

HOW RESEARCHERS CAN MAKE THE EVIDENCE THEY GENERATE MORE ACTIONABLE

REBECCA A. MAYNARD

Most professional evaluators seem highly motivated by the prospect that the results of their studies will improve outcomes for individuals and for society. However, recent work on the use of evidence shows a surprisingly large gap between these aspirations and the instrumental use of the evidence produced (Honig and Nitya 2012; Farley-Ripple and Jones 2015; Penuel and others 2017; Tseng and Coburn 2019). Concurrently, the public policy and philanthropic communities have increased substantially their emphasis on evidence-based policy and practice. In turn, this has fueled the imperative and seeded opportunities for rethinking approaches to evidence production to yield better and better used results (Haskins and Margolis 2015; Fedorowicz and Aron 2021; Zhang and others 2017). Practitioners and evaluators are having to adapt their usual practices to these new standards for production and use of evidence.

In this paper, I share snippets of my personal journey from a classically trained micro-economist to a “roll up your sleeves” evaluation partner of practitioners who are intent on improving the efficiency and effectiveness of their programs. I still use, primarily, the tools acquired in graduate school (updated as the evaluation field has matured). However, over time, I have found myself “flipping the script” to shift greater priority toward

the evidence needs of the program partners, while still ensuring I maintain high standards for evidence and meet essential requirements of funders. As a result, I have been able to sidestep many of the commonly perceived barriers to engaging in more actionable evidence building enterprises (Brooks and others 2019; Donaldson, Christie, and Mark 2015).

My graduate training was a fairly traditional mix of economic theory and its application in public finance policy, coupled with an atypical experience working on the Rural Negative Income Tax Experiments (commonly referred to as the NIT). This assistantship afforded me the opportunity to work hand in glove with faculty members, peers, and field staff who were breaking new ground. None of us had experience running large-scale social experiments or designing and carrying out complex study designs—for example, those entailing large amounts of primary data collection and the design and implementation of statistical programs to address the nuances of study designs that included multiple sites, children nested within families, longitudinal surveys, and merging data from disparate sources (Levine and others 2005; Maynard 1977). Federal funders, program administrators, field data collectors, senior faculty research directors, and graduate students all were breaking new ground in our jobs. This was a quick lesson in the value of stakeholder engagement and teamwork.

Working on the NIT provided the foundation for a somewhat unusual career path. I became skilled at designing evaluations that yielded credible impact estimates but rarely provided much detail about on-the-ground experiences of those in the study sample to contextualize the findings or guide improvements in the design or implementation of the policies. Over the course of my career, I have encountered a few defining experiences that caused me to rethink how I approach my work as an evaluator. Here are three examples.

Example 1. While at Mathematica, I worked on a study looking at the impact of supported work programs for ex-addicts, ex-offenders, school dropouts, and welfare recipients. In advance of our revealing the actual impact findings, field staff at MDRC who oversaw the ten supported work programs in our study ranked the programs in order of “expected” impacts—rankings that seemed reasonable to those of us on the study team. We all were shocked when the study findings revealed that these rankings were the opposite of reality. This was a reminder that easily observable markers of program performance may not be reliable indicators of program impacts (Gueron and Rolston 2013; Maynard 2015).

Example 2. I was working on evaluations of programs intended to lower the incidence of repeat pregnancies and improve economic prospects for teen mothers and their children. At the time, conventional wisdom among the social welfare and academic communities was that these young mothers needed supportive, nurturing environments (Quint and Riccio 1985). At the same time, policymakers were arguing that it was important to institute financial incentives and basic supports that would encourage young single mothers on welfare to pursue education and training that would improve their long-run economic prospects.

Unsurprisingly, we encountered strong community pushback on plans for a federally funded randomized controlled trial to determine the consequences of eliminating the current exemption from work requirements for first-time teenage parents applying for federal welfare benefits. In response, federal program staff, our research team, and the local welfare office staff co-developed a logic model to guide the study design and implementation. This model included engaging with the teenage mothers and their caseworkers and welfare office directors throughout the course of the study—hearing their perspectives and capturing their first-hand experiences in ways that proved to be critically important for making meaning of the study findings.

