. The American Public Welfare Association representing state welfare officials changed its name in 1998 to the American Public Human Services Association.

When the Republican-controlled 104th Congress finished its handiwork on the 1996 welfare reform act, it was touch and go as to whether President Clinton would sign it. The Personal Responsibility Act, featuring the block grant approach to federal aid to the states, was the product of a Congress with both bodies controlled by the opposition political party. Inside the administration, there was an intense debate for the soul of the President on this bill, with liberals strongly urging a veto, despite the President's campaign promise to end welfare and despite the fact that 1996 was a presidential election year when it would be awkward for him to veto legislation that was supposed to achieve a goal he had so vociferously championed for four years.

The Personal Responsibility Act, which of course Clinton did sign, sets work requirements that are ratcheted up over time so that when the law is fully implemented, half of all single parents must be working. In two-parent families, at least one parent is required to be working in 90 percent of the assisted families. The law tightly defines work activities, spelling out twelve qualifying types of activities. Also tightly defined is the way caseload calculations are to be made for fulfilling these and other participation requirements. Unwed teenaged parents are required to live at home or in an adult-supervised setting and to be in school in order to receive benefits. There are stringent requirements for establishing the child support responsibilities of noncustodial parents and collecting payments from them, including a requirement that states suspend the driver's and occupational licenses of individuals overdue in their child support payments. Sanctions accompany all of these requirements. For immigrants the law was especially strict, denying eligibility to large classes of people — provisions which later were significantly rolled back.

At the same time that the new law sets these and other personal behavioral requirements, it provides block grants to the states. In doing so, the TANF program eliminates the states' entitlement to federal payments for a fixed share of all welfare cash benefits, as was the case under the predecessor AFDC program. (Under AFDC, states were entitled to a percentage of all AFDC payments they made, the percentage depending on state personal income levels. The minimum state matching requirement was 50 percent in better-off states, and as low as 25 percent in the poorest states.)

Under the Personal Responsibility Act, states also receive a child care block grant, and are given the flexibility to transfer up to 30 percent of their TANF block grant funds to the child care block grant and up to 10 percent of their TANF block grant funds to the pre-existing social services block grant.

This description of the act, while not complete, suggests the spirit as well as the complexity of the new law, which is often treated summarily as a breakthrough in simplicity. The Personal Responsibility Act also represented a shift in management philosophy from a regulatory approach to an approach that the law says is supposed to emphasize results as opposed to administrative-process requirements. It does this by setting reporting requirements for activities like job placement and retention, the premise being that governments and the public can use such data to assess whether states are achieving the goals of the new law. What was formerly the quality-control ("Q.C.") approach to enforcement, whereby states had to report according to detailed federally prescribed categories of administrative processes, was supposed to be eliminated.*

---

* Actually, "Q.C." lives on. This is most clearly the case for the closely related food stamp and Medicaid programs, and pretty much the case, too, for determining eligibility for TANF cash assistance benefits.

For social policy and American federalism, the new law is a big deal. Most observers of U.S. domestic affairs probably would agree that the basic contours of welfare policy since the passage of the Social Security Act in 1935 involved the national government setting a framework for the states that on the whole has been liberal. This framework with its entitlement to the states to match cash assistance payments to poor families undoubtedly caused many politically conservative states to provide aid above what they would have provided on their own to this most controversial needy population, i.e., nondisabled, working-age parents and their children.

In 1996 the ideological tables were turned. The framework of the new law and the pressure it applies from Washington is conservative. Pressure is put on liberal states to terminate welfare benefits for cause or after a required time period in order to put teeth into the work, family, school, and reproduction behavior-changing requirements of the Personal Responsibility Act.

This turnabout should not have come as a surprise. American federalism is not, and never has been, intrinsically liberal or conservative. The ideological character of centralizing forces has varied over time. In the nineteenth century the federal government was dominated by pro-business, anti-government social values. Not until Franklin Roosevelt's New Deal in the mid-1930s was there a shift to a socially activist role for the federal government. Many people who follow U.S. domestic policies have lived their whole lives under this centrist-liberal regime. For them, there was always an easy assumption that liberal activism is dominant in Washington and that more conservative, anti-government forces have their political base in states and localities.

The welfare policies advanced by the Republican-led 104th Congress in Washington represented a turning point in these terms. The most dramatic signal was the five-year lifetime limit for a family head to receive federally aided welfare cash benefits. Welfare, so it was claimed, would no longer be a way of life — no longer a permanent condition.

To me, the most surprising thing about the 1996 law, especially compared to the 1988 law, is how much and how deeply it has penetrated America's governments and changed the character and operations of a wide range of social programs and public agencies. The next two chapters compare the effects of the two federal laws (1988 and 1996). The field evaluation study we conducted of the implementation of the Family Support Act of 1988 is described in chapter 9. A similar field evaluation we are presently conducting of the 1996 law is described in chapter 10.

Endnotes

1 "The President's State of the Union Address, Delivered before a Joint Session of the Congress, February 4, 1986," *Weekly Compilation of Presidential Documents* 22, no. 6, Monday, February 10, 1986. (Washington, DC: U.S. Government Printing Office), p. 138.

2 A number of caveats were added in the House. The law required that after nine months of participation in CWEP, workers should receive "prevailing wage rates" rather than minimum wage. It also prohibited requiring more that twenty hours of participation per week in CWEP from AFDC participants with a child under the age of six.

3 Remarks on signing the Family Support Act of 1988, October 13, 1988, *Weekly Compilation of Presidential Documents* 14, no. 41, Monday, October 17, 1988 (Washington, DC: U.S. Government Printing Office), p. 1313.

4 Charles Murray, "New Welfare Bill, New Welfare Cheats," *Wall Street Journal,* October 13, 1998.

9

Evaluating the Family Support Act of 1988

with Irene Lurie

This chapter on the Family Support Act of 1988 begins with a brief detour on the history of the law, focusing on its roots in public policy research. The description of the evaluation of the act which follows is based on an article by Irene Lurie, one of the two principal investigators for the field network evaluation of the 1988 law conducted by the Rockefeller Institute of Government. Jan L. Hagen co-directed this research.

❖ ❖ ❖ ❖ ❖ ❖

Ron Haskins, a Ph.D. developmental psychologist who served for thirteen years as the main policy advisor on welfare to Republicans on the Ways and Means Committee of the U.S. House of Representatives, played a major role in the enactment of the Family Support Act of 1988. In 1991, he published an article on the role research played in the enactment of this legislation. Overall, he said the role of research was "important but limited." He added that a large number of committee staffers were knowledgeable about the most pertinent public policy research. Haskins cited the Congressional Budget Office, the Congressional Research Service, and the General Accounting Office. Staffers in these offices, he said, "constantly seek out research on issues of interest to Congress and employ all conceivable means of conveying the findings to Congress." In Haskins's opinion, "analysts in each of these organizations are top-notch thinkers, well versed in the art of translating research into policy implications." He added his view, which I agree with, on the need for impartiality on the part of researchers.

> Researchers who would apply their knowledge to public policy have an obligation to avoid advocacy. Unfortunately, social scientists, like politicians themselves, are not neutral on the contentious issues of social policy. They must be watched and their ideas subjected to intense scrutiny by their colleagues and by the political process.[1]

In his article, Haskins discussed the studies of the Manpower Demonstration Research Corporation as well as other studies. To me, the notable thing about Haskins's role is that it is not uncommon. Hundreds of Congressional and agency staffers, in different ways and to different degrees, bridge social science and social policy, thus constituting the policy research movement in action. Haskins's article focused on the House of Representatives. It hardly needs to be added, as many of the events described in this book attest to, that the role of social science in the Senate in this period was appreciable because of the presence of Senator Daniel Patrick Moynihan, a ubiquitous and important figure in almost all of the events described in the book, and himself a noted social scientist.

❖ ❖ ❖ ❖ ❖ ❖

The description of the effects of the 1988 Family Support Act contained in this chapter is based on Irene Lurie's article, "A Lesson from the JOBS Program: Reforming Welfare Must Be Both Dazzling and Dull," which appeared in the Fall 1996 issues of the *Journal of Policy Analysis and Management.* Using a uniform analytical structure, Rockefeller Institute field researchers studied the implementation of the Job Opportunities and Basic Skills (JOBS) program in ten states — Maryland, Michigan, Minnesota, Mississippi, New York, Oklahoma, Oregon, Pennsylvania, Tennessee, and Texas — a representative cross-section of states in terms of their per capita income, poverty rate, level of fiscal stress, and geographical region. The research, conducted between 1990 and 1995, was funded by the U.S. Department of Health and Human Services.

Since 1967, when the Work Incentive Program (WIN) was created, federal law required states to operate employment and training programs for recipients of cash benefits under the Aid for Families with Dependent Children (AFDC) program. Congress intended that the WIN program would require recipients to work or prepare for work. But insufficient funding and inadequate motivation on the part of the agencies operating WIN dashed its expectations, despite repeated efforts at reform. States could impose fiscal penalties or sanctions on all welfare recipients registered for WIN services as a condition of eligibility for assistance unless they were exempted. But the federal government did not set goals for the states regarding actual employment or participation in an employment or training activity. So for many people the WIN program was nothing more than a paper process of registering for services that were never provided.

In drafting the Family Support Act of 1988, Congress expressed its intention that participation in the JOBS program "must be something more than simple registration for the program"[2] To encourage this, the act mandated states to serve a minimum percentage of their caseload, which rose from 7 percent of individuals who were required to participate in the program in 1990, the year of mandatory state implementation, to 20 percent of nonexempt individuals by 1995. Beginning in 1994, the act mandated considerably higher participation rates for two-parent families in which the principal wage earner was unemployed: 40 percent in 1994, rising to 75 percent in 1997. If a state failed to meet these mandates, the federal matching share of JOBS expenditures was to be cut from between 60 and 80 percent, depending on a state's per capita income, to 50 percent.

In writing the federal JOBS regulations, the critical task of defining the term "participation" was left to the federal Department of Health and Human Services (HHS). HHS viewed the participation mandate as an expression of Congressional intent that JOBS be a more "meaningful" program than WIN, not a "token effort," and toward this end defined participation in terms of a twenty-hour-per-week standard.[3] The "twenty-hour rule" set a high goal for the states, one that placed great pressure on them to offer more intensive services.

Three Strategies

In the ten-state Rockefeller Institute evaluation, the mandated participation rate and its twenty-hour rule were found to be at the forefront in the minds of state policymakers and administrators as they shaped their strategy for implementing the JOBS program. State welfare agencies pursued three strategies for achieving the participation goals. The first strategy, simply funding a supply of services for JOBS participants, was most straightforward and gave welfare agencies the greatest control over the design of their program. State welfare agencies used JOBS funds to provide JOBS services in-house or to purchase services from other organizations, generally designing programs to meet for at least twenty hours per week. Many organizations were willing and even eager to serve JOBS participants. The supply of services from local school districts, community colleges, nonprofit community agencies, and for-profit education and training firms was found to be quite elastic in response to increased funding.

But funding for JOBS was scarce. In 1993, only about one-third of the states nationwide spent enough to draw down their full federal allocation of JOBS funds; overall, states drew down 75 percent of the $1.1 billion in federal funds available. In the ten-state study, administrators argued that lack of funding was the most severe impediment to program implementation. In Oregon, which spent enough to draw down all of its federal allocation and even used unmatched state funds, the field researcher argued that "under-funding of the entire program is a persistent and chronic problem."[4]

A second strategy, one of cost shifting, was to obtain resources from other organizations through "coordination," a practice encouraged by the Family Support Act. In all ten states, to varying degrees, welfare agencies shifted costs to other organizations by negotiating formal interagency agreements or striking informal bargains. An array of federally aided programs assisted JOBS participants, including the Job

Training Partnership Act (JTPA) program, the Adult Education Act, the Perkins Vocational Education Act, the Higher Education Act's Pell grants, and guaranteed student loans, along with local public schools, state and local postsecondary educational institutions, and the programs of private nonprofit organizations.

Such interagency coordination, however, is often considered to be difficult, and likely to be hindered by turf battles, conflicting agency goals, differing constituencies, and varied eligibility rules. It is surprising therefore that in the ten states in the study, the degree of coordination emerged as one of the success stories of JOBS implementation. Collaborative planning was extensive. The JOBS program created or strengthened an array of interagency planning groups, task forces, advisory committees, policy committees, management committees, and coordinating councils. In some states, the program produced arrangements for pooling resources and sharing credit for positive outcomes so that each agency saw joint action as being to its advantage or, at least, leaving it no worse off. The combination of a lack of funds for JOBS and the federally mandated participation rate was a powerful incentive for welfare agencies at both the state and local levels to develop formal and informal linkages with other agencies.

The downside of reliance on other organizations for a supply of services was that in many areas the JOBS program lacked control over the availability of services. The people served and the types of services offered were based not solely on an assessment of each individual's need for services, but on the existing infrastructure of organizations and programs. Some sites did not even have the resources to assess all individuals, and instead devised ways to count people engaged in other programs as JOBS participants. In New York City, to take an extreme example, in 1995 two-thirds of JOBS participants were classified as being in "self-initiated" education and training programs that they had located on their own. So, rather than implementing a clear program model, the programs of many states were "service-driven," placing or counting recipients in services that were already available.

Finally, exempting recipients from mandatory participation and thereby lowering the denominator of the participation rate appeared to be a practice in a few of the ten states. Recipients were automatically exempt from participation on account of personal factors such as age, illness, or incapacity; they could also be exempt for quite different reasons such as a lack of state resources for funding necessary child care. Nationally, 56 percent of total adult recipients were exempt in 1994. Among the ten states, this proportion ranged from 36 percent in Michigan and 43 percent in Oregon, states that gave high priority to the JOBS program, to 67 percent in Texas and 78 percent in Tennessee, states that did not. This wide range suggests that some states exempted large numbers of recipients because they did not have the resources to serve them.

