Book Review: Fighting for Reliable Evidence, by Judith Gueron and Howard Rolston.

REBECCA A MAYNARD, University of Pennsylvania
Rebecca A. Maynard


_Fighting for Reliable Evidence_ is an incredibly powerful book, full of information that graduate schools do not teach and that would take a lifetime to learn on the job. Readers shadow the narrators—Judith Gueron, one of the founding employees and now President Emeritus of MDRC, and Howard Rolston, a long-time senior civil servant in the U.S. Department of Health and Human Services (HHS)—as they share their intersecting professional journeys in service of improving the welfare of our nation and its people. Over a 40-year period that began in 1974, a small band of warriors from the philanthropic, policy, and research communities, including Gueron and Rolston, fought and won many battles in a war to improve the availability and use of reliable evidence about what does and does not work to address the needs of poor and vulnerable populations. By shadowing the authors, readers learn the distinction between reliable and unreliable evidence and between evidence that is and is not useful for improving the development, implementation, and management of public policies. Readers also will gain an appreciation of the challenges to and rewards from strategic investments in generating such evidence.

The vision, commitment, and grit of a relatively few individuals, including the authors, were critical in moving us from a world in which empirical evidence was largely absent from the policy process to one in which the federal government is now experimenting with performance partnerships and increasingly often considering the extent and reliability of evidence in its policy deliberations (Haskins & Baron, 2014; Mervis, 2013). However, as the title of the book suggests, the war has not been won. There is still a dearth of reliable evidence to guide policy decisions. Importantly, this book demonstrates well the feasibility of generating reliable evidence across a wide range of settings by using social experiments and the power of successful experiments for accelerating the accumulation of evidence.

The story is rooted in the history of the use of social experiments to identify fixes for what, over the second half of the 20th century, was widely regarded as a seriously flawed welfare system. The system cost too much and did too little to incentivize and support poor or otherwise struggling individuals and families to improve their social and economic circumstances. As Gueron and Rolston explain, the beauty of experiments is that they largely sidestep big arguments about the appropriateness of the methodology used in and the credibility of findings from the “plenty of demonstrations, evaluations, and studies of welfare and employment and training programs...” (p. 11).
Experiments offer a very simple, easily understood way to know whether the policy or program in question made a difference that matters; you create comparable groups through the flip of a coin, expose one group to your program or policy innovation, and compare the outcomes for the two groups. The common arguments against social experiments include their high cost, long turnaround time, and knee-jerk reactions regarding whether or not they are ethical. Throughout this book, the authors describe examples that challenge these arguments and describe circumstances where the imperative of knowing the impact of a policy or program trumps temptation to eschew the experiment in favor of no or unreliable sources of evidence.

The tale begins before there was an MDRC or any such organization dedicated to rigorous evaluation of social policies and programs. The setting is the lead up to a decision by the Ford Foundation and six federal agencies in the early 1970s that it was time to “try small before you spend big” by “get[ting] the strongest possible evidence [using random assignment] about whether the program affected key outcomes and how benefits compared with costs” (p. 12). What the Ford Foundation and its federal agency partners decided to test was Supported Work—a program designed to move individuals with low attachment to the work force like former drug addicts and individuals with criminal backgrounds into the workforce through a process that gradually increased the expected performance on the job and reduced employment supports (Hollister, Kemper, & Maynard 1984). Both the Supported Work concept and the commitment to run the programs as social experiments were novel and sufficiently challenging that the Ford Foundation created a new organization, MDRC, to manage the trial. Together, MDRC, its funders, and the community-based organizations that operated the Supported Work Programs blazed a trail that, over the next 40 years, gave birth to a new arm of social science research—large-scale demonstrations and evaluations that rely on experimental designs to measure impacts.

One thread in the book concerns the organizational leadership principles and actions that were critical to the success of the Supported Work Demonstration evaluation, but that also contributed to the strength, endurance, and influence of MDRC on the standards and practice of evaluation research. A second thread tracks the inter-dependencies among the policy, research, and community contexts for social experiments. A third thread tracks the evolution of the science and practice of evaluation research at MDRC, beginning with the Supported Work Demonstration and continuing through the early part of this century.

ORGANIZATIONAL LEADERSHIP PRINCIPLES AND ACTIONS

In 1974, when MDRC was created to manage and evaluate the Supported Work Demonstration, we understood well the theoretical rationale for using randomization of individuals to the treatment or to a control condition as the basis for drawing reliable causal inferences (Campbell & Stanley, 1966; Campbell, Stanley, & Gage, 1966; Riecken, 1962) and there was budding interest in using experiments in the social policy arena (Rivlin, 1971). However, we were still in the midst of the first major application of social experimentation with the New Jersey negative income tax experiment (Conlisk & Watts, 1979; Kershaw, Fair, & Watts, 1977; Watts & Rees, 1977). Moreover, the mission of scholars who designed, ran, and evaluated the negative income tax experiments was quite different from that of the newly created MDRC. The former was a collection of academics who joined forces to advance economic theory and its application, while the latter was an organization dedicated to building knowledge to inform decisions about whether specific programs or policies
were worthy of public support—that is, whether they improved the lives of the focal
groups and at costs that are justifiable.

