

Presidential Address— Fostering Research Excellence and Impacting Policy and Practice: The Welfare Reform Story

Judith M. Gueron

Yesterday, this association's Policy Council adopted a mission statement that reads: "APPAM is dedicated to improving public policy and management by fostering excellence in research and analysis." These are inspiring words, but how do you foster excellence in research? How do you connect that research to the relevant decision makers so that it is likely to improve policy and management? And, what does this suggest APPAM's role should be, as distinct from that of APPAM members?

These are complex questions, but in my remarks today I want to provide an example of one route out of many, using a selective history of more than a quarter century of evaluations of welfare reform and employment initiatives. I choose this example not only because I know it firsthand, but because it produced a body of reliable information that is widely acknowledged to have played an important role in shaping policy and practice and, in the process, helped transform social science research more into a science.

So, why and how did this come about? As you listen, the story may sound so coherent that you may think this legacy was the gradual unfolding of some overarching master plan. But in the early 1970s, no one had laid out a centralized knowledge-building agenda.

The tale of how this happened could be told on many levels. I could focus on the special contextual factors that made welfare a fertile ground for doing research and having people pay attention to the results. Or I could emphasize, instead, the shifting locus of innovation and why the initiative moved from Washington to the states and back again. I could discuss the people involved and why they emphasized quality and rigor. Or I could focus on the continuing methodological advances, in the spirit of a history of technology. I could speak about the role of quantitative and qualitative research and the need for a synthesis of the two. Or I could point to the substantive findings and whether there was a logic to the progression of studies and lessons. Finally, I could describe the strategies used to affect policy and how and why they did or did not influence policy and practice. All of these themes deserve to be highlighted, but of necessity, I will be able to address only a few and hint at others.

My overall message is that there were elements—some obvious, others less so—in the welfare context that facilitated success. But there was also a fortuitous continuity and dedication among the key actors that help explain why they capitalized on this context to create an unrelenting forward momentum.

I once saw a rugby game that illustrates well the welfare research story: 15 people were moving forward down a field. At different times, different players carried the ball, but as they passed it, they kept moving forward. The players could represent

either the different welfare research innovators or the different dimensions on which research and policy advanced. The context—the rules of rugby and the competition—kept up the pressure to move forward and helped sustain the quality of the play, but individual players also shaped the outcome. In the real-world game of welfare research, this continuity and steady advance was crucial; it meant that people did not need to keep reinventing the wheel, since they were either part of the initial discovery or part of a group of people learning together. And, getting closer to home, APPAM played a role in creating and nurturing that community.

PRECONDITIONS AND CONTEXT

What were the special preconditions of the welfare context? It seems to me that seven features distinguish welfare from other domestic policy issues and help explain the continuing high level of federal and state focus on and use of exceptionally rigorous policy research.

First, until 1996, welfare was an open-ended, federal- and state-funded entitlement. Any low-income single mother could apply for assistance, though under terms that were set, to some extent, by the state. The resulting lack of any cap on expenditures, combined with the reality of increasing costs, meant that federal and state budget officials had to pay attention to welfare and were always interested in reforms that saved money. It also meant that people valued high-quality information on the effect of changes because they paid a clear cost for mistakes. Entitlement funding had another, more obscure, benefit. Because investing in research was a legitimate administrative cost, there was an uncapped source of matching federal funds to support state-sponsored research.

Second, welfare policy has always echoed basic values and attitudes about equity, social justice, and the role of women—all viewed through the prism of class and race. But in recent decades, public attitudes about these issues have changed enormously, reflecting among other things the dramatic increase in the labor force participation of women, including single mothers with very young children. This shift undermined the legitimacy of a program created in 1935 to pay some poor mothers to stay at home to raise their children. The combination of budget pressure and changing behavior and values made welfare policy a high-profile and intensely debated political issue. Governors—from Ronald Reagan, to Michael Dukakis, to Bill Clinton, to Tommy Thompson—won elections based, in part, on their reputations as welfare reformers.

Third, welfare supports some of the nation's poorest families. This means that people who care about issues of poverty and child well-being at all levels of government—but also in foundations, advocacy groups, think tanks, and research centers—have to be interested in ways to restructure the social welfare system to assist these families.