Little Fanfare or Rhetoric

The strong rhetoric of the authors of the Family Support Act — that it would instill a sense of obligation in recipients and would transform the AFDC program into an employment and training program that provided income support — raised lofty expectations with respect to cultural change in welfare systems. In implementing the provisions of the JOBS program, however, most of the ten sample states attempted to meet the letter of the law (as stated in the mandates), but not the spirit of the law (as stated by its authors). Most states introduced JOBS with little fanfare or rhetoric about the obligation of recipients to assume more responsibility for becoming self-sufficient, or about a changed role for welfare agencies.

Part of the fascination of welfare programs for policy and management analysts is the challenge of motivating the multiple layers of actors who play a role in determining the success of these programs. Motivation must be instilled at every point in a multilayered system: Federal leaders and policy must motivate elected and appointed state officials to design and fund programs; state leaders must in turn motivate welfare agency administrators at both the state and local levels; local welfare commissioners must motivate their staff to take the program seriously; and, finally, front-line staff must motivate the individuals in their caseloads to participate in these programs, place their children in child care if necessary, and seek and accept available employment.

Opinions differ about how to motivate welfare recipients to prepare for work and enter the labor force. Mary Jo Bane and Robert Behn argue that good management, with a high-energy staff and an employment-focused organizational culture, makes the threat of financial sanctions unnecessary.[5] Yet, we found states operating demonstrations that increased sanctions for failure to comply with the mandate to participate in the JOBS program, suggesting they either disagree with Bane and Behn or could not achieve their vision. Field researchers found that encouraging recipients who are not motivated to participate takes resources, in terms of both staff and JOBS services. Few states hired additional staff for their welfare agencies, reflecting the unwillingness of state and local legislators to devote more funds to what they often viewed as a "bloated" welfare bureaucracy. Training staff to focus on employment was limited, since it took time away from other pressing tasks. No states rewarded staff financially for good performance. Even imposing sanctions can be costly, since a welfare agency must devote staff to monitoring participation and to operating a conciliation and fair hearing process for people who refuse to participate without good cause. Some caseworkers thought the conciliation and sanctioning process was too time-consuming, too much of a hassle, so they made only a minimal effort to follow the rules.

In Riverside, California, one of the sites in the MDRC evaluation of the California GAIN (Greater Avenue for Independence) program, welfare administrators created an agency culture that achieved considerable success in moving recipients into jobs.[6] Welfare administrators from around the country visited Riverside to learn how they did this. Yet Riverside was found to be especially successful. The Rockefeller Institute's ten-state implementation study of the JOBS program, which did not include California, but did include Michigan, found that another MDRC evaluation site, Grand Rapid, Michigan, had a "tough" program, achieved high rates of participation, and was in both respects atypical of the Rockefeller Institute's three local sites in Michigan.[7] MDRC study sites agreed to be evaluated; administrators were confident enough of their programs to showcase them presumably because they had activist programs. In contrast, the sites in the Institute's JOBS study were, with only a few exceptions, selected by the field associates and central staff with the goal of including "garden-variety" programs.

A survey of front-line staff in these sites revealed generally broad support for the JOBS goal of promoting recipients' self-sufficiency, but a rather "lukewarm" agency context for implementing the program.[8] How to instill broad enthusiasm in a frequently overworked and underappreciated staff is the key challenge.

Although instilling a sense of obligation on the part of recipients was a goal of the Family Support Act, the act did little to change the financial rewards or penalties facing them. The only significant change was to reward work by providing transitional child care and Medicaid benefits for people leaving welfare for employment, thereby preventing the total loss of these subsidies when earnings made a family ineligible for welfare. The problems of what economists call "moral hazard," always inherent in the structure of the AFDC program, continued to exist.

A final lesson from the history of JOBS implementation concerns the politics of welfare reform. Governors and welfare commissioners like to take initiative in this area. When federal legislation in the early 1980s gave states the option to operate new welfare employment programs, many state leaders were enthusiastic about designing new programs and taking ownership of them. But when federal legislation required states to implement the JOBS program, embracing services that many states were already providing, JOBS became just another federal mandate requiring program changes and additional expenditures. Rather than devoting political capital to increasing expenditures for the JOBS program, state leaders obtained waivers from federal law to restrict welfare eligibility and benefits as a way of generating political capital. One state welfare commissioner explained that officials stand to benefit from this tough approach because "the public blames the welfare system for problems in their own life." A hard line on welfare policy appeared to offer state leaders an opportunity for political gain. In this respect, the Personal Responsibility Act of 1996 can be seen as a response to the findings of the JOBS program research and other studies of welfare employment programs, research which set the stage for debate about welfare reform strategies that eventually led to the 1996 Personal Responsibility Act.

Endnotes

1 Ron Haskins, "Congress Writes a Law: Research and Welfare Reform," *Journal of Policy Analysis and Management* 10, no. 4 (Fall 1991), pp. 617-631.

2 U.S. Congress House of Representatives, *Conference Report on the Family Support Act of 1988*, Report 100-998, September 28, 1988, p. 148.

3 *Federal Register*, 54 (197), (October 13, 1989), pp. 42,201, 42,203.

4 Jan L. Hagen and Irene Lurie, *Implementing Jobs: Case Management Services* (Albany, NY: The Nelson A. Rockefeller Institute of Government, 1994), p. ix.

5 See Mary Jo Bane, "Welfare Reform and Mandatory versus Voluntary Work: Policy Issue or Management Problem?" *Journal of Policy Analysis and Management* 8, no. 2 (Spring 1989), pp. 285-289; see also Robert D. Behn, *Leadership Counts: Lessons for Public Managers* (Cambridge, MA: Harvard University Press, 1991).

6 The GAIN program is discussed in chapter 7. See James Riccio, Daniel Friedlander, and Stephen Freedman, *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program* (New York: Manpower Demonstration Research Corporation, 1994).

7 Gayle Hamilton, *Executive Summary, The JOBS Evaluation: Monthly Participation Rates in Three Sites and Factors Affecting Participation Levels in Welfare-to-Work Programs* (New York: Manpower Demonstration Research Corporation, 1995).

8 Jan L. Hagen and Irene Lurie, *Implementing JOBS: Progress and Promise* (Albany, NY: Rockefeller Institute of Government, 1994), p. 205.

10

Evaluating the Personal Responsibility Act of 1996

with Thomas L. Gais

The Personal Responsibility and Work Opportunities Reconciliation Act of 1996 (called the Personal Responsibility Act in this book) has two main purposes. It is intended to discourage welfare recipiency by promoting work, and to reduce out-of-wedlock births and teen pregnancies. The first report on the twenty-state evaluation of the implementation of the Personal Responsibility Act conducted by the Rockefeller Institute of Government, based on field research conducted in 1997 and 1998, showed that governors and state legislators adopted enabling legislation for their welfare reforms that strongly discouraged welfare dependency and strongly signaled the importance of work.[1] And they followed through on a basis that often involved wide political and organizational support. We did not, however, find similar widespread political or administrative support for the act's anti-reproduction goals. Legislative proposals and tentative administrative policy initiatives to change sexual behavior were advanced in some states, but often they were dropped, watered down, or de-emphasized before state welfare reform legislation was passed or broad executive orders promulgated.

In the work area, the enactment of the Personal Responsibility Act in 1996 was a result both of the disappointment with the response to the 1988 Family Support Act and the results of research showing only modest effects of education and training (the human-capital approach to welfare reform) and much stronger results from work-first strategies.

The juxtaposition between the impacts of rigorous demonstration studies of the two approaches — one emphasizing education and training and the other emphasizing work — showing the latter to be more effective had a strong effect on the decision in the 1996 welfare reform act to downplay education and focus on work first. Questions about this choice of strategies came up frequently in committee hearings. Relevant research findings were a major subject of the deliberations in House Ways and Means Committee meetings to hammer out the final legislation. In particular, findings on the demonstration supported by the Rockefeller Foundation in San Jose, California, were mentioned, which concluded, "The findings of the evaluation suggest that immediate, job-specific training with a strong focus on getting trainees into jobs is a more effective way to improve the earnings of single mothers than are alternative strategies to seek to improve basic skills before offering job training."[2] Similar findings from the work/welfare demonstrations conducted by the Manpower Demonstration Research Corporation were cited in this context.

The 1996 Personal Responsibility Act seeks to modify two kinds of behavior — the personal labor force and family-forming behavior of poor family heads, and the bureaucratic behavior of the agencies that administer welfare and employment and other job-related social programs. The field network evaluation research conducted by the Rockefeller Institute focuses on the second of these two types of behavior modification — the effort to change the behavior of the *bureaucracy.*

The principal finding of our first look at the implementation of the 1996 act, conducted mainly in 1998, was that a lot was happening. As researchers who have been in this field for a long time, we had not seen (or expected to see) so much and such pervasive institutional change in social programs on an across-the-board basis.

Why did this change occur, especially since it is customary to label bureaucracies as hidebound and resistant to change? There were good reasons to be skeptical. Work requirements had been part of welfare since the early 1970s. The 1988 Family Support Act creating the JOBS program had been in effect since 1990 with little to show for it. Yet things were different under the Personal Responsibility Act.

Say what you like about the five-year time limit applying to federally funded cash assistance, it got the attention of the public and of potential and current welfare recipients in a big way. (Some states adopted shorter time limits, as the law allowed.) Moreover, as opposed to the response to the 1988 act, states did not simply layer on new responsibilities to public employees; in many places, they completely reorganized how they operated welfare and related employment and social services. The face of welfare changed for families as states altered the delivery of welfare benefits and services. The work first philosophy dominated, stressing immediate attachment to the labor force rather than training and education.

The flexible structure of the Temporary Assistance for Needy Families (TANF) block grant created in the 1996 act, fortuitously coming at a time when the national labor market was buoyant, had an immense effect on intergovernmental finances. The decision to use 1994 as the base year for calculating the amount of money to be provided in these state block grants produced a fiscal windfall for all states. This is because caseloads declined by close to 50 percent on average from the 1994 base year; all states had material declines. In addition to receiving TANF block grants from Washington, states are required under the 1996 act to maintain no less than 75 percent of their previous AFDC matching funds to be used to aid the TANF-eligible population, adding this money to the TANF block grant funds available to aid poor families. A fiscal study by the Rockefeller Institute of these new financing arrangements, initially conducted in four states, showed a major shift in spending away from paying cash assistance towards additional spending for child care and employment services.[3]

The organizational implications of the new law turned out to have widespread reverberations and multiple variations. In response to the act, new agency missions and arrangements were adopted. Delivery systems became more complex and diverse, and there was a redistribution of discretion, pushing downward to local offices, and ultimately to case managers. Local offices operating under new institutional arrangements, spurred by the federal block grant, came to have a wide range of tools and services available for assisting families and greater discretion in how to use them. A major consequence was the emergence of considerable diversity in local systems — diversity that has had an important effect on how people are treated, and that is influenced by a variety of factors, such as administrative resources, organizational style, and community capacity and values. This combination of programmatic flexibility and greatly increased local discretion posed huge management challenges for states to create and operate management information systems to provide accurate data to serve local case-management, as well as to meet federal and state needs for assuring accountability. All things considered, the task of moving a large and heterogeneous population out of dependency on public aid and into the workforce created a host of challenges with respect to program and case management, accountability, and program equity.

Changed Signals

We found that workers in social programs were not as opposed to the behavior change purposes of the 1996 act as many people had expected they would be. In fact, front-line workers were often strongly in support of the law's emphasis on work and reducing dependency. As one case manager said, "This is what I was trained to do. Now I have some leverage. I love it." Many ground-level workers themselves have a relatively low income and a strong work ethic, which in many instances are shared by welfare applicants and recipients. Devolution to the front lines is the subject of a special federally funded component of the Rockefeller Institute field evaluation focused on culture change and signaling by front-line workers: Did they get and transmit the new message? Irene Lurie directs this research.[4]

We found that signals about the importance of work and self support changed after 1996, despite the fact that work/welfare initiatives conducted under federal waivers had been under way in many states before the new federal law was passed. However, the other strong policy signaling in the 1996 act — that applying to pregnancy prevention — was found to have had little initial effect on bureaucratic behavior, although there are grounds for expecting that behavior in this area will change, too, as a result of the act. In this area, the change is more likely to be personal than bureaucratic. Poor family heads (most of them female and unmarried) face the new reality of time-limited cash assistance and strong requirements for work and participation in active job search and job-related activities. They have to participate for fixed amounts of time under "Personal Responsibility Agreements" specified in the 1996 law, which they are required to sign before a cash-assistance case can be opened. If there is a non-custodial parent (usually a male), there is a new social dynamic: "If he isn't required to do anything, why should I be; why shouldn't he be responsible too?" Many state officials predicted that the resentment by poor female single parents because of the reality of time-limited cash assistance will ultimately affect child-bearing behavior. There is in fact suggestive but no definitive evidence, that this is already occurring.

For policy implementation, the lessons are obvious. Changes in bureaucratic behavior are more likely to happen when there is a clear policy signal that is strong enough to be widely noticed both within the pertinent political subsystem and in the larger community. But this is not enough. Change, when it occurs, happens more easily and widely when it is in line with the dominant social values. In the case of the Personal Responsibility Act, we found bureaucratic behavior to be changing in one area and not another in a way that reflects the "political correctness" of one signal (the work signal) and not of the other signal about pregnancy prevention. These changes in signaling and bureaucratic behavior occurred in large part because many political leaders, especially governors, stressed that a work-based approach to welfare reform was not just one policy priority among many but the central objective of welfare reform. Furthermore, it commanded support from a broad and enduring political coalition.

