

# **Increasing the Impact of Program Evaluation: The Importance of Challenging Assumptions and Incorporating New Perspectives**

**Bridget Terry Long**

2022 Recipient of the Peter H. Rossi Award for Contributions  
to the Theory or Practice of Program Evaluation

Remarks for Publication in *Evaluation Review*  
Originally given at the APPAM Research Conference, November 17, 2022

Thank you to the award committee. I feel deeply honored to receive this prize. When I was first informed about winning the award, I took note of the past winners, and I am flattered to be in such esteemed company. The past winners include people who have made major contributors in research and methodology, such as Thomas Cook, Becka Maynard, and Howard Bloom. The list of past winners also includes people who have helped to advance the use of evidence in policy and practice, such as Judy Gueron, President Emerita of MDRC, and Russ Whitehurst, founding director of the Institute of Education Sciences (IES) at the U.S. Department of Education.

## ***Recognition of the Work of Peter Rossi***

This award recognizes the incredible work and impact of Peter H. Rossi. As I have learned more about Dr. Rossi, I have admired his efforts to “take to the streets” in order to study communities of interest, such as when he and a team of researchers documented the changing face of American homelessness in the 1980s by going into communities in Chicago. I also admire Dr. Rossi’s focus on evidence—focusing on what actually works is essential if we are going to help people and improve conditions. We do a disservice when we allow programs that don’t work to stay in place rather than exploring alternatives that might be effective.

It is another aspect of Dr. Rossi’s work that I will use as a jumping point for my remarks. Many are familiar with Dr. Rossi’s paper, *The Iron Law of Evaluation*. In it he wrote that “the typical impact assessment of a public social program finds that the program is either ineffective or only marginally effective,” and he largely blamed this on the fact that effective programs are difficult to design. Years later in a paper, entitled “The ‘Iron Law of Evaluation’ Reconsidered,” which Dr. Rossi presented at the 2003 APPAM conference, he provided an updated, and much more positive, perspective, noting that we are in fact “learning how properly to design and implement interventions that are effective.” Dr. Rossi credited the “impressive change in the evaluation field” to the “considerable growth in the sophistication of evaluators and in the methodology of evaluation.” He went on to say, “The best of evaluators simply know a lot more about how to design credible impact assessments and have at their command technical tools that make it possible to analyze data in much more sophisticated ways.” (p. 4)

This is a good starting point for us today, as I reflect on Dr. Rossi’s statements now nearly 20 years later. To give you a bit of a preview, I largely agree with Dr. Rossi on the tremendous advancements in evaluation and impact analysis due to improvements in our research design and methods. The capacity of the field has grown, and we have developed a wide range of tools and approaches to better understand the impact of programs, policies, and interventions. We have become much more thoughtful about how different types of evidence contribute to our understanding, and

there is also much more attention to being precise in how we talk about findings—whether they do or do not credibly support a particular theory or hypothesis.

However, advances in our research designs and methods are only part of the reason the field has progressed, and importantly, I'm not sure advances in our research designs will be the key to continuing to make leaps in program evaluation. We need to do more, and using examples from my own work, I want to emphasize the importance of three things: asking different questions and challenging long-held assumptions; welcoming new perspectives into the research community; and incorporating the voices of practitioners and communities we seek to study. These are approaches that could help to support the further advancement of not only evaluation and impact analysis but also the creation of programs and policies could improve outcomes.

### ***My Professional Development and Career***

My own development as a scholar has happened during amazing time of progress in program and policy evaluation. In education, there has been tremendous growth in the production and use of research as we have expanded the frontier of what we know and understand about how to improve student outcomes. I have benefited greatly from some of the people who pushed the frontier to develop clever ways of doing evaluation and impact analysis.