Fourth, welfare laws and practices have continually changed in response to some of these pressures over the past 30 years—from the Work Incentive (WIN) program of the 1970s, to the 1981 Omnibus Budget Reconciliation Act (OBRA) that increased state flexibility, to the Family Support Act of 1988, to the state waiver demonstrations of the 1990s, to the 1996 legislation that replaced the Aid to Families with Dependent Children (AFDC) entitlement with the Temporary Assistance for Needy Families (TANF) block grant. Two major trends are evident in the passage of these bills: the replacement of the unconditional entitlement to benefits with increasingly tough work mandates and other requirements, and ultimately by time limits on the receipt of federal benefits; and a shift in responsibility and initiative increasingly to the states.

Fifth, Congress, in 1962, added a fortuitous few words to the Social Security Act that had set the rules of the game for the federal-state welfare partnership since 1935. This obscure new provision of the law, Section 1115, gave federal officials the authority to allow states to waive provisions of the act to test welfare reform ideas likely to promote the law's objectives. Starting in the late 1980s, this authority was increasingly used to foster state innovation, but only if it would not increase federal costs and if the state demonstrations would be rigorously evaluated.

Sixth, these years spanned a special time in history. The U.S. Department of Health, Education and Welfare (today's Department of Health and Human Services, or HHS) had demonstrated in the pioneering Negative Income Tax (NIT) experiments that the medical research model of randomized clinical trials could be transported into social science research. At the same time, researchers were questioning the validity of a number of large-scale studies that used more traditional comparison groups. This prompted some academics and various advisory panels at the National Academy of Sciences and elsewhere to argue that if large-scale, longitudinal evaluations were to be launched, this new technique should be used.

A final factor was the visibility of the research—by Mary Jo Bane, David Ellwood, and others—that studied welfare dynamics and found that although some people stayed on the welfare rolls for long periods there was also substantial turnover (Bane and Ellwood, 1994). Apart from its substantive importance, this insight helped policymakers and administrators understand that using data on outcomes—for example, the rate at which people got jobs or left welfare, as had been done in WIN—to judge the performance of a welfare reform initiative would always overstate achievements, since to do so would wrongly assign credit to them for natural welfare dynamics. This strengthened the argument for seeking a reliable comparison group to assess the value added by any particular reform.

THE SHIFTING LOCUS OF INNOVATION

Major evaluations and demonstrations do not happen without very deliberate research entrepreneurship. Since the mid-1970s, the source of innovative approaches alternated among federal officials, the states, private research firms, and foundations. At the same time, the nature of how innovation was organized changed from relatively top-down, centralized structures to collaborative partnerships.

This shift reflected the devolving responsibility for policy innovation and the reality that studying welfare reform meant measuring the success of operating programs. Federal or foundation officials could fund research, but if they needed access to the program staff or to welfare recipients, they were rarely in a position to make a study happen. The people responsible for such programs had to agree to participate. That meant they had to be offered clear incentives to prompt them to cooperate with rigorous research protocols or intrusive observation.

Initially, special program funding was one lever used to gain cooperation. But as operational funding for demonstrations all but disappeared after 1980 and as evaluations increased in scale and political importance, the primary incentives became either the carrots of the gain in knowledge or the visibility that came from participating in a major study, or the stick of the Section 1115 requirement that a quality evaluation be undertaken in exchange for operational flexibility. Thus, the story of the devolution of initiative is also one about courtship, building partnerships, and leveraging federal power.

How did the shifts in initiative and the movement toward partnership come about, and how did this evolution in ownership affect the nature of the research and

the relationship between research and practice? I will share some insights into this process, using some of the key random assignment studies and tracing its occurrence over five distinct, though often overlapping, periods.

The first stage of the story starts with the 1974 National Supported Work Demonstration and was characterized by foundation innovation and centralized control. The entrepreneur was Mitchell Sviridoff in the Ford Foundation, but the federal role was also strong, because Sviridoff involved six federal agencies in a true funder partnership, including the U.S. Department of Labor (DOL) in the lead with HHS closely involved. The research team, from the Manpower Demonstration Research Corporation (MDRC), Mathematica Policy Research, and the Institute for Research on Poverty at the University of Wisconsin, included NIT veterans and newcomers (Hollister, Kemper, and Maynard, 1984).

