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
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