One of the intriguing characteristics of the politics of the implementation of the 1996 law was how little debate it generated. Partisan fights erupted in some states and on some issues, but such divisions were more the exception than the rule. The broadest support was for imposing time limits and strict and extensive employment obligations on adult recipients of welfare benefits. As Thomas Kaplan said in describing the Wisconsin field research for this study:

> A striking feature of this abrupt change in approach toward AFDC has been the apparent breadth of acceptance of that change in the state. From the governor's first proposals to reduce AFDC benefits in families with teens not attending school (Learnfare), many (though by no means all) representatives of the inner city poor in Milwaukee have endorsed the changes. Wisconsin's Democrats were the initial proponents of "ending welfare as we know it" in this state, and very little opposition has arisen to the requirement of W-2 [Wisconsin Works] that public assistance cash income must derive only from participation in work or work-like programs. . . . With the prominent exception of the Catholic archdiocese of Milwaukee, the central notion of Wisconsin's welfare reform — that "welfare as we knew it" was bad for both its recipients and the broader society — has been little questioned in this state, despite the sharp difference of this new consensus from the apparent consensus of 15 years.[5]

Wisconsin is the state with the most radical welfare reform, but the story was similar in other states. Despite the fact that the sample states enacted very different welfare laws, most were variants of the work first approach, which stresses moving a large proportion of adult recipients into jobs as quickly as possible. Florida enacted its work first program (entitled WAGES, standing for Work and Gain Economic Self-Sufficiency) with little controversy about the nature of benefits or their tight time limits.[6] The debate over welfare reform was "moderate in both tone and content," and the final bill passed both legislative chambers with unanimous support. In Texas, the policy process was "highly inclusive, deliberative, factually based, and surprisingly bi-partisan." In Kansas, policymakers "generally agreed that the major objective of reform was employment and reform." The political debate in Ohio saw some dissension over family caps and the two-year time limit rather than the federally mandated five-year limit, but after these and a child-support tracking requirement were modified, the final enabling legislation was enacted with only one dissenting vote in both legislative chambers. Broad support was also found for the work first philosophy in Michigan, Arizona, North Carolina, Washington, Utah, and West Virginia.

Not all states avoided conflict over welfare reform. But even where political divisions existed, the focus on jobs commanded broad support. Urban Democrats in Georgia unsuccessfully opposed the stringency of time limits and the strength of the sanctions in Democratic Governor Zell Miller's welfare bill and wanted the state to provide a more expansive safety net, but they did not repudiate the bill's work requirements. The debate in Rhode Island produced conflict over its reforms, and in the end the state rejected time limits and retained the entitlement status of welfare benefits, but work requirements were generally accepted. Flush state treasuries helped in Rhode Island and other states where there was political concern about the employability of parents. Demands that parents work were made more acceptable by giving working parents more help, such as a new entitlement in Rhode Island for unlimited child care and health care for children in families with incomes up to 250 percent above the poverty level.

There were other factors behind the shift in political support to work first. The 1994 elections were important in some states. In North Carolina, for example, Governor Jim Hunt's approach to welfare reform shifted from one that was centered on child well-being to one that emphasized work after the Republicans made large gains in the state legislature. In many states, administrators and policy staff may have been influenced by studies by MDRC and other researchers that showed the relative weakness (or at least lack of decisive strength) of job training and education-based programs in reducing welfare dependency. The five-year time limit in the 1996 act gave work-based reforms heightened seriousness, signaling that income support would no longer be provided on a permanent basis by the federal government for able-bodied, nonelderly, nondisabled parents. If state and local governments were going to support such families beyond the federal time limits, they would have to assume the full financial burden of doing so.*

As an outgrowth of the interest by national and state policymakers in changed signaling and agency culture change, a spin-off added to this

* Among the sample states, Michigan and Rhode Island took the position that the state would fund benefits for the most needy families beyond the five-year limit. In New York, a similar decision was made to do so through a system of vouchers to meet essential needs.

field evaluation, mentioned earlier, is an examination of the behavior of front-line workers in twelve sites in four of the sample states for the Rockefeller Institute's implementation study. This is the ultimate test of a new policy — how it affects people at ground zero. In a presentation on this research at a conference sponsored by the Administration for Children and Families of the U.S. Department of Health and Human Services, Irene Lurie suggested that the old welfare is not so much replaced as that a new one has been added on.

> So we've seen, and to the credit of the agencies we have visited, that to guarantee that the client does get the message about the importance of work that there is a process they go through where a specialized worker — maybe it's someone from the Department of Labor, maybe it's someone from a not-for-profit organization, or a for-profit, or a specialized worker in a specialized agency — has sole responsibility to tell people about work and about the importance of work. And we've seen this virtually every place we've been.[7]

New Partners

One of the most widespread strategies for changing the purposes and signals of welfare, as just noted, was to give employment bureaucracies greater control over program operations. This was done in various ways. Employment bureaucracies were seen as having the expertise, data, and institutional missions for getting people jobs, assessing skills, providing job-readiness training and other training services, working with employers, setting up work-experience opportunities, and finding and overseeing local job service contractors. One way of viewing what was happening here is that political leaders decided that if they wanted to change the signaling of welfare — they had to change the signal*ers*. New arrangements between employment bureaucracies and welfare agencies changed the character of welfare. By locating welfare offices in job centers and/or having employment agencies serve as the front door or first step in the application process for cash assistance for poor families, states signaled administratively that people could and should get a job and that in many places welfare cash assistance was part of a new service system for job seekers, rather than a program unto itself.

Wisconsin was the most active state in transforming administrative structures in order to put welfare in an employment context. It began the process of creating closer connections between its public assistance agency and its Department of Labor (which used to be called the Department of Industry, Labor, and Human Relations) when it co-located county welfare offices with what are called "Job Service" offices in job centers. Co-location gave way to consolidation in 1996 when the state's Wisconsin Works (W-2) program was enacted and all public assistance functions except Medicaid were transferred out of the state's welfare department into a renamed Department of Workforce Development.

Not only were labor and employment bureaucracies given greater responsibilities for carrying out welfare programs, in many states their activities were integrated with cash assistance and social services. This local presence, in terms of staff playing a central role in the program, is much greater than in the past, when such job-stressing efforts were often half-hearted. Georgia was particularly successful in incorporating work activities and expertise in its TANF program. According to the Georgia field researcher, in the AFDC/JOBS era (1990-1995) there was a "complete separation" of eligibility determination and employment services. "AFDC and JOBS were located in separate offices, often several miles away from each other."

> Although there was no stated animosity between these two staffs, state officials acknowledged that the two programs seemed to exist in separate worlds. Eligibility workers appeared concerned only with the determination of benefit eligibility for individuals; employment services were only focused on securing a job for a client. In fact, it ran counter to the organizational interest of local AFDC offices to place clients in jobs — since diminished caseloads resulted in a need for fewer employees.[8]

Formerly, there were eight AFDC caseworkers for every employability specialist in the Georgia JOBS program, creating a situation, as one senior official put it, comparable to "eight people loading the truck and only one person unloading it." In 1997 this situation changed after the Department of Human Resources contracted with the Department of Labor to place 18,000 TANF recipients in jobs and provide training and education services to an additional 40,000 clients. Under this contract, Department of Labor staffers were "physically located within welfare offices throughout Georgia, handling the job placement tasks formerly performed by the Department of Child and Family Services." In some counties, front-line staff were organized into teams; in others, staff were cross-trained as generic workers with proficiency in work assessment and employment monitoring, as well as case management and child care functions. In one local site, there were nearly five workers who tracked work-related activities for every one who determined eligibility and benefits.

Such new partnerships for the job-service strategies of the 1996 law and precursor state work/welfare initiatives put a high premium on *connectivity*. The term refers to three important types of connections — those between programs; these between institutions (i.e., jurisdictions, public agencies, and contracted service providers); and connections between information systems. The third area (information systems) became another special subject for examination by Rockefeller Institute field researchers working jointly with evaluators from the U.S. General Accounting Office. Together, the two organizations have studied six state systems in depth and another nine states less intensively as part of the program of the Working Seminar on Information Systems for Social Programs.[9] Increasingly, the development of information systems for managing work-focused social systems has emerged as critical for policy oversight and program planning, and, as noted, is especially needed by front-line workers to link and track benefits and social services for individual cases.

"Diversion"

Although not mentioned in the 1996 welfare reform law, "diversion" emerged as a major new term of reference and a strategy for administering human services that produced a great deal of expanded discretion and variation in operations at the local and caseworker levels. The term is used in several ways. One way is to refer to one-time cash payments or other inducements given to families in exchange for limiting their eligibility for continuing cash assistance. The term, diversion, is also used to refer to activities intended to have the effect of diverting families from receiving regular cash assistance. The initial field reports showed that this latter form of diversion is much more important than lump-sum payments. As a result, it is not always easy to identify what is diversion and what it is not. An activity like eligibility screening may be a form of diversion when it is carried out in certain ways, whereas sometimes it may simply be part of an eligibility review.

In Florida, as an illustration, two types of diversion were explicit — non-cash and cash. Non-cash diversion involves telling applicants of their responsibility to work if their application is approved and requiring them to look for work before their eligibility for assistance is determined. Cash diversion on the other hand can provide two months worth of cash assistance in exchange for which the client must agree to forego ongoing cash assistance for a specified length of time, the length being set by the local office. However, there was little cash diversion initially in Florida at the outset because the state's information system could not keep track of who had received cash grants in order to be able to prevent people from receiving repeat diversion payments. At the same time, non-cash diversion in one region of the state was reported to have affected as many as 20 percent of the applicants.

As compared with lump-sum cash diversion, states generally make greater use of non-cash diversion on a basis that increases the discretion exercised by local agencies and front-line workers. One of the most elaborate diversion programs we found is called "Texas Works." It aims to get people a job or needed services before they apply for TANF benefits in order to make sure that "the only people coming onto TANF are those who need more in-depth help." Clerical staff may route potential applicants to employment counseling with a Texas Works advisor, to a group intake interview, or to an employment resources center. Advisors may also make referrals to public or private agencies, such as those for family violence intervention, local women's shelters, disability benefits, drug and alcohol treatment, protective and regulatory services (for child abuse and neglect), and various health services. The state's welfare program thus has a new emphasis. The commissioner of the Department of Human Services awards a "Commissioner's Cup" to the region that has diverted the largest number of clients.

In sum, diversion can serve different purposes; actual practices often depend on local interpretation and implementation. Some states use diversion to erect a fortress-like welfare system, instead of expanding the options available to families. In doing so, they use diversion to minimize caseloads — to discourage applicants, whatever their needs happen to be. Because diversion can be ad hoc, dependent on local resources and on front-line workers' judgment, knowledge, and objectives, it has a highly devolutionary effect. Local offices and case managers have decisions to make and leeway in making them. Diversion also produces considerable variation in how people are treated, and therefore poses enormous challenges for monitoring the ultimate effects of welfare and social service programs on families, as families move across a wide variety of public and private institutions.

In some states, we found personal responsibility agreements to be the main tool for structuring the relationships between families and welfare agencies. In other states, these agreements are little more than a standardized list of program requirements and benefits, along with a vague commitment by the state to provide TANF benefits to eligible persons. However, in some cases, these agreements can be specific and highly individualized. Promises may be made by the state to provide particular services, with the resulting mutually signed agreement used to review progress and guide case management. Such agreements are not only important signaling devices regarding client obligations; they contribute to the shift by local agencies from a rule-driven administrative culture to one that is structured around solving problems and preventing dependency. In Ohio, for example, TANF participants are required to sign what are called "self-sufficiency contracts." They specify what kinds of work activities the head of household plans to perform, for how many hours per week, and when the activities will begin and end. Specific expectations may include paternity establishment, securing child support, reporting "everything known" about absent parents, and complying, where relevant, with a reunification plan developed by the Ohio Children's Service Agency. The plans are reciprocal; they are viewed as binding contracts between county agencies and participants. They are taken seriously enough that counties are "strongly encouraged" to seek the advice and approval of county prosecutors if they choose to design their own contractual arrangements, which they may do. Participants are advised of their rights, including the right to assistance in locating employment within the state's thirty-six-month time limit, filing grievances about work assignments, appealing county actions to the state, and maintaining eligibility for Medicaid, child care, and food stamps even if they are no longer eligible for cash welfare benefits.

Sanctioning

In the first round of the Rockefeller Institute field research, we found that personal responsibility agreements and program requirements were enforced by the more frequent, albeit selective, use of sanctions. In site visits and field researcher interviews, caseworkers and agency personnel maintained that sanctions were critical to the new message, and that the threat of full family sanctions in particular was important in getting clients to show up for orientation meetings and job interviews and to meet their minimum-hour participation requirements. At the same time, the evidence was unclear as to whether sanctions actually resulted in case closings in large numbers — that is, whether they were used essentially as signals of the seriousness of work and other requirements or more strictly to discipline cash assistance recipients.

States tended to impose less than full sanctions on families, perhaps because these sanctions are regarded as commensurate with the issues in question and thus more appropriate for bringing recipients into compliance. They were likely to impose graduated, calibrated, or even "vanishing" sanctions to focus parents' attention on program requirements without removing them from the program. Graduated sanctions are increased if violations are repeated or ignored. In the Rockefeller Institute sample, fourteen states were found to have adopted provisions for some form of graduated or part-time reduction of benefits. Minnesota, for example, reduces a participant's grant by 10 percent for at least one month as a first sanction; the second sanction results in taking away whatever remains of the cash grant after rent (and in some cases utilities). Less common were time-calibrated, participation-related sanctions. Wisconsin was the only state in 1998 using this approach. Under Wisconsin's W-2 program, cash payments are calculated in relation to the number of hours of program participation.