My start began as an undergraduate at Princeton, where I took an economics course with David Card. David modeled for me clear thinking with a grounding in economic theory and methods but also creativity and a strong interest in the real world. He, of course, won the Nobel Prize in Economics last year with Josh Angrist and Guido Imbens, both of whom I have also had the opportunity to learn from. Together, they pioneered “natural experiments” and “quasi-experimental” methods—essentially, using real-world events and shocks to design credible ways to isolate causal relationships. During my time as an undergraduate student, David was also working with Alan Krueger, who must also be mentioned with this group of luminaries, to study the relationship between the minimum wage and employment by exploiting a 1992 change in the minimum wage in New Jersey (Card and Krueger, 1994). Needless to say, if this is what economists did, then I was hooked. I came up at a time that appreciated the real world as a laboratory for understanding social phenomena.

I then went on to Harvard to get my Ph.D. in Economics, working with Caroline Hoxby. Among her early studies, her “rivers” paper astounded with the idea of using exogenous variation in topography—i.e., rivers—to derive instruments that partially determine district size (Hoxby, 2000a). In other work that was completed during the time I was a graduate student, Caroline studied the effects of class size and composition on student achievement using national population variation (Hoxby, 2000b). The Economics of Education grew quickly as a result of new data and new methods that could be applied to examine important questions.

After graduation, I entered the academic profession having seen a rich body of work focused on critical issues related to inequality and social policy using a rapidly advancing methods and research designs. I too was committed to engaging in the real world to understand a long list of issues related to education, and so I became an Assistant Professor at the Harvard Graduate School of Education, a place where I could develop a rigorous research agenda while also being valued for the impact of my work and my engagement with educators, policymakers, and families.

My research agenda as a professor started with secondary data analysis, trying to utilize changes and differences across states to understand the impact of different policies. The data sets

available through the National Center for Education Statistics (NCES) were the foundation of so much of my and others' research on education.

For my second chapter, as additional data sources became more readily available with the expansion of administrative data sets, I moved on to study the impact of programs and investments on student outcomes. With access to comprehensive data systems for Ohio, Florida, and Tennessee, and working with Eric Bettinger, I was able to back out quasi-experimental approaches by learning about the educational context of each state. We worked to contribute to the field's understanding of the effects of postsecondary developmental and remedial education and the impact of different kinds of instructors, among other issues.

And then for my third chapter, we started to carry out experiments. We designed and tested programs and interventions focused on helping families with financial aid forms and starting college savings accounts. Using randomized controlled trials, partnering with organizations in the field, and capitalizing on administrative data, we conducted studies that tested new theories as well as provided practical and policy solutions to formidable barriers.

It has been during my latest chapter, my current "day job" as Dean of the Harvard Graduate School of Education, that I have been able to take a larger perspective on the "big picture" of evaluation and research, as I attempt to support, advance, and steer an institution of scholars and practitioners to create, use, and reflect on evidence.

### ***Lessons and Looking Ahead***

What have I taken away from this grand tour of impact analysis? First using many different approaches in my own work, and then what have I observed as Dean?

Based on my own experiences and what I have seen in the field more broadly, this has been an incredible time when our methods and research have advanced substantially. However, I would say that this is only part of the reason the field has progressed. Missing from a narrative that only emphasizes our research designs and methods are other important approaches that are key to us continuing to push the frontier of this field. Stated another way, research strategies and techniques alone will not get us answers to the ultimate question of what works, and more than that, *what could work*.

The most impressive work I've seen from my colleagues and from scholars across many fields are about more than just innovative research design. This tells me that we need to do more to advance the field—things will push us to reconsider our theoretical frameworks and the hypotheses we test, push our understanding and questions about mechanisms, and spur us to consider new ideas of what programs and policies might work.

Before jumping into the specific points, let me first acknowledge the degree of difficulty but also urgency of this work. One thing the current state of evidence clearly tells us is that the challenges many of us focus upon are stubborn—whether they be in education, health, workforce, family and child policy, poverty, environmental policy, or migration. In education, there are persistent gaps in opportunity and success at all levels of education—from the insufficient availability of high-quality early childhood education; to inadequate supports and rigor in K-12 classrooms; to uneven access and low completion rates in higher education.