The project was a resounding management and research success in the respect that it scored well in fostering research excellence. But it was a stunning flop in terms of its effect on policy and practice. The initiative had functioned as a stealth project, involving only the funders and the community-based Supported Work corporations. While this helped keep random assignment below the radar screen, the research activities were very top down. When the findings came out, there was no constituency in the then-existing employment and training or welfare systems—or in Congress or the states—waiting for the results and primed to act on them. That sobering lesson had a major effect on the structure of subsequent evaluations.

In the second stage, beginning in 1977, the federal government fostered partnership when Merwin Hans, the senior federal official in DOL responsible for moving people from welfare to work through the WIN program, launched the WIN Research Laboratory Project (Leiman, 1982). Hans thought that good research studies were often undermined by local program staff who did not see the value of the new approach being tested and thus had no stake in the study. To counter this, he proposed transforming a number of local WIN offices into laboratories that would function as programmatic and research partnerships among federal agency staff, local WIN practitioners, and evaluators. He hoped that this would lead to smarter program innovations, stronger research based on better questions and higher-quality implementation of research protocols, and a more enthusiastic reception and use of the findings by the staff in the four participating local offices and in the wider WIN system. Hans proved to be right.

In the early 1980s, the next stage of research entrepreneurship emerged in several locations, including at the Rockefeller Foundation, which launched the Minority Female Single Parent Demonstration, and at DOL, which sponsored the AFDC Homemaker–Home Health Aide Demonstrations and evaluations of the Comprehensive Employment and Training Act. But the project from this period that most shaped future welfare reform research was the Demonstration of State Work/Welfare Initiatives, which moved research from the relatively protected environment of the WIN laboratories out to the states and forged larger public-private partnerships.

This project was prompted by the 1980 presidential election, which brought Ronald Reagan into office with the controversial goal of changing welfare into mandatory workfare, requiring people to work in exchange for their benefits. Congress did not enact national workfare, but for the first time, it did give states the option to implement such programs, as well as new flexibility to reshape their WIN welfare-to-work programs. It became clear, however, that HHS did not intend to evaluate what states did under the new law. Indeed, the department abandoned an initial plan to do so and both HHS and DOL reduced their commitment to policy research in general, leaving the enterprise of research-based policy innovation in

jeopardy. Faced with the loss of federal resources and impressed with the opportunity to address important questions, staff at MDRC asked the Ford Foundation for a challenge grant to support the evaluation—to my knowledge the first major national evaluation to be launched with no direct federal funding.

The Demonstration of State Work/Welfare Initiatives raised by several notches the bar reached by prior studies. Abandoning the earlier mantra of “test small before spending on a large scale,” the vision for this project was to transform the impressive energy and innovation that swept states and localities in response to the new legislative flexibility they had been given into social experiments that would test the effectiveness of the new approaches. There was no special funding and no research laboratories; the studies were conducted in ordinary welfare offices in eight states. The scale was unprecedented and the resulting questions were different, but the successful implementation of the project provided strong confirmation of the value of Merwin Hans’s insight in the WIN labs (Gueron and Pauly, 1991).

If researchers want to implement a complex and demanding study in a real-world, operating program, people in the relevant agency must have a reason to cooperate. The argument that has repeatedly proven the most convincing is that the study will address questions that they care about and will provide results that are directly useful to them. Under those conditions, key state and local actors have a stake in the successful completion of the study, and once committed to the endeavor, they have shown themselves willing to work with researchers to overcome the inevitable obstacles that arise over the many years of a longitudinal evaluation.

However, for this collaboration to be genuine, the researchers must deliver something of clear value. It can be assistance in designing or operating a more effective program, answers to administrators’ questions about policy alternatives, a sense of participating in an elite learning community with a network of colleagues in other states, state and national visibility, or a combination of several of these. Moreover, these benefits have to be delivered in real time, not 10 years down the road. If researchers can meet this standard and get buy-in from the outset, the resulting involvement of the key stakeholders not only increases the likelihood of the study’s successful implementation, but it will provide partners for selling the results when they come in.