Political "Detoxification"

The work first approach, combined with time limits, affected not just what state and local welfare agencies said, but what they did: Agencies could do more because case managers had fewer TANF clients owing to the declines in caseloads. Local agencies could also do more as a result of changes in the political climate. We observed a new politics of welfare — indeed, what might be called a "detoxification" of welfare politics in the sense of its no longer being a big-time political issue. This is because of the greater public acceptance of welfare programs that help people go to work, stay at work, and thereby limit their time on cash assistance. The combination of changed signals, smaller caseloads, and new politics enables many social agencies to do more and to work together more easily on a range of service programs involving employment assistance, child care, transportation, health, food stamps, and other social services.

Field researchers were asked to describe the mix of services and benefits offered at the local sites they studied. We found an extraordinary range of often highly specialized services. The most common were child care, employment services, aid in child support enforcement, and transportation assistance; but we also found a number of sites offering such services as:

- educational services, the most common being remedial education, GED preparation, and English as a Second Language;

- substance abuse counseling, usually via contracts with community-based organizations;
- family support services, including a Teen Parent Center and Pediatric Care Center that a local state agency developed on its own (in Bibb County, Georgia);
- emergency housing, ranging from state-subsidized housing assistance to contracts with local churches to assist with recipients' housing needs;
- domestic violence and emergency intervention; and
- mental health services.

One-stop job centers in Wisconsin offer an especially wide range of services. The Wisconsin Works program (W-2) is fully integrated with other services at the most developed job centers, such as in Kenosha County. Participants in W-2 not only draw on employment and training services, employment support, and other services and benefits specific to the program, they also can use the job center to access a wide range of services, such as child care, child health checks, Children First (an employment and training program for noncustodial parents), a dislocated worker program, economic support programs (such as Medicaid and food stamps), Head Start, the Job Corps, Job Training Partnership Act (JTPA) programs, paternity establishment, Supplemental Security Income advocacy, and a feeding program for infants. In some of its local offices, Kenosha County's W-2 program provides these services by using "integrated service teams" of co-located staff members.

States are not merely adding services and forming new alliances; they are also making new financial commitments. Minnesota, for example, decided to fully fund a universal sliding fee for child care for the first time. The program had offered subsidies to welfare and low-income working families since 1986, but in the past waiting lists had been as long as three years. The coalition behind welfare reform, however, along with a large budget surplus in 1997-98, and the belief by the governor and the legislature that affordable child care was crucial for welfare recipients, led the state to spend an additional $60 million on child care.

While many innovations were found in the initial round of field research, the overall national picture for program finances highlighted not spending, but the lack of spending or the slow pace of the use of available funds. This led to the adoption of the fiscal-effects component of the Rockefeller Institute research on the welfare reforms.[10] States in 1999 had accumulated large balances of unspent and unobligated federal TANF funds. As a result, they were admonished by the chair of the House Subcommittee on Human Resources that "unless states begin spending more money, we will eventually lose the battle to protect it here in Washington." In the Institute project to find out how state social services spending priorities have changed in the wake of welfare reform, we focus on three key questions about state spending.

- *How has overall state spending on social services changed as a budget priority?* Is social services spending a greater proportion or lesser proportion of the state budget than before? What insights can we glean about the extent to which states have used a welfare surplus for other state priorities outside the realm of social services, such as tax cuts, education spending, or other purposes?*
- *How have priorities shifted within social services?* In what ways have states shifted their spending among program categories within social services? What are they spending more money on, and what are they spending less on? Why? Why are they finding it difficult to spend all of their TANF grants?
- *How have federal-state financial relationships changed?* How has the federal share of spending on various programs changed? Are states supplanting state money with federal money in some programs? How does this vary across states?

To answer these questions requires assembling a data set with two main features. First, we need data for the pre- and post-welfare reform period. Second, we need to put the entire social services pie on the table so that we can see how states have shifted spending among social

* Chapter 11 deals with an earlier field network evaluation, focused in this way on the issue of supplantation of federal aid funds for state expenditures and/or tax reductions.

service categories, how they have shifted reliance from state to federal funding or vice versa, and so that we can gain insights into how states used welfare savings to provide general budget relief. Many of these questions cannot be answered using federal data; they require field research.

Second Order Devolution

Although devolution to the states has been widely cited in describing welfare reform, the big story of welfare reform is not the states, it is local. Even in what are defined as state-administered welfare systems, we found major changes to devolve welfare and social program responsibilities to local entities.* This local devolution includes more than counties. In many states, new or relatively new regional entities now are responsible for welfare and related workforce services — second order devolution. One might be skeptical about this finding. By law, states over the years have assumed increased responsibility for welfare. Thirty-eight states now have what are legally "state-administered" welfare systems. The remaining states (including some very big ones — California, New York, and Ohio) have state supervised/county administered systems, under which states set policy goals, prescribe administrative arrangements, and provide funds, but the basic legal responsibility is at the county level.

The driving force behind second order devolution is intrinsic to the nature of the tasks involved. The 1996 shift to work first and a service strategy for welfare inexorably pushes decision making downward. This is because so much of what needs to be done to prevent welfare dependency and to keep people off of cash assistance has to be decided, arranged, and carried out locally. The biggest increase in discretion under the new regimes for welfare policy occurs at the point of contact between local case managers and the individual applicant or recipient.

In past periods in U.S. history, the idea that such discretion should be assigned to local workers was resisted. Reformers often complained

* In most states, welfare is both state supervised and also state administered. In other states, it is state supervised and county administered.

about harsh treatment by local agencies where poor people were treated badly, sometimes simply given "bus money" to leave town. If the kind of behavioral fine-tuning envisioned in the Personal Responsibility Act is to be achieved, however, there is no way to avoid assigning primary responsibility to local administrators and, most of all, to front-line case managers. As stressed earlier, this places a high premium on the development of information systems for case management to connect and track social program benefits and social services, and also for oversight and evaluation purposes.

Adaptability of the Research Process

In the way just cited and in others, field evaluation studies over the years have been adaptable in the sense of enabling findings to reflect institutional learning. As of the time of the writing of this book, many factors lead us to believe that the changes taking place in the late 1990s in welfare and job programs have staying power. The strong work orientation of the new welfare resonates with widely held beliefs by the majority of Americans about what welfare systems ought to do and emphasize. This may be the most lasting effect of the reforms, producing a new basis for political legitimacy that was lacking under AFDC. In effect, what we are seeing in many places is a shift in opinion to a new consensus about welfare systems with a more politically acceptable mission.

The second order devolution of welfare and job programs and the diversity we found in signaling and systems requires a similar devolution by researchers. For us, the critical independent variable institutionally is local "regimes," not programs in the customary sense. By regimes, we mean combinations of values, signals, agency cultures and connections, and operational capacity. Future Rockefeller Institute implementation research on welfare reform will concentrate on local systems: Are social programs connected for TANF benefits, job services, food stamps, Medicaid, child support, child care, transportation, and related social and health services (such as those for substance abuse treatment, the prevention of family violence, mental health, and family planning)? Local systems in this way are a major independent variable. Research ingenuity then needs to be applied to use Census data (2000 fortuitously is a decennial census year), administrative data, available state studies, and other information about what is happening to poor and working poor families as crucial dependent variables.

Almost all states are now conducting their own studies, using both administrative records and survey data, to ascertain what is happening to people who leave the TANF rolls. A number of states are also studying what is happening to other populations — people on the rolls as long termers, the population diverted from cash-assistance recipiency, and also what can be called (but is very hard to identify) the "deflected" population of eligible persons who are discouraged from applying for work-conditioned welfare benefits. Other research projects are also producing large data sets. Two of the major studies are sponsored by the U.S. Bureau of the Census and the Urban Institute, in the latter case focused on a representative sample of thirteen states.[11]

Still, in the final analysis, there is a limit to what social scientists can know about the effects of so complex and diverse a phenomenon as the plethora of current welfare reforms. Albert Einstein once said: "I have little patience with scientists who take a board of wood, look for the thinnest part, and drill a great number of holes where the drilling is easiest."[12] To evaluate program implementation in American federalism, the object should be to answer hard questions in as scientific a manner as possible; we shuld drill where the drilling is not easy. The purpose and argument of this book can be re-stated in these terms. Demonstration research can help politicians decide what to do for welfare reform. Then, after major decisions are made about new policies, two kinds of questions come into play: Were the new policies carried out? That is, did the program procedures and institutional systems change the way they were supposed to? And second, what were the effects of these changes on people? Like the sand of the sea, which the Prophet Hosea said, "can neither be measured or numbered,"[13] there are limits to what we can learn in both respects. But we can learn a lot and in so doing aid the policy process.

Endnotes

1 Richard P. Nathan and Thomas L. Gais. *Implementing the Personal Responsibility Act of 1996: A First Look* (Albany, NY: The Nelson A. Rockefeller Institute of Government, 1999).

2 John Burghardt, Anu Rangarajan, Anne Gordon, and Ellen Kisker, *Evaluation of the Minority Female Single Parent Demonstration. Volume I: Summary Report* (Princeton, NJ: Mathematica Policy Research, Inc., October 1992), p. xiii.

3 Deborah A. Ellwood and Donald J. Boyd, *Changes in State Spending on Social Services Since the implementation of Welfare Reform: A Preliminary Report* (Albany, NY: The Nelson A Rockefeller Institute of Government, February 2000).

4 Irene Lurie, "Evaluating Welfare Reform," Department of Health and Human Services Administration for Children and Families, Office of Child Support Enforcement, Arlington, Virginia, May 11-13, 1999.

5 This quote and other references to findings for states contained in the Rockefeller Institute's field research sample are taken from Richard P. Nathan and Thomas L. Gais, *Implementing the Personal Responsibility Act of 1996.*

6 There is an overall four-year time limit for adults that further restricts them to not more than two years of benefits in any five-year period, with a three-year exception for long-term participants with limited job skills.

7 Irene Lurie, "Evaluating Welfare Reform," op cit., p. 27.

8 Richard P. Nathan and Thomas L. Gais, *Implementing the Personal Responsibility Act of 1996*, p. 17.

9 U.S. General Accounting office. Welfare Reform: Improving State Automated Systems Requires Coordinated Federal Effort (Washington, D.C>: U.S. General Accounting Office, April 2000), Report to Congressional Committee, GAO/HEHS-00-48. See also Terrence Maxwell, *Information Federalism: A History of Welfare Information Systems, Working Paper* (Albany, NY: The Nelson A Rockefeller Institute of Government, 1999).

10 Deborah A. Ellwood, and Donald J. Boyd. *Changes in State Spending on Social Services.*

11 The Census Bureau study is longitudinal, called the Survey of Program Dynamics. Information about the Urban Institute thirteen-state study can be obtained from its website. For a discussion of the role of these studies, see Peter Rossi, "Research on PRWORA and TANF: What Can Be Learned from Four Projects Currently Underway," paper presented at an AEI conference on "Child Well-Being Under Welfare Reform,"Washington, DC, December 8-9, 1999.

12 As quoted in *Science News* 115 (March 31, 1979): 213.

13 *Old Testament,* Hosea 1:10.

11

Lessons from Evaluations of Employment and Training Programs

A major idea about public policy research advanced in this book is the need for multidisciplinary approaches. The field of employment and training research offers a good opportunity for examining this idea. Labor economists have played the strongest role in the design and conduct of evaluations of employment and training programs. In this chapter, I consider research both by economists and political scientists. I begin with the study conducted by the field network evaluation group at the Brookings Institution, which I directed, of the public service employment program established under the Comprehensive Employment and Training Act (CETA). This study produced a lively controversy about research methods.

The CETA Public Service Employment Program

The Comprehensive Employment and Training Act was enacted in 1973 in the form of a block grant as part of Nixon's "New Federalism" domestic program. Among other purposes, the aim of the law was to consolidate federal grants-in-aid in order to give greater discretion to state and local governments. Initially, most of the funds provided were for job training. Republicans resisted the idea, which many Democrats in Congress were advocating at the time, to provide funds for job creation.

Such a public service employment program was authorized as a title of the CETA law, but was funded at a very low level. Later, this public service employment program was expanded to combat the deep recession in 1974-75. Again, this was despite resistance from Republicans and from the Ford administration. As the recession deepened, however, spending for public service jobs was increased and Republican resistance subsided. This expansion of the public service jobs program occurred over three years from 1976 to 1978. The public employment program continued in effect until 1981, when it was eliminated under Ronald Reagan, who was vociferous in his criticism of this policy.[1]

At its peak in April of 1978, the public service employment program employed over 750,000 people at an annual spending rate of $7 billion. This made it by far the largest federal program in the employment and training field, and at the time one of the largest federal grant-in-aid programs to states and localities.* The expanded version of the public service employment program was an obvious target for Reagan's budget cutters in 1981; CETA job creation had few friends and lots of enemies when Reagan came into office and ended it.

In the usual way, the CETA job creation program of the late 1970s had multiple purposes which were not precisely stated or weighted. Among its major objectives were to provide jobs for the unemployed in recession periods; aid disadvantaged persons in developing skills and obtaining job experience; and assist state and local governments in providing needed public services. In 1976, when the public service employment program was reauthorized, Senator Henry L. Bellmon (R-OK) added an amendment to study an issue related to the first of these three objectives, job creation. The Bellmon amendment became the source of the interdisciplinary controversy featured in this chapter.