This story shows how the laser focus of MDRC’s leadership on the mission of
learning truths about the impacts of new programs, policies, and practices and
on communicating these truths to people who could use them to achieve better
outcomes served both the organization and the public well. For the organization
it meant channeling its human capital toward a central goal of building trusted
knowledge about what did and did not work to mitigate social welfare concerns
and resisting temptation to follow the money or political winds. The result was the
emergence of an organization with brand recognition for designing and carrying
out rigorous social experiments to test important ideas and for communicating the
findings well to diverse audiences.

INTERDEPENDENCIES AMONG THE POLICY, RESEARCH, AND COMMUNITY CONTEXTS

A second strand of this story is a version of “when the time is right.” In chap-
ter 2, Gueron explains how the imperative for testing the Supported Work con-
cept derived from growing divergence between the vision for and the reality of the
well-intentioned New Deal policies. The result was eroding support for the New
Deal programs and widely different views about the best policy response. The Sup-
ported Work Demonstration was conceived as an effort to “find out whether a
promising program could be replicated . . . to show the value of . . . ‘try-small-before-
you-buy’ . . . and [to] test a new technique for management and oversight of large
scale . . . research efforts” (p. 23). Variants of these themes are present throughout
this journey—meeting the needs of real people, finding effective and cost-effective
strategies for supporting public welfare objectives, and maintaining an effective
research enterprise that can conduct the needed research and communicate the
findings well.

This journey likely would have had a very different ending had MDRC (with
support from the Ford Foundation) not recognized and crafted a response to the
struggles of frontline program administrators who, in the early 1980s, were launch-
ing controversial welfare reforms in response to Congressional actions that created
new options for state welfare agencies to operate work programs, absent evidence to
guide decisions about whether and, if so, how to exercise these options. This was the
occasion for MDRC to message what it had demonstrated through evaluations con-
ducted during its early years: “You do not know whether . . . programs will work; typ-
ical outcomes will not answer that question; and you need to find out the . . . impacts”
(p. 119). This timely messaging, combined with federal requirements for some form
of evaluation of emerging welfare reforms and generous matching funds from the
Ford Foundation opened the door for a “new [evaluation] paradigm”—experimental
design evaluations that were integrated into normal program operations. This was
the breakthrough needed to establish that random assignment could be ethically
implemented in the context of operating programs and, in Gueron’s words, to learn
that random assignment was “the dog that did not bark” (p. 128).

However, because these work/welfare demonstrations were high-profile public
programs, it was notable that the impacts were small relative to the outcomes that
politicians were citing as evidence of their successes. In chapters 5 and 7, Rolston
describes how both Congress and HHS became critical forces in moving beyond the
knee-jerk focus on outcomes over impacts as the success indicator. Key actors in
both arenas now understood the feasibility and value of experimental evaluations
of controversial and high stakes policy issues, like Learnfare and family caps on
welfare benefits. Importantly, experimentation came to be viewed as a responsible
way for the federal government to incentivize states to come up with cost-effective strategies for administering welfare.

By the mid-1990s, a number of governors also were on board with, if not advocates of, random assignment, in large part due to the persistence of the frontline warriors at MDRC and their growing band of comrades at HHS. As Rolston explains, “we learned how to present random assignment as a fair test, . . . we became adept at . . . convince[ing] staff that . . . [random assignment] would avoid unknown and incalculable uncertainty” (pp. 256–257).

**THE EVOLUTION OF THE SCIENCE AND PRACTICE OF EVALUATION AT MDRC AND BEYOND**

I had the privilege watching the evolution of MDRC, sometimes from afar and other times closer up. Yet, there was much about the inner workings of the organization that I had not seen. This book exposed the debt those of us in the policy and evaluation communities owe to that band of warriors within the organization who stayed focused on their mission of generating and using evidence to improve the lives of people. This meant continually improving the science and craft of field experiments as new opportunities and challenges emerged. There is now widespread acceptance within government, among a growing range of practitioner groups, as well as the evaluation community of three truths identified by the authors: (1) “large-scale field experiments are feasible,” (2) “experiments answer important questions,” and (3) “high quality evidence matters” (pp. 424–426). This does not mean that it is easy to generate support and funding for experiments, to recruit and create the enduring partnerships with practitioners that are essential for running experiments, or even to design and carry out the evaluations themselves. However, we now have a rich set of case studies and deep technical expertise to draw on that was largely absent in 1974.

In the final chapter of the book, Gueron and Rolston distill the lessons that are embedded throughout this journey from the perspective of MDRC and HHS, respectively. I urge you not to miss this chapter. If you get to the end of the book feeling pretty smug about how far we have come in the fight for reliable evidence, keep reading. The coda will reground you in the work that remains. *Fighting for Reliable Evidence* deserves a place on the required readings list for students of policy analysis, public administration, applied social sciences, and organizational leadership. It also is a must-read for thought leaders and practitioners who are intent on channeling scarce resources to achieve the greatest good.

**REBECCA A. MAYNARD** is the University Trustee Chair Professor of Education and Social Policy at the University of Pennsylvania, 3700 Walnut Street, Philadelphia, PA 19104 (e-mail: rmaynard@gse.upenn.edu).

**REFERENCES**