By the end of the study period, a majority of the young mothers subjected to the new requirements for continuous engagement in school, work, and/or training as a condition of receiving the full welfare benefit reported feeling the policy was not only fair but helpful. The reasons they gave all related to the fact that the requirement was accompanied by supportive case management and other services to help them raise their child and improve their own lives (Maynard 1995; Maynard and Rangarajan 1994). Many also saw the requirement as an “escape hatch” from controlling partners or family members who preferred the young mothers remain dependent.

By broadening our evaluation agenda to include a rich (though not outrageously costly) qualitative study component, we were able to piece together a coherent explanation for why a policy that initially seemed draconian became preferred by both welfare workers and program participants. Most notably, we conducted periodic rounds of case conferences with welfare office staff (that is, meetings where case workers provide status updates on selected cases and invited reactions/input from peers and supervisors); focus groups with teenage mothers; and periodic meetings with case managers

and office directors (for example, to discuss policies, practices, challenges, successes) (Polit 1992).

Example 3. A third defining experience was the design and implementation of evaluations of four of the Title 10 Abstinence Only Education Evaluation programs (Devaney and others 2002). This pulled on every lesson of my then twenty-something years of program evaluation experience. The Title V Abstinence Only Until Marriage legislation (Section 510 (b) of Title V of the Social Security Act, P.L., 104–193, provided major funding for health and sex education programs that promoted abstinence until marriage as the only way to avoid sexually transmitted diseases and unwanted pregnancies, and it prohibited any teaching related to contraceptive effectiveness or access. This policy had exceedingly strong support from the Christian Coalition and other conservative groups like Focus on the Family, and equally strong opposition from the health and sex education advocates and organizations like Planned Parenthood and the Alan Guttmacher Institute (Darroch and others 2000).

To successfully design and launch an evaluation, it was critical to fully understand the logic behind the views of those who endorsed the policy, those who opposed the policy, and those in between, and to ensure the perspectives of all three groups were well represented in the study design. Moreover, the study design needed to be recognized by all sides as nonpartisan; the goal of the study was to learn about and understand the consequences of the policy relative to the status quo and to be able to communicate the findings to all sides in a manner that was respectful and useful to them. This meant we needed to design a study that was centered on the common goal of supporting the sexual and social emotional health of youth.

To achieve this goal, we needed to invest heavily in stakeholder outreach at three levels. First, we needed to engage with the various constituents—for, against, and neutral toward abstinence-only education—and incorporate their beliefs and fears into a program logic model. Second, we needed to recruit and engage communities that had received Title 10 funds to partner with us in experimentally testing the abstinence-only programs against their usual health and sex education practices—a task that required building trust and a shared commitment to the evidence-building agenda. Third, we needed to make sure we maintained sufficient communication with sex education providers, students, and school administrators to ensure we would understand and be able to communicate to others the mech-

anisms through which the Title 10 programs altered (or did not alter) outcomes for students. This required a lot of listening, learning, and documenting to arm the study team with knowledge to support a cogent, evidence-based interpretation of the findings—whatever they were. Community members needed assurances that our goal was not to “prove” anything but, rather, to learn what difference the choice of sex education strategy made for students and to advance understanding of health and sex educators, parents, and community members about what it was about one approach versus the other that made the difference (assuming there was one). In the final analysis, there is neither evidence that abstinence-only education improves outcomes for youth nor evidence that it is harmful relative to comprehensive sex education (Ott and others 2007; Trenholm and others 2007).

Through these types of experiences, I have developed a vigilance around designing studies that will be useful to the practitioners and policymakers regardless of whether the tested concept yields the expected result. First and foremost, the goal is to learn something that will help them improve their program, policy, or product. This was an explicit motivation behind the development and improvement research I worked on in partnership with colleagues at Abt Associates, doctoral students at the University of Pennsylvania, and Year Up staff (Britt and others 2021; Fein and others 2020). This work produced concrete illustrations of some of the most important lessons of my career—lessons for researchers and researchers in training, lessons for practitioner partners, and the inevitability of limitations in any evidence-building enterprise.

The following are some of the major take-aways from the one slice of evidence building we featured in the Year Up case study.

1. Engaging in a collaborative process to prioritize investments in evidence building was extremely beneficial. It not only grounded the research team in the intricacies of the program—its goals, culture, participants, staff, partners, opportunities, and constraints—it paved the way for collaboratively identifying and prioritizing “rooms for improvement” in Year Up’s Professional Training Corps program model that would be the focus of quick-turnaround evaluations to inform short-run improvement efforts and lay the groundwork for longer-term plans for evidence building.