And the pandemic only served to shed new light to long-standing inequities. As Paul Reville, my colleague at the Harvard Graduate School of Education and former Massachusetts Secretary of Education has said, “It’s as though a big wave has pulled back the sea revealing the ocean floor and all its disturbing realities that had heretofore been hidden beneath the surface of the water.” More than that, what we have experienced during the last several years in education is not just dramatic—the decline in achievement levels, as measured by NAEP scores for instance, has been historic. Trends suggest that the pandemic erased the decade of progress in the late 1990s to the early 2000s. And the reductions in test scores are just the tip of the iceberg—we should also worry about broken connections, a vulnerable profession, and deepening inequality.

There is incredible urgency in the need to find what works and to have that evidence influence our policies, programs, and practices, but the challenges facing education—and many other areas of social policy—are persistent and stubborn. Many things have been tried, and we lament the same sentiment Dr. Rossi expressed in 1987 and again in 2003: “the majority of impact assessments end up with findings of no effect or substantively marginal effects” (p. 4).

I would posit that while an abundance of data and the clever use of research methods will help us to understand multiple dimensions of the problems, we need more to make things better. But that does not mean that we are without hope. Let us not underestimate the many programs and policies that have been found to produce positive effects, especially in recent years. Perhaps part of the reason I have been given this award is because of my contributions to the positive side of the ledger.

What will it take to continue advancing the field? I preface this by acknowledging that these ideas are not new. In fact, I have benefited from others who have emphasized the importance of these approaches. Please take these examples from my work as a few illustrative data points to help underscore my suggestions.

### ***Reconsidering Our Assumptions and Models***

The first point is the importance of asking good questions. Of course, that’s not revolutionary, and we know good research is guided by good questions. But I want to call attention to the fact that too often, we shy away from questioning what has been accepted as fact or law. This can include the models and assumptions that dominate our disciplines and fields. Let us also recognize that too often we simplify the issues that we are studying to such a degree that it prevents us from identifying new insights. To take leaps forward as a field, we need to question and test whether new paradigms are relevant.

To provide a bit more context for the first point, let me take you back to the early days of my career as a faculty member. My early work focused on the impact of financial aid, which largely came from a personal suspicion that unequal college access was rooted in issues of affordability. Researchers long before me had examined the effects of a range of government and institutional financial aid programs, and yes, money seemed to sometimes matter. However, one takeaway from the literature is that the largest financial aid program, the federal Pell Grant, seemed to have little impact on increasing college enrollment, though it is incredibly difficult to isolate its effects for causal analysis. Additionally, data revealed that individuals from low-income families who were eligible for need-based financial aid often did not apply for or use the benefit.

How do we interpret those findings? You could assume that low take-up rates suggest that many low-income students just didn’t want to go to college. Of course, any rational person eligible for a

\$3,000 grant to go to college would take the money if it was offered to them, wouldn't they? And if you are satisfied with that statement, then the next step in the logic is that if we want to improve educational outcomes, we shouldn't waste resources on high school seniors or college-age adults. Perhaps we should focus on younger children instead. And for a while, that narrative pervaded the discourse about how to improve education.

Now, I am not at all saying that we shouldn't invest seriously in early childhood education. We absolutely should, and at a high level, but consider for a moment the strong assumptions that are being made when the prevailing narrative in the field is to not bother investing in young adults—that low-income families don't care about education at older ages, that they don't care about opportunity. That's not what I saw in my family, where the outcome of whether one of my cousins went to college or not were far more complicated than that. It's not what I saw in the communities I've worked. For instance, at the main branch of the Boston Public Library, which is near my home, I've seen the long lines of people waiting to use the public computers, including many youth who looked like they were high-school age. The assertion that low take-up rates were due to lack of intent struck me as false. Are gaps in college access really due to preferences and choice? Is it really too late to do anything for older children and adults?