Senator Bellmon's concern was that public service jobs were being used by state and local governments to displace workers who would otherwise have been on state or city payrolls. Opponents of the public

* It was, however, smaller than the Works Progress Administration (WPA) of the Depression years, which at its peak employed three million people.

service jobs program charged that it was simply a hidden subsidy to state and local governments and that it did not really increase employment. Bellmon's amendment directed the National Commission for Employment Policy, a permanent research and advisory group chartered under federal law, to make a study of the "net employment effects" of the public service employment program. This commission, chaired by Eli Ginzberg, contracted with the Brookings Institution to conduct this research through a field evaluation study of the CETA public service jobs program. The study was initiated at Brookings, and was later completed at the Woodrow Wilson School of Public and International Affairs at Princeton University.[2]

The choice by the National Commission for Employment Policy to adopt the field network evaluation approach is interesting. Previously, there had been a number of theoretical studies by labor economists of the displacement issue for public service employment programs. The research was done in the form of statistical modeling studies, using econometric techniques based on program, national economic, fiscal, and demographic data. A 1974 paper written for the U.S. Department of Labor by labor economist George Johnson reached the tentative conclusion that the job-displacement impact of a federally funded public service employment program would be very large — in fact that in the long run the displacement effect of grants-in-aid for public service job creation would absorb all, or nearly all, of the funds appropriated for this purpose.[3]

Johnson's econometric approach is not the only way to try to answer the hard-to-answer question about the job-displacement effects of public service employment grants to state and local governments. By contracting with the Brookings-based field network to conduct the study mandated under Senator Bellmon's amendment, the National Commission was bringing an alternative methodology to bear. Researchers refer to this as *triangulation,* conducting studies using different methods to see if their findings converge.

It was in this setting that the Brookings-based field network turned its attention in 1976 to evaluating the effects of public service employment. Thirty field researchers (in this case approximately equally divided between economists and political scientists) conducted four rounds of field observations in forty state and local jurisdictions. The sample included three types of program sponsors: (1) large local units (over 100,000 population) that qualified as "local prime sponsors"; (2) consortia of local units designated as "local prime sponsors"; and (3) state governments responsible for administering CETA programs in "balance of state" areas. Within these jurisdictions, funds for job creation were channeled to thousands of public and nonprofit agencies, thus greatly expanding the screen for field analysis and requiring a sampling procedure for selecting the organizations for detailed analysis. The study jurisdictions at the time accounted for 10 percent of all enrollees in the CETA public service employment program.

The research design included a number of categories of effects. It was not limited to job displacement. It dealt with questions involving the fiscal effects of CETA job funds and also the types of jobs provided, the public services affected, structural matters involving the agencies that administered the program and their implementation processes, the training provided to participants, and the effect of these funds on nonprofit organizations, which sponsored many of the local programs.

As it turned out, the bottom-line finding of the Brookings-Princeton study of the job-displacement issue was very different from that made by George Johnson. His paper indicated a high (eventually total) job-displacement effect of the CETA public service employment program. By contrast, we found what many observers regarded as a surprisingly low level of job displacement.[4] For the sample as a whole, approximately one-fifth of the positions studied were assigned by the field researchers, in consultation with the study's central staff, to job displacement. These findings were issued in a series of reports published by the National Commission for Employment Policy, and summarized in books, testimony, papers, and articles. These publications described the research methodology and provided illustrations, drawing on statements in the reports submitted by the field researchers regarding the kinds of determinations that were made in assigning the positions studied to the various employment effect categories used in this field research.

A number of reasons appeared to explain why local officials (most of these funds went to local governments) were reluctant to use public service employment funds for displacement. One reason, and it is a reason which I believe other experts in the field did not take seriously enough, is that the law prohibited displacement. A number of provisions in the law were intended to prevent it from occurring. A second reason for the relatively low level of displacement we found was that local officials in many cases shared the national objective of creating jobs to relieve unemployment.

In even more basic terms, however, field researchers reported that many officials in the study sites were reluctant to use these job creation funds for displacement because of the possible long-term consequences of doing so. These local officials had been burned before by federal policy changes. In the case of the CETA public employment program, there was an understandably high level of uncertainty about the future intentions of the federal government. A major concern on the part of local officials was that when this federal largesse was ended, or if the rules for its use were changed (both events in fact occurred), they would be left holding the bag. They would face strong political pressures to lay off local workers or to increase taxes in order to continue to pay employees who were being supported with CETA funds if they had used these funds for job displacement. This was a risk that most local officials simply did not want to take.

There was one important exception to this finding. In jurisdictions facing serious fiscal pressures (mainly economically distressed central cities in our sample), we found displacement to be higher than in other jurisdictions, which we attributed to the need, or at least the perceived need, to use every available dollar to maintain basic services.

Another important finding was that many public service positions were not filled by local governments or other public entities at all; they were instead suballocated to nonprofit organizations. Over time, an increasing percentage of these positions were contracted out to nonprofit (mostly community-based) organizations that provided social services and administered community development programs. In the first round of the field research, one-fourth of the positions accounted for in the sample were subcontracted to nonprofit organizations; by the second round of the field research, this proportion had risen to nearly one-half.[5]

The reports we issued showing relatively low-level displacement were widely noted in the relevant policy communities because they were contrary to what many people (including some of the researchers in our field network) had expected. Two meetings about this controversy stand out in my memory. The first was with Senator Henry Bellmon. I was told by the staff of the National Commission for Employment Policy that he had read our initial report and that he wanted to talk to me about it. I was apprehensive, since Senator Bellmon had earlier indicated his opinion that CETA public service jobs funds were highly substitutive. At our meeting, and greatly to his credit in my opinion, Senator Bellmon said that, although he had been skeptical about our results when he first heard about them, after reading our report he was convinced by the conclusions and by our explanation about how we arrived at our findings.

The second memorable meeting at which the results of this research were discussed was arranged by the staff of the National Commission for Employment Policy to examine the two approaches used to analyze the net employment effects of the public service employment program. George Johnson and I were the speakers. The audience was relatively large for a meeting on research methods (about sixty people). It included both researchers and experts in the field of employment and training. True to academic style, this meeting featured a hot, intense debate about the competing research approaches. Those of us working on the field network evaluation study came away from this confrontation (not an inaccurate description) with a good feeling about the reaction to our presentation.

The "Complementarity" Approach

A major premise of the field evaluation approach as we have applied it over the years is that the choice among research methods is not an either/or proposition. In two of our studies — the public service jobs study and the earliest field evaluation we conducted, that of the revenue sharing program discussed in the next chapter — members of the field research group conducted what we refer to as *complementary* statistical studies. In the case of the public service employment program, the results of the complementary statistical study, using a pooled time series approach, were published in the *Journal of Human Resources*, thus subjected to the scrutiny of academic peer reviewers.[6] The idea of this complementarity approach was to use insights from the field research to build a better mousetrap for a statistical analysis. As it turned out, the high degree of subcontracting to nonprofit organizations under the CETA public service employment program made a substantial difference both in the field and in the statistical research we conducted.

The first step in conducting this complementary statistical analysis was to review the earlier econometric studies to identify lessons from the field research — in this case, especially the importance of subcontracting CETA public service jobs — that could cause distortion in a statistical analysis. In the early econometric studies, it was assumed that all of the CETA job creation funds allocated to a jurisdiction were paid as wages by that government. If, instead, some (maybe the bulk) of these funds were subcontracted to nonprofit organizations (or if they were suballocated to other public entities as was often the case, for example, to local school districts) this could have the effect of creating what appeared to be — but was not — a job-displacement effect in a statistical analysis.[7]

Another important aspect of the field research that was discovered while doing the statistical analysis involved the timing of the receipt and expenditure of CETA funds. Data from the U.S. Census of Governments did not fit well with the receipt-and-spending cycle for CETA job funds. It was found necessary to adjust for these timing differences in our statistical analysis. In the statistical study we conducted, this was done with the help of the Bureau of the Census, using unpublished worksheets from the Bureau that showed the actual receipt-and-spending cycle for CETA funds for the thirty municipal governments in our statistical analysis.[8] Fortunately for us (or I wouldn't be making so much of the point), the results of the complementary statistical analysis tied to the field research turned out to be similar to those found earlier in the field evaluation. The interdisciplinary approach used in this case involving economists and political scientists seemed to produce added value.

Studies of Individual Impacts under CETA

So far this chapter has described the fiscal effects of the CETA public service jobs program on jurisdictions.[9] However, these effects on jurisdictions were not the foremost concern of most researchers and policymakers. The dominant concern was the impact of CETA programs on *individuals*. In this research arena, more than any other, the experience with studies of the impact of large ongoing programs shows that measuring their impact on individual participants is fraught with problems.

Several characteristics of both training and public service employment programs produce high research hurdles. One hurdle is the federalism barrier reef discussed earlier. Employment and training programs of the federal government have been in an almost constant state of flux for a very long time. The CETA law was succeeded in 1982 by the Job Training Partnership Act, which in turn was succeeded by the Workforce Investment Act in 1998. However, the important point is that the CETA block grants and all successor federal programs have been administered on a highly decentralized basis by state and local governments. This involves fifty states, thousands of localities, and literally hundreds of thousands of local agencies and nonprofit organizations. The data problems of such a program would give any self-respecting researcher indigestion, but this is not all.

Another problem of individual-impact research in this field is that these programs are so widespread (indeed, effectively universal) that politics mitigate against the use of research designs based on random assignment whereby some persons are assigned to a program and others to an untreated control group. It would have been extremely hard for even the most supportive CETA managers of public service employment programs at the local level to arrange to have a randomly selected, untreated control group under these conditions. As we have already seen, random assignment is not the only way to get at questions about what works for individual participants in government programs, although it is the best way to do so. Without it, the results of studies are often unconvincing in part because of controversies about their methodology.

A third and related hurdle of evaluation research on the impact on individuals of federal grant-in-aid programs for job training is what might be called the "marginality problem." This was described earlier as the research hurdle of the null hypothesis. The typical CETA intervention (usually a job training program) was not that long or large. There is reason to question whether job training or job counseling services by themselves could have been expected to have a discernible, lasting impact on the lives of participants, that is, considering the great number of forces and factors (both public and private) that impinge on the way human beings develop and change in a technologically advanced society like ours.

In the 1970s, I served on the advisory committee for a large survey research project designed to evaluate the impact of CETA programs on individuals. It was called the Continuous Longitudinal Manpower Survey (CLMS), and was funded by the U.S. Department of Labor beginning in 1974. For the CLMS, the Labor Department contracted with the U.S. Bureau of the Census to have the Bureau survey a succession of cohort groups entering various components of CETA programs. Altogether, more than 6,000 people were included in this study. There was a baseline interview and in most cases four follow-up surveys for people in the sample.

But even after interviewing all these CETA participants, one faced the perennial question — *compared to what?* Observations about the experience of participants during and after CETA programs could not reveal the extent to which the programs did or did not assist them. There is value, of course, in knowing what CETA services were provided and what types of people received them. Still, it is reasonable to ask whether a research investment as large as this one should have been undertaken if these were the only research outputs. The advisory committee wrestled with the problems of establishing the counterfactual for this study, as did staffers of the Labor Department and the researchers with whom the Department contracted from the Westat Company of Rockville, Maryland.

The solution decided upon was to rely on data from existing files as the basis for comparison, mainly the Current Population Survey conducted by the Census Bureau, which in March of each year includes an enlarged sample on the employment experience of respondents. These data, however, are time limited. To obtain a longer time horizon, data from the Current Population Survey were linked with the earnings records of people in the CETA sample obtained from the U.S. Social Security Administration on a confidential basis. In this way, a "matched file" was constructed with characteristics similar to those of people in CETA programs.

Labor Department technical reports on this research were candid in discussing the drawbacks of this approach for establishing a comparison group. Contamination was one problem: There was no way of knowing whether the people in the matched file had received employment and training services. Furthermore, there was evidence that the Social Security earnings records (critical to the analysis, because earned income after the CETA program was the key outcome variable) did not include some types of earnings, for example, wage and salary payments to many state and local government employees and also federal government employees.

One of the early reports on the CLMS contains this depressing statement: "One simply cannot say at this time how comparable the comparison groups are to the participant groups in the absence of the CETA program."[10] The Labor Department continued to try to work out the wrinkles, and some progress was made. Nevertheless, a report issued two years later on this study contains a caveat similar to that just quoted: "As in all program evaluations where it is not feasible to randomly assign potential participants to the program or to a control group, there is some uncertainty about the amount of selectivity bias in the estimates of net impact."[11] Other experts in the field were not so restrained.

When the Reagan administration won enactment of the successor program to CETA (the Job Training Partnership Act, JTPA) in 1982, an elaborate data collection system like the CLMS was designed as the basis for research on the impact of the JTPA program. But in the end, it was decided not to use such data for research purposes. Instead, the U.S. Department of Labor decided to conduct studies with random assignment to ascertain the impact on individuals of JTPA services. The Manpower Demonstration Research Corporation was one of the organizations involved in this research. In sixteen areas around the country, MDRC along with Abt Associates studied 20,000 randomly assigned adults and out-of-school youth who applied for JTPA services and were assigned either to a treatment or control group.