2. Designing evaluations to inform program improvement regardless of the study findings can improve both the efficiency of evidence generation and the likelihood that the resulting findings will be used. This begins with exploring what questions practitioner partners would want answered under a range of plausible study findings—for example, if a strategy for boosting program retention proved not to be effective or even harmful.
3. It was important to balance staff and participant burden with its potential return. For us, this meant prioritizing essential data needs with “nice to know” data dreams. Having strong, trusting practitioner partners, we were able to fill many data gaps through opportunistic encounters and very targeted “mini data collection efforts.” We also worked with program staff to plan implementation strategies that were minimally disruptive to operations and respectful of the informants, including providing tokens of appreciation such as modest gift cards or food.
3. We held fast on study design features critical for to generating credible evidence to inform program practices features, while being flexible in their implementation. For example, we insisted on testing program improvement strategies using a randomized controlled trial. Yet, we were very flexible on issues like sample size, assignment ratio, and timing of randomization, we provided options for “pairing” and “separating” participants, and we accommodated staff recommendations to exclude certain participants from randomization.
4. We not only encouraged programs to iterate on their improvement strategies between testing cycles, but we actively facilitated information sharing among program staff within and across sites between cycles regarding their professional judgments about what was and was not working. We also encouraged staff to modify their improvement strategies based on the shared experience. We did not provide interim impact findings, which (by design) were intended to inform summative conclusions about the progression of improvements over time.
5. All study findings were shared first with Year Up staff working directly with the study team in “draft” briefing documents—one package targeted on senior management and a second package

targeted at site staff. Each packet consisted of three parts: 1) a very summary “pre-read,” which was a high-level overview of the study, findings, and recommendations; 2) a slide-deck to support an on-line briefing; and 3) a more detailed slide-deck sent out immediately following the briefing containing more detailed information about the study design and findings.

6. We collected artifacts created and/or used by program staff to support their program improvements and, with encouragement of the Year Up national staff and site program directors, created an indexed compendium of tools to support Year Up’s use of the study findings to improve outcomes throughout its network of local programs.

The research team has prepared detailed technical reports on the study background, methods, and findings (Fein and others 2020), presented on its work at numerous professional conferences, and published academic and policy articles on the study methods and findings (Britt and others 2021).

Variations of these evaluation strategies are reflected in many other recent and ongoing evidence generation efforts by social scientists trained in various disciplines and research traditions—a sample of which is reflected in the twelve case studies commissioned as part of the Actionable Evidence Initiative supported by the Bill and Melinda Gates Foundation. For example, Julie Martin and Elisabeth Stock (2021) conducted a rapid cycle evaluation of an interactive digital learning strategy for improving academic mastery and social emotional learning skills of high poverty middle school youth; D. Bradley and S. Burkhauser (2021) report a partnership among the Mid-West Regional Educational Laboratory (REL Midwest), the Minnesota Department of Education, and four state-approved alternative education programs for students judged to be at-risk of not graduating from high school in a Networked Improvement Community aimed at identifying evidence-supported strategies for improving instructional practices through small-scale, quick-turnaround testing of strategies for collecting and using data to make real-time shifts in strategies that improved student outcomes; and P. York (2021) reports on a partnership among two programs serving youth aging out of foster care—First Place for Youth and Gemma Services—to use evaluation and data science to build predictive and prescriptive models that use real-time data to guide case management and support of youth in ways that improve their outcomes.

Increasingly, many in the evaluation community have been moving toward more nimble applications of their research tools than is common. Historically, evaluators have gravitated toward using their tools in service of independent evaluations aimed at general knowledge development or applying them in service of “thumbs up or down” accountability. Increasingly, evaluators are recognizing the power in applying their skills and resources to support practitioners to improve their decision making—thus, increasing their emphasis on actionable evidence. Through effective curation and dissemination of the products of such evidence building, we likely also will accelerate the pace and impact basic education and social science knowledge. This does not require a shift in methods; rather, it requires asking the right questions, gathering and using credible data, matching the questions to the methods, and reporting in a timely and accessible format. The odds of producing actionable evidence and having it used goes up significantly when practitioners are invested partners in the effort (Brooks and others 2019).