This model of a voluntary state-researcher partnership in learning set the style for the studies for the next 15 years. It showed that there was a small group of state and local administrators who were intensely interested in building knowledge about what worked and were willing to take political risks to participate in the research enterprise.

Many people have commented on the unusual impact of these 1980s studies on state policy and on the Family Support Act of 1988, attributing this to factors as diverse as random assignment, timely lessons viewed as impartial and aggressively disseminated, the replication of similar results across very diverse environments, and the nature of the findings themselves (Baum, 1991; Weaver, 2000; Wiseman et al., 1991. See Bane, 2001, for a discussion of the contrasting role of research in the 1996 welfare reform debate.)

The Demonstration of State Work/Welfare Initiatives helped to create a constituency for knowledge among a group of state and county administrators and legislative staff who found the findings useful in developing policy and sometimes even politically advantageous, since positive results often garnered extensive press coverage. Most gratifying, the experience led to some locations becoming repeat participants in random assignment studies; administrators in these locales were convinced that the mechanics of the lottery were a minor intrusion compared to the benefits. (In one case, the county commissioner actually continued random assignment after the research ended, seeing this as a fair way to contain staff caseloads.)

But I do not mean to imply that this courtship was easy or the partnerships always enduring. While it is nice to say that one learns equally from positive and negative findings, I do not think I would be talking about a 30-year track record if the research had produced only negative results. State administrators could often live with modest positive impacts, but they are by occupation very political consumers. Widely publicized bad news from some independent researchers about a highly visible reform program does not endear the research enterprise to a public official. This can prompt a “shoot the messenger” response and a “not in my backyard” attitude toward future studies. Related to this is another problem: The stronger the research design, the more likely you are to find modest impacts. Suffice it to say that there were many moments during this period when researchers were selling success based on modest net impacts, while governors were touting their achievements based on much more impressive gross outcomes. It took patience and perseverance to educate people in the press and practitioners on how to interpret these two sets of results. This remains an uphill battle.

During the next stage, starting in the mid-1980s, research entrepreneurship again advanced in parallel at the federal, state, foundation, and research firm levels. At the federal level, HHS (in what is now called the Administration for Children and Families [ACF] and in the Office of the Assistant Secretary for Planning and Evaluation [ASPE]) began evaluations of saturation programs in Philadelphia and San Diego, the evaluation of the welfare-to-work programs implemented under the Family Support Act (the National Evaluation of Welfare-to-Work Strategies study), the Teenage Parent Demonstration, the Post-Employment Services Demonstration, and other projects. Meanwhile, DOL launched a major evaluation of the Job Training Partnership Act and partnered with a number of foundations in evaluations of programs for youth and teen parents. Howard Rolston, at ACF, and Ray Uhalde, in DOL, were key entrepreneurs promoting these and other studies.

This new round of research built on the lessons from the prior ones, with each study’s designers, to pick up the rugby analogy, catching the ball and advancing down the field. Since there was little or no funding for sites and programs, and since their participation was voluntary, there was a deliberate effort to build partnerships, following some of the techniques developed in the earlier efforts.

At the same time, states were proposing increasingly ambitious welfare reforms. The initial focus was on building stronger welfare-to-work programs, with expanded services, tougher mandates, or both leading to evaluations in Massachusetts, California, Washington state, Florida, and elsewhere. But starting in the late 1980s and accelerating until the 1996 welfare reform law, the magnitude of change increased.

President Clinton had campaigned on the promise to end welfare as we know it, and governors responded by proposing increasingly dramatic changes, including various combinations of supplementing the income of working families, using “learnfare” to encourage teens to stay in school, increasing the harshness of sanctions on noncooperating families, setting time limits on welfare, freezing grants when women had additional children, and replacing welfare with a child assistance program. The resulting reforms were highly visible, sometimes forming the centerpiece of a governor’s political agenda.