Although called a national study, this is somewhat misleading as JTPA participants were not selected nationwide; they were chosen from the sixteen local sites willing to participate in this study, none of which were large cities. (The largest city included in the sample was Oakland, California, population 372,000.) Other barriers were that it was not possible for specific types of treatments to be tested rigorously, and school districts would not participate. Commenting on this effort to use randomization to evaluate an ongoing program, Cornell labor economist Vernon M. Briggs, Jr., said, "The problem is that this is a methodology only an economist could love." He added, "What it should be able to accomplish in principle is at variance with what it can accomplish in fact."[12]

The authors of the report on this study themselves came to the reluctant conclusion that the study design should not be replicated, noting that the first results from the national JTPA study were not available until six years after it was initiated, and also observing that it was "relatively burdensome to program staff in the study sites."[13] Other problems were political. To the extent measurements could be made, they showed very small impacts, and in some cases differences between the treatment and control groups that were so small as to be insignificant. Youth groups in particular showed no long-term gains. Other groups did show gains, notably adult men and women. However, the overall tone of the discussion of the findings from this study was downbeat from the point of view of advocates of this public program. Government officials tried to explain away disappointing findings or to downplay them. When the dust had settled, it has to be said that this was not a good moment for applied social science.

Endnotes

1 The Reagan administration later replaced the training portion of the CETA law with a new block grant, mainly distributed to the states, called the Job Training Partnership Act (JTPA), and then still later, in 1998, it was replaced with the Workforce Investment Act, which stresses the creation of "one-stop" job centers. This has been an unsettled, changing area of domestic policy for a long time.

2 I moved to the Woodrow Wilson School as a faculty member in 1979.

3 George E. Johnson, "Evaluation Questions for the Comprehensive Employment and Training Act of 1973," Framework for Evaluation Paper No. 2 (Washington, DC: U.S. Department of Labor, Office of Assistant Secretary for Policy, Evaluation, and Research, July 1974).

4 The design and findings of the Brookings study are summarized in Richard P. Nathan, Robert F. Cook, V. Lane Rawlins, et al., *Public Service Employment: A Field Evaluation* (Washington, DC: Brookings Institution, 1981). For a later and fuller account, see Robert F. Cook, Charles F. Adams, Jr., and V. Lane Rawlins, *Public Service Employment: The Experience of a Decade* (Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, 1985). Four rounds of field research were conducted for this study: in July 1977, at the beginning of the Carter administration's expansion of the program; in December 1977, at the height of this enrollment buildup; in December 1979; and in December 1980.

5 See Nathan, Cook, et al., *Public Service Employment*, chapter 2.

6 Charles F. Adams, Jr., Robert Cook, and Arthur J. Maurice, "A Pooled Time Series Analysis of the Job-Creation Impact of Public Service Employment Grants to Large Cities," *Journal of Human Resources* (1983): 283-94.

7 Later statistical studies of the intergovernmental effects of the CETA public service employment program corrected for this factor. See Lauri J. Bassi and Alan Fechter, "The Implications for Fiscal Substitution and Occupational Displacement under Expanded CETA Title VI," Technical Analysis Paper No. 65 (Washington, DC: U.S. Department of Labor, Office of the Assistant Secretary for Policy, Evaluation, and Research, March 1979).

8 This research had to be limited to relatively large municipal governments because Census Bureau worksheets were only available for these jurisdictions.

9 For examples of the implementation studies, see Grace A. Franklin and Randall B. Ripley, *CETA: Politics and Policy*, 1973-1982 (Knoxville, TN:

University of Tennessee Press, 1984); and William Mirengoff and Lester Rindler, *The Comprehensive Employment and Training Act: Impacts on People, Places, Programs, an Interim Report* (Washington, DC: National Academy of Sciences, 1976).

10 "Continuous Longitudinal Manpower Survey: The Impact of CETA on Participant Earnings," Working Paper No. 2, Entrants during the First Half of 1975 (Washington, DC: U.S. Department of Labor, Employment and Training Administration, June 1980, prepared by Westat, Inc., Rockville, MD), p. xviii.

11 "Continuous Longitudinal Manpower Survey," Net Impact Report No. 1, Supplement No. 1, "The Impact of CETA on 1978 Earnings: Participants in Selected Program Activities Who Entered CETA during Fiscal Year 1976" (Washington, DC: U.S. Department of Labor, Employment and Training Administration, July 1982, prepared by Westat, Inc., Rockville, MD), pp. 1-3.

12 Vernon M. Briggs Jr. review of "Does Training for the Disadvantage Work? Evidence from the National JTPA Study," by Larry Orr, Howard S. Bloom, et al., *Industrial and Labor Relations Review* 50, no. 3 (April 1997): 531.

13 Larry Orr, Howard S. Bloom, et al., "Does Training for the Disadvantaged Work? Evidence from the National JTPA Study." An Abt Associates Study (Washington, DC: The Urban Institute Press, 1996) p. 23. Commenting on the scope of the study, Vernon M. Briggs, Jr., said: "It is hard to understand how any evaluation of this type of public policy initiative can boast of being a 'national' study if it does not include any major city in its scope," in *Industrial and Labor Relations Review*, p. 530.

12

The Beginning of the Field Network Evaluation Methodology

This penultimate chapter, which continues the consideration of the field network methodology, is out of sequence chronologically in order to preserve the flow of the narrative on welfare reform issues and studies. The chapter deals with the first program evaluated using the field evaluation methodology, the general revenue sharing program enacted in 1972. The program is gone now. It is of interest for the book because of the way the early decisions were made about how to conduct this type of study.

Federal revenue sharing was enacted under President Nixon to provide general purpose grants-in-aid ($5.3 billion per annum initially) to state governments and some 39,000 city, county, and township governments, and Native-American Tribes. I had been involved in the design and enactment of revenue sharing as assistant director of the U.S. Office of Management and Budget during the first term of the Nixon administration. I had previously chaired the transition task force in 1968 that recommended a revenue sharing program to the new administration.

Revenue sharing was the keystone of Nixon's "New Federalism" domestic policy. The program was in existence fourteen years until it was ended in 1986 under the Reagan administration. Altogether, the program paid out nearly $80 billion over the fourteen years of its existence. Only Gerald Ford among Nixon's successors fully supported the program. Neither Carter nor Reagan (despite the fact both were former governors, or maybe because of this) shared Nixon's enthusiasm for the revenue sharing idea. Under Carter, state governments were eliminated as aid recipients (up to then they received one-third of the total amount of money distributed), and under Reagan the remaining shared revenue payments to local governments were ended.

When it was enacted in 1972, supporters of the revenue sharing program backed it for a number of reasons. Some saw it, as Nixon did, as an instrument of decentralization. Nixon frequently complained, using conventional Republican rhetoric, about the fragmentation and complexity of the federal grant-in-aid system. Its many hundreds of categorical grant-in-aid programs were seen as undermining the priority-setting processes of state and local governments and weakening their role vis-à-vis what was perceived as the increasingly intrusive role of the federal government in domestic affairs. Many Democrats, too, criticized the rigidities of categorical federal grant programs in this period. Walter W. Heller, a Democrat who supported revenue sharing as chairman of the Council of Economic Advisors during the early years of the Johnson administration, urged President Johnson to adopt this idea. Heller, always a great phrasemaker, referred to the federal grant-in-aid system as suffering from "hardening of the categories."

The answer for Nixon to the problem of multiple program categories was to change the form of federal grants, adopting two types of grant-broadening instruments — revenue sharing and block grants. Revenue sharing was essentially unrestricted aid to state and local governments. Block grants, which still exist, are flexible grants in a broad functional area, allocated on a formula basis. Nixon first proposed revenue sharing in 1969 at the relatively meager level, by Washington standards, of $500 million per year. This was not enough to get Congress's attention, so in January 1971, in a State of the Union message devoted almost exclusively to domestic affairs, Nixon upped the ante. He proposed an $11 billion combined program of revenue sharing and block grants — $5 billion per year for revenue sharing and $6 billion in block grants. (Total federal spending for grants-in-aid to states and localities at this time was $34 billion; total federal spending was $231 billion.)

Nixon's 1971 federal aid initiatives were estimated to increase total aid by 25 percent. "The time has come," he argued, "to reverse the flow of power and responsibilities from the States and communities to Washington, and start power and resources flowing from Washington to States and localities and, more important, to the people, all across America. . . . I reject the patronizing idea that government in Washington, D.C., is inevitably more wise, more honest and more efficient than government at the state and local level."[1] Most of the funds for Nixon's proposed block grants were obtained from consolidating existing categorical grants into broader programs, with some additional funds added called "sweeteners." Nixon's two most important block grants, both of which were enacted, were for employment and training and for community development.

Unlike the earlier and much smaller revenue sharing plan Nixon had put forward, his new and enriched $5 billion-per-year version of revenue sharing did get people's attention. It activated state and local officials to lobby for this initiative. In 1972, in the midst of Nixon's re-election campaign, revenue sharing was enacted. The President journeyed to Philadelphia to sign the law in the presence of a large group of state and local government officials, and claimed that revenue sharing would "renew" the American federal system.[2] The first payments were made just before Election Day.

Soon after this legislation was enacted, officers of the Ford Foundation approached the Brookings Institution to propose that Brookings conduct an evaluation of this program. It is interesting to look at the questions faced in taking up this research challenge, which of course you have to do when the Ford Foundation or another big foundation comes calling. The key and underlying question was: What difference did it make to have this new and more flexible fiscal flow from the national government to states and localities? Brookings formed a research team, which I headed, to consider this. From the outset, we decided that an evaluation of the revenue sharing program should focus on two main types of effects: (1) those that emerged as the most prominent effects in the implementation of revenue sharing; and (2) those effects (whether prominent or not) that were important to politicians because they were strongly reflected in the goals of the new program. The second type of effect is particularly hard to deal with. In textbooks on research methods, readers are told that step one in an evaluation study is to clearly define a program's objectives. Yet in the case of the revenue sharing program, as for a great many other public programs, there was not agreement among policymakers on the goals of the program.

In addition to Nixon's goal of decentralization, supporters of the revenue sharing program stressed its role in achieving other goals. Among those purposes were: (1) providing fiscal relief to local governments; (2) equalizing fiscal capacity among states and localities; (3) providing new funds to deal with important public sector needs; (4) serving as a stimulus to innovation on the part of recipient state and local governments; (5) stabilizing and reducing state and local taxes, particularly the property tax; and (6) altering the nation's overall tax system, putting more emphasis on income taxation (predominantly federal) as opposed to property and sales taxes.

Which of these goals should be featured in an evaluation? How should these goals (seven in all) be defined and weighed? Views of the various players in the policy process differed. In fact, the same players often emphasized different objectives at different times. Moreover, even if we could have assigned weights to all of the goals of revenue sharing on the basis of a close reading of the legislation and the legislative debates, this would not have been enough. Ideas about policy goals change all the time. The ultimate decision about the success or failure of a program depends on the way these goals are regarded by a particular person or group at a particular point in time.

Two other important factors complicated this picture. One was the fact that the program was *universal*. Every state and some 39,000 local governments received a share of the funds according to a distribution formula stipulated in the law. There was no way to use states and cities that did not receive revenue sharing funds as a comparison group for research purposes. Compounding this challenge was the federalism barrier reef. Tremendous variation exist in the structure of the state and local governments that received shared revenue, and also in the ways these jurisdictions defined and carried out their functions and kept their financial accounts.

The Research Approach

We decided to base our evaluation of revenue sharing on parallel and connected, uniformly structured case studies of a representative sample of state and local governments. The field research was conducted by an interdisciplinary group, ultimately thirty-two on-the-scene academic social scientists (political scientists and economists in roughly equal numbers) serving on a part-time contractual basis as field researchers.

Four broad categories were used to define the program effects to be studied — fiscal, programmatic, institutional, and distributional. The distributional effects of revenue sharing were studied centrally at the Brookings Institution in Washington, based on demographic and financial data from the U.S. Bureau of the Census and program data from the Department of the Treasury, which was responsible for administering the revenue sharing program. The analysis of the three other types of effects (fiscal, programmatic, and institutional) was done by the researchers in the field. Researchers in sixty-five state and local governmental jurisdictions (including one large Native-American Reservation) spent on average of thirty days on this study for each of three rounds of field observations. Before each round, the field evaluators participated in a research conference at Brookings at which the central staff and field researchers discussed the conceptual framework and the research plan for that round of field observation. The aim of these conferences was to have all of the members of the field research group on the same wavelength.

Field researchers submitted _their_ analysis for each round of observations. The reports consisted of answers to both closed-ended and open-ended questions according to a standard reporting format. Field researchers drew on state and local records and reports, and interviewed key participants in the decision making processes regarding the state and local use of revenue sharing funds.

It needs to be stressed that the analysis of the effects of revenue sharing was done by the field researchers. It was not based on the answers to closed-ended survey-type questions asked of particular state and local public officials, although other studies of the effects of the revenue sharing program used this survey-research approach. The Brookings central staff for this study reviewed and combined the analyses made by the field researchers into summary research reports. The draft reports were circulated for comments to the field researchers; it was very much a group effort. The sample for the study included eight state governments, twenty-nine municipalities, twenty-one counties, six townships, and one Indian Reservation. It overrepresented larger jurisdictions. Overall, the sample jurisdictions accounted for about 20 percent of the total funding of the revenue sharing program. The research was longitudinal. It began soon after the first revenue sharing checks were received by states and localities and continued for the same jurisdictions over six years.

The decision to start out quickly in the field research turned out to be a wise one. Although there was a temptation to devote more time to research design, we expected that the period in which the initial payments of revenue sharing funds were made would be the time when the decision making processes regarding the use of revenue sharing funds would be easiest to observe.