It has been amazing to watch even strong, long-established organizations learn and improve as they work to craft a strategic evidence plan. Even more rewarding has been the excitement among program partners once they have prioritized their needs for evidence and arrived at creative options for generating and using it.

REFERENCES

- Bradley, D., and Burkhauser, S. “Lessons from a Minnesota Networked Improvement Community.” Boston, MA: Project Evident. September 21, 2021. <https://static1.squarespace.com/static/58d9ba1f20099e0a03a3891d/t/61426c66ac7327450a8141a7/1631743085075/MN+SAAP+NIC+Actionable+Evidence+Case+Study+Sept21.pdf>.
- Britt, J., Fein, D., Maynard, R., and Warfield, G. “Improving Academic Success and Retention of Participants in Year Up’s Professional Training Corps.” Boston, MA: Project Evident. July 7, 2021. www.projectevident.org/updates/2021/7/26/improving-academic-success-and-retention-of-participants-in-year-ups-professional-training-corps.
- Brooks, J. L., Boulay, B. A., and Maynard, R.A. “Empowering Practitioners to Drive the Evidence Train: Building the Next Generation of Evidence.” Boston, MA: Project Evident. March 27, 2019. https://static1.squarespace.com/static/58d9ba1f20099e0a03a3891d/t/5cf81b7b4e259a0001c7a4a5/1559763835719/Empowering+Practitioners+to+Drive+the+Evidence+Train+_+Building+the+Next+Generation+of+Evidence.pdf.

- Darroch, J. E., Landry, D. J., and Singh, S. "Changing Emphases in Sexuality Education in U.S. Public Secondary Schools, 1988–1999." *Family Planning Perspectives* 32, no. 5 (2020): 204–11, 265. PMID: 11030257.
- Devaney, B., Johnson, A., Maynard, R., and Trenholm, C. *The Evaluation of Abstinence Education Programs Funded Under Title V Section 510: Interim Report*. Princeton, NJ: Mathematica Policy Research, Inc. 2002. <https://eric.ed.gov/?id=ED471712>.
- Donaldson, S. I., Christie, C. A., and Mark, M. M. (Eds.). *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations*. Los Angeles, CA: Sage Publications. 2015.
- Farley-Ripple, E., & Jones, A. R. (2015). Educational Contracting and the Translation of Research into Practice: The Case of Data Coach Vendors in Delaware. *International Journal of Education Policy and Leadership* 10, no. 2, n2. <https://files.eric.ed.gov/fulltext/EJ1138590.pdf>.
- Fedorowicz, M., and Aron, L. Y. *Improving Evidence-Based Policymaking: A Review*. Washington, DC: Urban Institute. 2021. www.urban.org/sites/default/files/publication/104159/improving-evidence-based-policymaking-a-review_0.pdf.
- Fein, D., Maynard, R., Baelen, R., and others. *To Improve and to Prove: A Development and Innovation Study of Year Up's Professional Training Corps*. Rockville, MD: Abt Associates Inc. 2020.
- Gueron, J. M., and Rolston, H. *Fighting for Reliable Evidence*. New York: Russell Sage Foundation. 2013. <https://files.eric.ed.gov/fulltext/ED610239.pdf>.
- Haskins, R., and Margolis, G. *Show Me the Evidence*. Washington, D.C.: Brookings Institution Press, 2015.
- Honig, Meredith I., and Venkateswaran, N. "School-Central Office Relationships in Evidence Use: Understanding Evidence Use as a Systems Problem." *American Journal of Education* 118, no. 2 (2012): 199–222. www.journals.uchicago.edu/doi/full/10.1086/663282?casa_token=-u6UPJnBiToAAAAA:McLtNdE4zBDz4unl7v1KL25UgBoJNmuIeBHhehp13-PGuRCm0r0-wYT_OUzpeXRnc-ylCpok9TqA.
- Martin, J., and Stock, E. *Examining Family Playlists' Impact on Student Social Emotional Learning and Science Mastery through Short-Cycle RCTs*. Boston: Project Evident. 2021. <https://projectevident.org/wp-content/uploads/2022/08/PowerMyLearningActionableEvidenceCaseStudySept21.pdf>.
- Maynard, R. "Teenage Childbearing and Welfare Reform: Lessons from a Decade of Demonstration and Evaluation Research." *Children and Youth Services Review* 17, no. 1–2 (1995): 309–32. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.574.7021&rep=rep1&type=pdf>.