For the most part, states simply wanted to change policy, not to study it. But in contrast to the early reforms in work programs, which could often be implemented under the normal flexibility permitted by the Social Security Act, the newer state proposals clearly went beyond the provisions in the law. States could move forward only if HHS agreed to waive some of the AFDC rules. The obscure Section 1115 provisions had

been invoked as far back as Supported Work (to allow for something called “grant diversion”). However, this was the moment during this 25-year period when the much-maligned group of bureaucrats at HHS (backed up by their colleagues in the Office of Management and Budget [OMB] who insisted on cost neutrality) resisted the intense political pressure simply to grant states greater flexibility and used the Section 1115 weapon to require rigorous evaluations, insisting wherever feasible on random assignment. By doing so, they converted this explosion of variation into an opportunity to learn. The result was a flowering of high-quality studies and a major expansion of the community of scholars involved.

Today, we find ourselves in a new stage, with HHS once again playing a major role in setting the agenda. In the 1996 legislation, two of the preconditions I mentioned at the beginning of my remarks were eliminated. Welfare was no longer an open-ended entitlement, and the long set of federal rules waivable only under Section 1115 were dropped. They were replaced by a system of block grants that give states enormous flexibility to design their own welfare policies. The result is greater variation, but the federal government no longer has the clout to assure that the differences among state programs will be assessed, and there is no longer the open-ended supply of federal matching money to support state-sponsored evaluations.

A few years ago, I wrote pessimistically about the diminished potential to launch new random assignment studies in this environment (Gueron, 2002). The period immediately following 1996 saw a profusion of valuable descriptive studies funded by foundations and government, along with efforts to assess the new law using complex designs, but opportunities for continuing the 25-year history of using random assignment to assess state reforms looked grim.

Astonishingly, the environment has recently become much more positive. While it is still too early to celebrate, it appears that HHS has figured out a way to continue the research tradition, albeit in a different form. The secret to this turnaround seems to be the creation of a new set of incentives that encourages states to return as voluntary partners to the research party. Rather than assess the states’ core, high-visibility TANF programs, the newest evaluations are studying the innovations, enhancements, and strategies states have designed to address emerging issues and problems raised by those programs. For example, recent state studies of people leaving welfare have shown that many got low-wage jobs and others returned to the rolls, raising the question of how to help people retain jobs and advance to better ones and move out of poverty. HHS staff in ACF and ASPE responded by targeting several key objectives—helping people in urban and rural areas retain and advance in jobs, reaching the hard-to-employ, building strong families, and providing child care subsidies—that represent areas where states want to try new approaches and where there is little experience or evidence. These projects are less like evaluations and more like demonstrations. HHS selects the general topic, an invitation is extended to states to participate, a courtship ensues, and a relationship is built. Because the projects offer real technical assistance and training and since the states and research contractors work together to build programs that seem strong enough to merit evaluation, states and counties see this as an opportunity to learn to use resources better. It is too soon to tell how these studies will evolve, what innovations they will produce, or whether the carrot of knowledge will prove as resilient as the stick of Section 1115.

ADVANCES IN METHODOLOGY AND IN UNDERSTANDING WHY PROGRAMS SUCCEED

A third way to tell the welfare story is through the continuing advances in research methods and the effort to understand why programs succeed or fail. On the

methodological front, the story about the use of random assignment in welfare studies is one of successive leaps forward and increasing ambition. Supported Work took the first step. The NIT experiments had already demonstrated the feasibility of randomly assigning eligible individuals to a test program or a non-treatment control group, but those were researcher-controlled tests of new welfare payment systems that did not involve integrating random assignment into operating service programs. The innovation in Supported Work was to take this concept and, for the first time, apply it to applicants for employment programs run by community organizations across the country. The use of random assignment was key, but more was needed to ensure strong findings. A sizable group of very disadvantaged people had to be enrolled in the study, and the survey team asking about their subsequent behavior had to succeed in collecting high-quality data—some on very sensitive and confidential topics—from a large percentage of them over a number of years.