The Brookings study combined two methods, field research and a complementary statistical analysis.* Although the focus was on institutions (recipient state and local governments) as opposed to the effects of shared revenue on individual citizens, the research challenge was the same. We needed to model the counterfactual in order to determine what would have happened in the absence of revenue sharing so that we could draw conclusions about the effects of this new form of federal aid. This analysis, as noted, was done by the field researchers; the evaluation, like the program itself, was decentralized. In essence, field researchers modeled the counterfactual in that most powerful of all computers, the human brain. This modeling process used many

* This complementary field and statistical approach was also used in field evaluation of the public service employment program described in the previous chapter.

variables in elaborate ways. These were in effect similarly structured case studies integrated by the central staff into a combined analysis. This was not a new methodology for program evaluation research, but its extensive use in this and later studies did constitute a departure for evaluation research on the effects of U.S. domestic programs.

The central staff reviewed each field report. We did not change the assessments made by the field researchers without consulting them about the possible reasons for modifying their findings. In most cases, collegial relationships obtained. The role of the central staff was to elicit the reasoning behind the analytical findings made by the field researchers in order to make certain, to the best of our ability, that the group as a whole (all of the field researchers and the central staff) were using a similar analytical framework and approach across the sample. Other methods were used for checking the field analyses. One check involved changes in the assignments of the field researchers. In some instances, field researchers moved or were unable to continue to work on this study. Two field researchers died during the first round of this study.*

Our use of the field network evaluation approach was based on what I believe is the most sensible way to take into account the basic federalism terrain. One cannot unravel state and local decision making processes about the use of a fiscal subvention without collecting observations from the field. Field reports prepared in this way are much more detailed than the statistics that are available on state and local finances. The U.S. Census Bureau collects state and local financial, program, and employment data; however, they are not sufficiently detailed and comparable with respect to the way grant-in-aid funds are treated or with respect to the techniques and time periods used for accounting for them by different state and local jurisdictions. To study the effects at the margin of a program like revenue sharing, we needed a more sensitive

* Fortunately, this did not happen in any of the other field network evaluations. Transfers of responsibility for the state and local field research for other reasons (e.g., people moving or on sabbatical) did provide a check on the findings and the consistency of the application of the research methodology for some jurisdictions.

data-collection methodology that permitted researchers to probe about fiscal and program effects of this new subvention.

In addition to the field research, later when statistical data became available on state and local finances for the period during which we conducted the revenue sharing field research, a group of the researchers who participated in this study undertook a statistical analysis of the fiscal effects of the program on a complementary basis where the two research methods were closely connected.

Knowledge obtained in the field on the workings of the revenue sharing program enabled the researchers doing the statistical analysis to conduct a statistical study that took into account the nature and limitations of national data on state and local finances. The method used for this complementary statistical analysis was to apply regression equations comparing revenue and expenditure patterns before and after revenue sharing funds were received in order to make inferences about the uses of shared revenue, with observations from the field research guiding this analysis. As it turned out, the complementary statistical study of the fiscal effects of shared revenue was limited to major cities because of data availability considerations, in this case involving data from the U.S. Bureau of the Census. This part of our research, which was conducted as a special study for the U.S. Treasury Department and vetted with public finance experts, yielded findings in line with those from the field research.

A practical point about the advantage of using these two approaches together concerns the timing of program evaluations. Statistical studies of new national programs like revenue sharing, even under the best of circumstances, almost always involve a substantial time lag before the needed national data become available. In this case, this meant that the statistical research results were not completed in time for their consideration in relation to the Congressional schedule for legislative reauthorization.

On an overall basis, timing considerations had a significant bearing on the publication and use of our revenue sharing research. The initial findings from this field evaluation were published in books, articles, and papers and presented in Congressional testimony early in 1975.[3] The original revenue sharing law expired at the end of 1975, so the congressional renewal process was just getting under way as these findings were becoming available. In the House of Representatives, the Committee on Governmental Operations, which had jurisdiction over the revenue sharing program, began hearings in June 1974. In the Senate, where jurisdiction for this program was assigned to the Finance Committee, hearings were held in April of the following year. Brookings researchers testified at both hearings. In the House, we previewed our findings. By the time the Senate hearings began, the first Brookings book on this research was available.

A major point that emerged early in the discussion of the effects of revenue shared involved the *fungibility* of shared revenue: All dollars are green. Federal grant-in-aid dollars are very hard to trace. The law required that these funds be assigned to certain spending categories enumerated in the law (for example, public safety, environmental protection, transportation, health, social services). We found, however, that these categories had little effect on recipient behavior — sometimes no effect. These "official" designations of the uses of shared revenue, which were compiled and published by the U.S. Department of the Treasury in the early program years, were highly political, and in our view, and also that of the U.S. General Accounting Office, illusory. We found that because many local officials regarded police protection to be a popular area of governmental activity, they perceived political benefits from officially assigning revenue sharing dollars to public safety. However, in such cases our field researchers often determined that the effect of this infusion of federal aid funds was very different from what was officially reported.

For local officials especially (two-thirds of this money went to localities) grant-in-aid revenue is received from many federal and state sources, including in this period the revenue sharing program. In setting priorities, local officials could allocate broad-gauged revenue sharing dollars to any function they wanted for purposes of reporting to the federal Treasury, as long as they had spent at least that amount on that function. So, if they felt police protection, as stated above, to be a popular function that would appeal to constituents and to the Congress as a use of shared revenue, they could report that they used their shared revenue for the police. But did they actually use the funds this way? That is, was this extra money actually used for more policing? That was a question we asked our field researchers to answer, using a uniform analytical framework to report in the best way they could and with a detailed justification of their analysis. Over time, we developed increasingly better, and we felt more sensitive, ways for explaining fungibility in relation to the use of revenue sharing funds. In our opinion, the official data from the U.S. Treasury Department on the uses of shared revenue, based on the approved spending categories listed in the law, were misleading. A consensus emerged on this point, and as a result the official reporting categories were dropped in the first renewal of revenue sharing legislation.

Basically, shared revenue could be used for new purposes or to offset tax cuts, that is, substituting this money for what would otherwise have been raised through taxation. If the funds were used for substitution purposes (also called fiscal displacement), another question arose: In what types of jurisdictions was this most likely to occur? And why? Likewise, to the extent shared revenue was used for new-spending purposes, policymakers wanted to know what kinds of new purposes.

Although the aim of this chapter is to describe the rationale of the field network methodology, several points about the findings are useful here. Substitution effects were not found to be widespread in our research. One-time (often capital) expenditures dominated the new-spending effects. We determined that the reason for this was that government officials, especially local officials of smaller jurisdictions, were leery of absorbing this new aid money into their regular finances. They feared, and rightly so, that someday the federal government would turn off the water. They did not want to have to raise taxes or lay off regular employees when this happened, so they tried to keep shared revenue separate from other revenue. A good way to do this was to build something. Since many politicians have an "edifice complex," revenue sharing turned out to be good for them in these terms.

Despite the fact that our sample was stratified and not random, and despite the fact that we based our findings on field analyses that could not easily be replicated, we were able to present findings in ways that helped policymakers understand the effects of the revenue sharing program in the diverse and fragmented setting of American federalism. In every situation, we tried to be as clear as possible about our research method — its strengths and its limitations.

Other studies of the revenue sharing program were also conducted, several of them under the auspices of the National Science Foundation (which also provided support for the Brookings research) and the U.S. General Accounting Office.[4] The other approaches used in the evaluation studies of revenue sharing were: (1) surveys of state and local officials; and (2) statistical studies that compared the post-revenue sharing experience with a counterfactual established through econometric modeling.

Researchers using the three different methodologies, as is to be expected, had strong opinions about the weaknesses of approaches other than their own. I am no exception. I particularly have reservations about the use of self-reported survey data in this kind of setting. State and local officials have many options about what they could report as their use of this federal aid. Under such conditions, they are likely to tell researchers what they think the U.S. Congress or their constituents want to hear, whether or not that represents the way federal aid funds actually affected the finances of their government. The behavior, which is not unreasonable, bears out a statement by former New York Yankees catcher Yogi Berra that is a good axiom for studying governmental behavior in a setting as fluid and complex as American federalism. Said Berra: "You can observe a lot just by watching."[5]

Endnotes

1 The President's State of the Union Address Delivered before a Joint Session of the Congress, January 22, 1971. *Weekly Compilation of Presidential Documents* 7, no. 4, Monday, January 25, 1971 (Washington, DC: U.S. Government Printing Office), pp. 92, 94.

2 The President's Remarks Upon Signing the Bill Providing State and Local Fiscal Assistance in a Ceremony at Independence Hall in Philadelphia. October 20, 1972. *Weekly Compilation of Presidential Documents* 8, no. 43, Monday, October 23, 1972 (Washington, DC: U.S. Government Printing Office), p. 1536.

3 See Richard P. Nathan, Allen D. Manvel, et al., *Monitoring Revenue Sharing* (Washington, DC: Brookings Institution, 1975). See also Richard P. Nathan, Charles F. Adams, Jr., et al., *Revenue Sharing: The Second Round* (Washington, DC: Brookings Institution, 1977).

4 For a discussion of other studies, see Nathan and Manuel, chapter 12.

5 Yogi Berra et al., *The Yogi Book: "I Really Didn't Say Everything I Said"* (New York: Workman Publishing, 1998).

PART IV

CONCLUSIONS

13

Public Policy and Policy Research: Limits and Possibilities

Demonstration studies to test new policy ideas and evaluation studies to assess the effects of ongoing public programs can focus on the same units of analysis — individuals, groups of individuals, communities, and various different kinds of institutions and organizations. However,demonstration studies are best suited to the first two types of units of analysis — individuals or groups of individuals — although it is necessary (indeed essential) in demonstration research also to know a great deal about program operations, that is, the institutional behavior of a public program being tested. It does not make sense to test for the outcomes on individuals of a particular new public policy or major program departure if researchers do not know a great deal about how it operates.

Evaluation Research — The Frontier of Applied Social Science

My special interest in this book is in evaluation research on the implementation of ongoing programs. Once a new policy or program has been adopted, politicians are likely to be especially interested in its implementation. This is particularly the case in American federalism, as in any decentralized governmental setting. The experience of the field network evaluation studies of program implementation shows the importance of this point in the welfare field. We found that the Family Support Act of 1988 was not implemented aggressively or effectively, and on the other hand that the institutional effects of the Personal Responsibility Act of 1996 have been pervasive, deep, and fundamental.

This distinction between evaluation and demonstration research has important implications both for policy research and for public policy. Public policies can be of two main types: Governments can attempt to influence the *society* — that is, to influence how people in the society think, feel, and view the world. They can also attempt to cause public agencies and other organizations to take actions that change *individual behavior*. Many public policies have both types of purposes, and the lines between them can be blurred.

Welfare policies send signals. They seek to transmit values in the society and to change the way responsibilities to meet needs are viewed and treated. They also direct public agencies to impinge in particular ways on people's lives. Put in the most basic terms, public policies can touch the human spirit and they can touch a human life. They often seek to do both. They seek to achieve both societal effects and individual outcomes without being explicit about which types of goals are paramount or how they are to be weighted and related.

Many public policy researchers care very much (sometimes totally) about having crisp, clear measures of the effects of public policies on individuals. Policy researchers on the other hand tend to be less interested in the effects of public policies on societal values and institutional behavior. But the catch is that the way a public policy affects institutional behavior, though elusive, has to be understood for the purpose of measuring its effects on individuals treated as research subjects.

Realistic, well-designed evaluation studies can assess, although generally not definitively, the effects of complicated and often imprecise governmental actions where the aim is to assess their effects on societal values and institutional behavior. But they cannot easily assess — sometimes they cannot assess at all — whether such a public policy once adopted has caused a certain effect in the life of a particular individual or family. For example, teaching Johnny or Jane to read and act responsibly as a citizen by influencing school districts and schools is an important purpose of political actors for many education policies. But not every political action done in the name of such a purpose is translatable into actions by school districts and schools that can be defined precisely, treated as an independent variable, and measured in terms of its effects on an individual child.

Public policies operate in complex, noisy environments in which a great many factors are also operating. Tightly specified causal analyses of policy effects, no matter how desirable, often simply are not possible to undertake. What I am concerned about is that despite this situation, many policy researchers reject inferences about causality that are not based on rigorous, often only randomized, policy research. The perfect should not be the enemy of the good.

As we have seen in the discussion of demonstration research in Part II of this book, even under the best of conditions, the real world environment in which demonstration studies are carried out to test new public policies presents many hurdles to the specification of treatment and control groups in a social experiment. Moreover, because of the complexity of the environment in which they operate, many social experiments show only modest results compared to what the advocates of a tested new policy seek to achieve. In chapter 2, I referred to an article by two policy researchers, Gary Burtless and Robert H. Haveman, who reviewed three social experiments and reflected on the modest results obtained. They concluded dourly, "if you advocate a particular policy reform or innovation, do not press to have it tested."[1]

My concern is that the overemphasis of public policy researchers on individual, as opposed to institutional effects, loads the deck against government. It does so because it does not acknowledge that institutional and signaling effects can change values in pervasive ways that touch the human spirit greatly, but touch individual lives much less discretely and discernibly. This point about signals changing attitudes and thus affecting institutional and individual behavior indirectly is a strong theme in social science. Economists emphasize theories about signaling. Indeed, John Maynard Keynes stressed this psychological dimension for understanding how economies fare and function.

Welfare is by no means unique among areas of public policy in which value changes and changes in institutional behavior are intrinsic to what politicians are seeking to accomplish. In the post-World War II period, civil rights legislation has been the most notable area of American legislative reform in the field of domestic policy. Yet, civil rights laws are not instrumental in the micro-sense whereby politicians are trying to change individual behavior in precise and specific ways. Civil rights laws seek to change the way institutions behave — the labor market, governments, private accommodations, schools, and the electoral system.