- Maynard, R. A. "Review of *Fighting for Reliable Evidence* by J. M. Gueron and H. Rolston." *Journal of Policy Analysis and Management* 34, no. 1 (Winter 2015): 243–47. www.jstor.org/stable/43866094?casa_token=Qhghttp23KlgAAA:AAA:lRAs6djZQOf3K6cRxcuCPGi8c-8BM3ATmKrhU5enZgnutQu9bqbTi3CPepp77WyZPHdHrdF53X8YfIOjBm23zJ7VnCS50xtuEvTK7YHRqVV4_SzfP7g.
- Maynard, R., and Rangarajan, A. "Contraceptive Use and Repeat Pregnancies among Welfare-Dependent Teenage Mothers." *Family Planning Perspectives* (1994): 198–205. <https://aspe.hhs.gov/sites/default/files/private/pdf/74136/report.pdf>.
- Maynard, Rebecca A. "The Effects of the Rural Income Maintenance Experiment on the School Performance of Children." *American Economic Review* 67, no. 1 (1977): 370–75. <https://aspe.hhs.gov/sites/default/files/private/pdf/74136/report.pdf>.
- Ott, M. A., and Santelli, J. S. "Abstinence and Abstinence-Only Education." *Current Opinion in Obstetrics & Gynecology* 19, no. 5 (2007): 446–52. <https://doi.org/10.1097/GCO.0b013e3282efdc0b>.
- Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., and others. "How School and District Leaders Access, Perceive, and Use Research. *AERA Open* 3, no. 2 (2017). 2332858417705370. <https://journals.sagepub.com/doi/pdf/10.1177/2332858417705370>.
- Polit, D. "Barriers to Self-Sufficiency and Avenues to Success among Teenage Mothers." Princeton, NJ: Mathematica Policy Research, Inc. 1992. TPD# 2584. <https://aspe.hhs.gov/reports/barriers-self-sufficiency-avenues-success-among-teenage-mothers>.
- Quint, J. C., and Riccio, J. A. "The Challenge of Serving Pregnant and Parenting Teens. Lessons from Project Redirection." 1985. <https://files.eric.ed.gov/fulltext/ED275777.pdf>. Updated version 1988. <https://files.eric.ed.gov/fulltext/ED299444.pdf>.
- Levine, R. A., Watts, H., Hollister, R. G., Williams, W. and others. "A Retrospective on the Negative Income Tax Experiments: Looking Back at the Most Innovative Field Studies in Social Policy. In *The Ethics and Economics of The Basic Income Guarantee*. Ashgate, 2005. <https://works.swarthmore.edu/fac-economics/347/>.
- Trenholm, C., Devaney, B., Fortson, K., and others. *Impacts of Four Title V, Section 510 Abstinence Education Programs, Final Report*. Princeton, NJ: Mathematica Policy Research. 2007. www.mathematica.org/publications/impacts-of-four-title-v-section-510-abstinence-education-programs.
- Tseng, V., and Coburn, C. "Using Evidence in the US." In *What Works Now*, edited by A. Boaz and H. Davies, 351–68. Policy Press: 2019. <https://>

- wtgrantfoundation.org/library/uploads/2019/09/Tseng-and-Coburn_What-Works-Now.pdf.
- York, P. “Generating On-Demand, Actionable Evidence for Front-Line Practitioners at First Place for Youth and Gemma Services. Boston, MA: Project Evidence. 2021. www.projectevident.org/updates/2021/7/26/generating-on-demand-actionable-evidence-for-front-line-practitioners-at-first-place-for-youth-and-gemma-services.
- Zhang, X., Griffith, J., Pershing, J., Sun, J., and others. “Strengthening Organizational Capacity and Practices for High-Performing Nonprofit Organizations: Evidence from the National Assessment of the Social Innovation Fund—A Public-Private Partnership. *Public Administration Quarterly* (2017): 424–61. www.jstor.org/stable/pdf/26383392.pdf?casa_token=yLdRkiQ4D_AAAAAA:7-B5AdFYdjNLX2HyCeIKgeZNT3dDaWTKtaPTj3oVPfq41qUZJRg-ZG2BQ_xkHs5iRo7cQA2jFGNmFyYgkC-pt6TIGRuYjU2ZTntrBMBdY17lvAcufWQ.