All of this seems routine now, but at that time, lodging random assignment in real-world programs was a revolutionary idea. Most people doubted that program staff could be persuaded to use a lottery to determine who would be served, thinking instead that they would react like doctors being told to deny a patient a known benefit and refuse to participate in the study. Ultimately, by paying careful attention to ethical and legal issues, by balancing research requirements against operational feasibility, by using arguments about the limited number of program slots, by emphasizing the opportunity to gain funding and technical assistance, by highlighting the visibility that would come from participation in the study, and by honestly acknowledging the uncertainty of program success, the demonstration managers convinced 10 pioneering community organizations to join, and the study was successfully implemented. The result was the first convincing evidence that a transitional work program made a positive difference in the long-term employment of welfare recipients, but just as importantly, that it did not succeed for the other three target groups (Hollister et al., 1984).

If Supported Work showed it could be done, the WIN labs and the Demonstration of State Work/Welfare Initiatives advanced the ball by making several critical innovations. The first was to show that random assignment could be conducted in operating welfare offices without overly disrupting program intake or burdening staff. A second was to integrate sample enrollment into a mandatory program. A third was to use routine administrative records—in these instances, records created when employers reported to the Unemployment Insurance system and welfare agencies paid benefits—to track employment, earnings, and welfare behavior.

This change from the survey strategy of Supported Work (and, to my knowledge, other early longitudinal evaluations) to a reliance on existing records was a breakthrough essential for tracking the behavior of large samples within the relatively low budgets of subsequent very large-scale studies. And as the welfare reform studies unfolded over the next 15 years, large scale was critical because the new state programs being evaluated were often low-cost and in some cases not dramatically different from the old system open to controls. This meant that any resulting differences in behavior between people in the program and the control groups (the programs' impacts) might be modest but nonetheless important to detect, since modest impacts multiplied over tens of thousands of people and several years could have substantial human, budgetary, and political impact.

Later projects made further breakthroughs, including using additional administrative records, locating random assignment in different delivery systems (for example, the courts as part of the child support enforcement process), and randomly assigning groups of people (for example, whole public housing developments) in a community initiative where individual random assignment was not feasible.

By the mid-1980s, it was clear that random assignment could play a vital role in answering the first-order question of whether programs worked. But researchers were doing less well in understanding why some programs were more effective than others for particular groups of welfare recipients—insights that were vitally important to the state and local partners participating in the research. If the studies were to inform underlying theory and, through this, deliver on their potential to improve policy and practice, progress would have to be made toward determining the sources of program effectiveness.

The Supported Work Demonstration had included quantitative and qualitative efforts to get inside the “black box” and understand what features of program management, implementation, or the local context contributed to the success and replicability of the model. The qualitative research provided some insights. New ground was also broken on the quantitative front, thanks to a pioneering effort by researchers at Mathematica Policy Research to quantify the relationship between program inputs and outputs using data in some ways more detailed than that compiled in any subsequent welfare study. Nonetheless, this approach ultimately could not deliver on its goal of creating a useable production function for this complex social program. Twenty-five years later, this objective remains elusive.

There has been real progress, however. Subsequent studies have kept at these issues, adding ethnographic work, focus groups, and surveys with participants to try to understand behavior and develop theories of change. Researchers are also using more refined measures of participation to quantify the intensity of program treatments and making greater use of systematic qualitative and quantitative measures of program treatment and caseworker and staff management practices. Another major breakthrough was the use of increasingly complex, multigroup random assignment designs to measure the impact of various program components (for example, time limits or work incentives) or to compare directly the success of distinct philosophies (including different messages and services) operating back-to-back in the same sites. Prominent examples of this are the HHS-sponsored National Evaluation of Welfare-to-Work Strategies and the study of the Minnesota Family Investment Program (Hamilton, 2002; Knox, Miller, and Gennetian, 2000).

More recently, as studies have accumulated, there has been a growing effort to learn from the potential offered by the 25 years of virtually nonstop accumulation of high-quality evidence. Researchers can now move beyond each separate study to attempt different levels of synthesis—from assembling and comparing findings from many studies to undertaking formal meta-analysis, to pooling the primary data from several studies to address new questions with greater power (Bloom, Hill, and Riccio, 2001; Bloom and Michalopoulos, 2001; Greenberg, Michalopoulos, and Robins, 2001; Grogger, Karoly, and Klerman, 2002; Gueron and Pauly, 1991; Hamilton, 2002; Morris et al., 2001). Much remains to be done to create a theory and model of how such programs work, but the large number of studies that use comparable research designs and have collected identical data on aspects of program implementation and management provide what may be a unique opportunity to do this. In short, we now have the potential to deliver on the goal of quantifying the relationship between program inputs, outputs, and local context and build a real science. From where I sit, this appears to be an exciting new frontier, and I am encouraged that some recent studies suggest the potential payoff.