While there is a large literature on civil rights policy, there was not in the period of the civil rights revolution in America a large amount of social science research on how and how much institutional change occurred in response to civil rights laws and policies. There are historical and advocacy literatures about civil rights in America, but perhaps because the research terrain is so amorphous, the sponsors and leaders of public policy research did not gravitate to this area the way they have to studies of more finely tuned social programs that appear to lend themselves to the application of statistical methods and random assignment to study individual outcomes.

In the case of civil rights, the impetus for change can be said to be liberal in the way that word is commonly used. Returning to the education policy used earlier for illustrative purposes, there has been a similar strong (although historically not as consequential) effort recently to reform schools and school systems. Here, too, the push has been to change values and institutional behavior, although in this case much of the recent hard charging has been done by conservatives. By advocating charter schools and vouchers, reformers seek to change education as an enterprise. Their aim, often explicit, has been to create competitive pressures by dint of their support for alternatives to public education so that the educational industry as a whole is stimulated to raise standards.

The best measure of the success of charter schools and vouchers is not whether a single student does better over a lifetime of earnings than a student educated in a conventional public school. This may be a test we would like to apply, but for the sophisticated school reformer it would not be the most sensible one, or even a sufficient one. The real test is whether the enterprise of education is infused with a new spirit of competitiveness to raise standards of excellence.

In short, the mindset of randomized experimental studies is useful and good, but its attractiveness to social scientists has caused many participants in governmental policy-making to view many policies too narrowly. Evaluation studies of the institutional effects of ongoing public programs may not be glitzy to social science researchers who like to use their latest statistical bells and whistles. Nevertheless, such studies are of great value to politicians who want to know if and how their new policies got implemented. Hard as it is to change reward systems for large and multifaceted undertakings such as applied social science in government, I think the public would be well served by trying to do this.

Many activities of governments in the United States could be made more effective if we knew more about what happened to public policies after they were adopted: Did they get implemented in the ways they were supposed to in terms of changed signals and changed organizational structures, institutional behavior, and administrative processes? My essential argument is that this type of social science knowledge-building (both quantitative and qualitative and involving multiple social science disciplines) should have more standing for public policy research and public policy researchers. Significant benefits could flow from this in many fields. The signaling, institutional, and administrative effects of efforts to reform public welfare, public schools, hospitals and health systems, and child care and child welfare systems, to name just a few examples, are all cases in which sophisticated, independent public policy research to evaluate the performance of institutions could enhance the nation's governance capacity and enrich our understanding of the operations of America's governments and American federalism.

The Demand for Policy Research

The focus of most of the discussion in this book has been on the supply side of public policy research, describing and comparing certain kinds of applications of social science to social policy. I shift in this final section to the demand side. Two main types of institutions — governments and foundations — are the main sponsors and funders of public policy research. Both have turf to protect and purposes to advance that are rooted in politics, although different kinds of politics. In this consideration of the demand for public policy research, the discussion again is organized according to three types of applied social science — demonstration studies to test potential new policies, evaluation studies to assess the effects of ongoing public policies, and studies of conditions and trends.

Demonstration Studies. I have already stated a preference for service-type demonstrations as opposed to demonstration studies of the effects of income maintenance programs, or similar universal entitlements. But even in the former case of service-type programs, I believe demonstration research should be selective. Demonstration studies are best suited to situations in which three conditions apply: Politicians and administrators are genuinely interested in a particular goal; there is a lack of consensus on how best to achieve it; and there is a willingness to devote the time and resources necessary for finding out about the effects of preferred policy approaches. Random assignment, which is the favored approach of most demonstration researchers, has the special advantage in such situations of helping (or at least making it easier) to factor out bias in the conduct of demonstration studies.

Generally, it is governments that sponsor and pay for demonstration research. For example, in the case of the supported work demonstration, the first demonstration study conducted by the Manpower Demonstration Research Corporation, it was the Ford Foundation that launched this ship, but it floated on federal money. The negative income tax and other demonstration studies of income maintenance programs conducted in the fields of health, housing, and education were paid for almost entirely by the federal government.

State governments and sometimes local governments and foundations also undertake projects to test new program ideas. Note the use of the word "projects," as opposed to studies, in the last sentence. By this, I mean to convey the idea that sometimes demonstrations are not so much a test of what works as an effort to prove that something does work. For social scientists, this can produce a dilemma: Should social scientists help to bring as much expertise as possible to bear in such projects, or should they restrict their activities to well designed, genuine, and independent pilot studies of programs and major new program approaches? My inclination is for social scientists to get involved, but not to conduct research or to sign and issue reports, unless there is a willingness on the part of the sponsors of demonstration projects to be scientific and nonpolitical about the results achieved and the methods used.

Increasingly, public policy experts trained in university graduate schools of public affairs are the people who play the key roles on both sides of this transaction. They serve as government and foundation officials responsible for sponsoring demonstration research and they also serve as public policy researchers. They both do applied social science and sponsor it, in which latter role they have a big hand in determining what kind of public policy research is conducted. People trained primarily in economics have played the main roles in applying social science to social policy, and tend to dominate public policy education. But, as I argue, while economists should be credited for this, there is a need for a broadened disciplinary perspective in order to add other social science disciplines and researchers — sociologists, political scientists, psychologists, and ethnographers — to the teams conducting demonstration research studies. This is done in some demonstration studies, but needs to be done more widely and more actively.

Two important topics also need to be considered here — access to data and the problem of squishy research. I learned from the experience of the Manpower Demonstration Research Corporation that months, even years, can be taken up working out study designs and getting them adopted, and even then that problems can occur in getting access to needed data. Sometimes a plan for a demonstration study aborts because the political stars are not aligned; government officials decide, for whatever reason, to prevent access to needed data. The best course of action in dealing with such situations is to be up-front and to avoid the political thicket when it is just too thick. It is in the interest both of the sponsors of applied social science and of policy researchers to avoid situations in which the results of demonstration research will not be credible. This is not to say that agencies and advocates should or could be prevented from using whatever data are available and making claims that a program works or doesn't work. However, government officials in the field of policy research and social scientists should try — and this is hard to do — to make a distinction between situations in which an independent and scientific test of a new policy or program idea is feasible and those in which politics operate in ways that make it unwise to invest in attempting to do this.

The word "research" cannot be patented. There are lots of reports called "research" written by individuals and groups that are not seeking to answer a question, but rather to advance a particular point of view. This is "advocacy research." It is a cousin to policy analysis, but I believe it is distinct from good policy analysis. Advocacy studies (both research and policy analysis projects) have a role to play. It concerns me, however, when advocacy studies are cited too easily and authoritatively as "research" in the media where quick answers and easy-to-describe arguments are always needed.

Evaluation Studies. Evaluation research, which is highlighted in this book, is harder to deal with than demonstration research when it comes to the problem of studies called "research" that aren't legitimately such. The tendency to claim success for an ongoing public program, or that the other side's program is a failure, is common in political discourse — very common. Evaluation studies that are legitimate research undertakings require sponsors and researchers who are credible and seen as such. Sophisticated as they may be, government officials who have a stake in their programs may not be able to resist asking researchers: "Really, tell me if my program is working." Even without intending to do so, they may be saying, "Tell me what is good about my program and how to make it better."

One solution is to have outside organizations (that is, outside of the agency involved or outside of government altogether) evaluate an ongoing program. This can include oversight agencies such as the U.S. General Accounting Office, the Congressional Budget Office, and the Congressional Research Service. Although it is often the case that within operating agencies, research offices lack adequate political insulation, officials in such positions (i.e., heading agency research offices) frequently feel compelled to show their independence by sponsoring evaluations that are unbiased. But this is not easy. It is very hard for organizations that do things also to assess them.

Even when outside organizations conduct evaluations, the politics of policy research can be hard going. To stay in business, a research organization (public or private) has to generate a steady flow of income. This requires a delicate balance in order to have a critical mass of support for the work one wants to do and at the same time maintain a high level of scientific integrity.

Outside evaluators are of three types — university research centers, independent think tanks, and private companies. The first two (universities and think tanks) receive grants from foundations, governments, and sometimes from corporations, whereas profit-making companies typically do not receive grants, although they do compete under "requests for proposals." As opposed to government research offices, foundations as the sponsoring agencies of public policy research often and understandably seek to use their (at least in their view) "limited" resources to change things rather than to study them. The special plea I would make to foundation grant officers is that applied social science of the type recommended in this book — *action research* — can help them. To change the world, it is often necessary as a first step to understand it.

The three types of research organizations are pulled in different directions. University-based social scientists interested in conducting applied research often have a hard time coexisting with their academic colleagues whose main interest is in theory building and who often eschew applied work. Think tanks, on the other hand, are not all of one genre. They are varied in their character. Some are advocacy organizations. Others are dedicated to independent inquiry. But even for the latter group of research-oriented think tanks, they often get (and sometimes want to be) co-opted — that is, positioned to reflect a certain point of view and set of values about the areas in which they conduct demonstration or evaluation studies. As for private corporations that conduct governmental research, their stockholders may not care about the purposes of the programs being evaluated, but they may find the often specialized and elaborate requirements of scientific inquiry incompatible with a corporation's other lines of business.

We need to look at the pluses and minuses of each of these three auspices. On the plus side, universities are likely to attract intellectually strong, creative social science researchers and to have inexpensive support personnel (students). Think tanks can be more specialized, hiring people with just the right subject-area background and research skills for a particular line of research, but they may not be as grounded intellectually in social science as university researchers. Private companies, on the other hand, can be more cost-consciously competitive and may have a somewhat easier time being apolitical, or at least not being as much affected by policy preferences as nonprofit policy researchers.

On the minus side, universities are difficult environments for managing undertakings like a large evaluation study. As already stated, the leading scholars in the pertinent academic departments often and understandably resist being pulled, or having their junior colleagues and students pulled, into applied tasks. The downside for think tanks, again as already suggested, is that they may have difficulty attracting and retaining highly qualified lead social science researchers for studies, not so much as consultants and advisors, but as the crucial experts that give cohesion and character to their enterprise. Another downside factor for think tanks is that in recent years policy experts at some think tanks with ostensible research missions have tended to yield to the temptation to engage too often in op-ed like punditry. Finally, the fact that private companies frequently are not comfortable homes for some types of social science researchers may say more about the kinds of evaluations private contractors are best suited to bid on than about their overall suitability as evaluators.

Studies of Conditions and Trends. The third type of applied social science, as defined in this book, is studies of conditions and trends. Even though this subject is not treated in depth, it needs to be included in discussing the demand for applied social science. Activities under this heading of studies of conditions and trends include "advocacy research" on public problems, which I argue is not an appropriate type of work for academic applied social science. This practice of essay-writing to advance a point of view is best left to advocacy organizations in the political process. There are times when social scientists testify, opine, and, as citizens, back ideas or causes they care about. But it is necessary, and I believe possible, to distinguish these occasions, as compared to those in which trained social scientists are acting in their professional capacity.

This discussion is not meant to gainsay the value of research and policy analysis done by social scientists to study conditions and trends, particularly those studies that involve creating and using large data sets. References were made in the discussion of policy research on welfare reform to state government studies of people who leave the welfare rolls or are diverted or deflected from applying for cash assistance. Federal agencies and national foundations have also undertaken large data-gathering studies of the conditions of poor and vulnerable populations in efforts to link such studies to changes taking place in welfare and related social programs. The largest such studies recently have been sponsored by the U.S. Census Bureau and the Anne E. Casey Foundation. These studies are examples of applied research on social conditions using large data sets, often multiple data sets (statistical data, administrative records, and survey data), in creative ways. What I am opposed to is not these types of studies, but rather the allocation of substantial amounts of money that could be used for applied social science to essay writing on what the social scientists themselves define as "problems."

Concluding Comments

The types of applied work that social scientists do in and around government can be distinguished by their degree of scientific seriousness. The most activist is *advocacy* research, which moves into, and I believe should be viewed as, a form of political action. It is not inappropriate for social scientists to function in this role. In truth, it would be unrealistic to consider preventing them from doing so. Although surely not everyone in the field would agree, I believe the amount of attention devoted to this advocacy role by social scientists should be reduced. And in the most scrupulous way possible, it should be kept separate from the professional practice of social science.

A second activist category of applied work that social scientists do is *participatory*, whereby social scientists are engaged in working with, advising, and assisting political actors in the design and execution of public policies. Again, there are social scientists who serve in this role and see great value in it. Such work can be technical assistance or can involve a stronger advocacy orientation. But again, such a participatory role should be distinguished from the professional practice of social science.

Henry James once advised an aspiring novelist that the theme of a book should come through like an iron rod. I hope the theme of this book that comes through in this way is that credible, applied social science is different from, and should be kept separate from, advocacy. The book argues for more and stronger emphasis on evaluation research — multidisciplinary studies of the effects of new and changed public policies on institutions. In the long run, such studies can enrich both theory building and scholarship in the social sciences.

I think of both demonstration research to test new program ideas and evaluation research to assess the effects of ongoing programs as "action research" to inform and assist the governmental process. Its role, as stated in the opening paragraph of the book, is to educate not advocate. Such a role for social scientists performing in their professional capacity needs to be distinguished from basic social science where the principal aim is theory building. However, in the final analysis the two activities are compatible, in fact very much so. They can reinforce each other. The relationship between applied and basic social science should be a two-way street. The conduct of applied social science research should not only be a matter of what social science can do for the real world. It also should be a matter of what the real world can do for social science.

Endnotes

1 Gary Burtless and Robert H. Haveman, "Policy Lessons from Three Labor Market Experiments," in *Employment and Training: R & D Lessons Learned and Future Directions.*

Index