FOSTERING EXCELLENCE AND IMPACTING POLICY AND PRACTICE

As I mentioned earlier, this history of welfare evaluations looks, in hindsight, like the smooth unfolding of a unified knowledge-building agenda. It was, in fact, the

result of a series of independent actions. Then why was the progression so linear, with each initiative building on lessons from the preceding ones? What drove the forward progress?

The contextual factors outlined earlier clearly played a role. The fact that reforming welfare had a continuing high profile throughout the period drew funds and made the search for solutions a priority. The reality of ever-changing federal legislation prompted new studies and put states increasingly in the driver's seat, leading to a stream of new research questions and building a potential audience for reliable information on what worked and what was cost effective. The fact that Section 1115 required states to get a federal signoff to launch major reforms provided an opportunity to insist on high-quality research.

But to capitalize on those factors, there had to be determined entrepreneurs, willing partners in the states, and sophisticated funders, and there had to be a fellowship that united the players to ensure that they would not continually reinvent wheels. None of this was automatic; some of it was probably just chance; some of it was the phenomenon of successful studies generating momentum for more success.

As I have outlined, the entrepreneurs were at different times in different organizations—ACF and ASPE in HHS, foundations, DOL, the states, and a range of research institutions—and they were often motivated by the support of staff in Congress, in OMB, and people in other federal and state agencies. I know only one of these institutions well, but I think that the unrelenting three-decade-long commitment by MDRC to identify and answer the next generation of questions, employ random assignment whenever possible, and use its findings to inform policy and practice was shared by the broader community of researchers, federal officials, and a select group of foundation officers who valued the legacy of policy change that could be leveraged by strong research. These people were united in their vision that evidence-based social policy was a key to progress and that, if the methodology was unimpeachable, policy research had a better chance of being seen as believable and not as just noise from yet another pressure group.

While the list of influential actors is long, it seems to me that it was officials at HHS who drove the process at crucial times over these almost 30 years, sometimes as the initiator and sometimes as the reactor, and particularly during the 10-year period when HHS staff responded to the state waiver requests. Although they could easily have taken a more permissive stance, they instead used the leverage of Section 1115 to insist on high-quality research. I do not really know what motivated this, but from where I sat, it looked like it was inspired by a search for truth and a conviction that the objectivity of random assignment was their best protection against research ending up as a mere tool of politics.

A year or two ago, I was concerned that the demise of Section 1115 and the 1996 block granting of welfare would spell the end of this renaissance. But impressively, without the clout but with the allure of a shared fellowship of learning about the key concerns, HHS is successfully launching a major new group of ambitious studies. This bodes well for continuing the legacy.

The second key group was the state administrators. As I said earlier, researchers and funders alone cannot create this type of real-world policy research or change practice on the ground. It is the state and local administrators and practitioners who provide the playing field and who decide whether to act on the lessons the research produces. The relatively small number of pioneering welfare commissioners and staff in state and local agencies who opened up their programs to objective scrutiny, cooperated with the research protocols, provided the data, and sometimes ultimately became fans of high-quality evaluations were indispensable

to the fostering of research excellence. To me, their involvement has always been the most moving. While there were clear rewards to being in these studies, the burdens and risks were just as clear. Running welfare programs is difficult; line staff are very hard pressed. Administrators could easily ask, Why me? Why not just control the numbers and the public relations for my state and take a free ride on evaluations conducted elsewhere?

While the studies were sometimes forced by HHS, the state-researcher partnerships almost always became genuine over time. I don't want to exaggerate this, but the result, after all these years, is that, in a number of states and localities, there is a real constituency for this work. I used to say that no one wants to be in a random assignment study, but I now believe that is not true. I have to assume that this is prompted by administrators who think that this experience helps their agencies improve policy and practice.

APPAM, too, had a part in this story, and I want to turn to that now. A relatively small number of people have been key actors over the full span of the welfare research story, others have joined over the years, and many of us meet yearly at APPAM conferences. These are not our only occasions to get together, but they have been among the rare opportunities we have had to listen to each other's methodological breakthroughs and track the emerging findings. I am impressed that so many of us do that. If you were to ask this group of scholars, administrators, and entrepreneurs about their conference of choice and their research community, many would answer APPAM. The Journal has also played a part, publishing individual studies and, occasionally, symposia on the research itself or on its impact on policy.

APPAM conferences have served another role through their overt attempts to force an exchange between practitioners and policy analysts. Many of us recall vividly sessions where state officials told researchers what they valued and what seemed arcane or too late to be useful. This feedback has improved our work. Hopefully, the practitioners have also benefited from the exchanges. Together, I think this has helped foster excellence, improved policy and management, and moved the whole line of scholars and practitioners forward.

In conclusion, this history is remarkable and, in the social sciences, may be unique. Recent changes are heartening, but another quarter century is certainly not assured. It will take the dedication of the next generation of entrepreneurs to keep moving the ball forward. Those of us who believe that research excellence is the key to improving social policy and practice—and ultimately the well-being of low-income people—must continue to demonstrate and advocate for its benefits.

JUDITH M. GUERON was President of APPAM in the year 2002, and is President of Manpower Demonstration Research Corporation (MDRC), New York.

REFERENCES

- Bane, M.J. (2001). Presidential Address—Expertise, advocacy and deliberation: Lessons from welfare reform. *Journal of Policy Analysis and Management*, 20, 191–197.
- Bane, M.J., & Ellwood, D.T. (1994). *Welfare realities: From rhetoric to reform*. Cambridge, MA: Harvard University Press.
- Baum, E.B. (1991). When the witch doctors agree: The Family Support Act and social science research. *Journal of Policy Analysis and Management*, 10, 603–615.
- Bloom, D., & Michalopoulos, C. (2001). *How welfare and work policies affect employment and income: A synthesis of research*. New York: Manpower Demonstration Research Corporation.

- Bloom, H.S., Hill, C.J., & Riccio, J. (2001). Modeling the performance of welfare-to-work programs: The effects of program management and services, economic environment, and client characteristics. New York: Manpower Demonstration Research Corporation.
- Greenberg, D.H., Michalopoulos, C., & Robins, P.K. (2001). A meta-analysis of government-sponsored training programs. Baltimore: University of Maryland Baltimore County.
- Grogger, J., Karoly, L.A., & Klerman, J.A. (2002). Consequences of welfare reform: A research synthesis. Santa Monica, CA: Rand Corporation.
- Gueron, J.M. (2002). The politics of random assignment. In Mosteller & Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 15–49). Washington, DC: Brookings Institution Press.
- Gueron, J.M., & Pauly, E. (1991). *From welfare to work*. New York: Russell Sage Foundation.
- Hamilton, G. (2002). *Moving people from welfare to work: Lessons from the National Evaluation of Welfare-to-Work Strategies*. Washington, DC: U.S. Department of Health and Human Services.
- Hollister, R.G. Jr., Kemper, P., & Maynard, R. (Eds.) (1984). *The National Supported Work demonstration*. Madison: University of Wisconsin Press.
- Knox, V., Miller, C., & Gennetian, L. (2000). *Reforming welfare and rewarding work: Final report on the Minnesota Family Investment Program*. New York: Manpower Demonstration Research Corporation.
- Leiman, J.M. (1982). *The WIN labs: A federal/local partnership in social research*. New York: Manpower Demonstration Research Corporation.
- Morris, P., Huston, A.C., Duncan, G., Crosby, D.A., & Bos, J.M. (2001). *How welfare and work policies affect children: A synthesis of research*. New York: Manpower Demonstration Research Corporation.
- Weaver, R.K. (2000). *Ending welfare as we know it*. Washington, DC: Brookings Institution Press.
- Wiseman, M., Szanton, P.L., Baum, E.B., Haskins, R., Greenberg, D.H., & Mandell, M.B. (1991). Research and policy: A symposium on the Family Support Act of 1988. *Journal of Policy Analysis and Management*, 10, 588–666.