If you know anything about my research, you know that is not the case. You know that there are cost-effective ways to increase not only college access but also college success. So let's go back to our models. They are important tools to help us try to understand complex phenomena. We start by simplifying things, but we must never forget that they are an extreme simplification of the world. Looking more closely at our models about decision making, they all start from a place of assuming the decision-maker has information about the decision they need to make. In fact, the basic model assumes perfect information, but is that reasonable? The answer is clearly no. Moreover, Richard Thaler and Cass Sunstein in their 2008 book, *Nudge*, pointed out a host of ways people do things that seem counterintuitive and the fact that choice architecture is important.

Making the assumption that low-income people just didn't want to go to college a woeful example of letting our basic models overpower our sense of curiosity. Assuming that there was only one explanation for college-going behavior is too simplistic. We need to ask more insightful questions. We have to question and test our assumptions, even those long held.

My design and implementation of an intervention that focused on the federal financial aid form (or FAFSA) illustrates this point. College access organizations had long complained that the FAFSA was too complicated and deterred students from applying for financial aid. College administrators lamented that students often lost their aid awards due to not understanding that they need to submit renewal applications each year. Therefore, working with Eric Bettinger and Phil Oreopoulos, we conducted a randomized control trial in partnership with H&R Block that significantly streamlined the process of completing the FAFSA. We created a process where a family's tax information was pre-populated on the FAFSA, so instead of taking hours and hours to complete, the FAFSA was filled out in an average of eight minutes or less. The result was that the graduating high school seniors whose families received the treatment—assistance completing the FAFSA—were 28 percent more likely to go to college than the control group, which did not receive the special help (Bettinger, *et al.*, 2012) —which demonstrates just how much of a barrier this form is for most families. The problem was not motivation or intent; in fact, we tracked the students and found that they persisted at much higher rates than the control group as well.

To be clear, I'm not being critical of the use of models. They are a necessary learning tool. They help us to organize and isolate issues we are trying to study. But we also need to make sure we recognize them for what they are—just a simple beginning—and be conscious of the need to move beyond the simple as we apply our theories and models to real people and try to understand their behavior. And as our knowledge grows, we must revisit what were previously considered baseline models and assumptions.

### ***The Importance of Diverse Perspectives***

Next, it is important to recognize that in order to improve not only our questions, but also our designs, methodologies, and interpretation, we must welcome new perspectives that have traditionally been underrepresented in our disciplines and fields. Past research can sometimes reflect a narrow set of perspectives, ideas, and assumptions given who was at the table when the work was conducted, but it is through a broader range of lived experiences, viewpoints, and perspectives that we can improve the relevance and insights our work provides.

To illustrate what I mean about the narrow set of perspectives that typically sit around our tables, let me give you an example from my research. While presenting a paper on how colleges and universities attempt to support academically-underprepared students, I discussed the prevalence of developmental and remedial postsecondary courses. Audience members were surprised at just how many students in higher education were not prepared for college-level material (40-45 percent, and even higher at community colleges). I was asked the question of why this was important—surely not everyone needed calculus, right? It was then that I had to describe how developmental ed could include not only geometry but also arithmetic. It struck me then how none of the people around the table were even aware of the common experience of most students in higher education, having come from a select group of schools with advanced preparation in quantitative fields. It took a while to help most academics understand why focusing on developmental education was of critical importance.

More broadly, my life experience is one that is severely underrepresented in this research domain, and that has had an effect on how I approach my work. When I enrolled in graduate school in the Economics Department at Harvard, I was one of only eight women out of a class of over 45. Moreover, I was the only African American student in the entering cohort for economics that year. I found that experience to be incredibly jarring. But over time, I also realized that there was something special about the perspective that I was bringing to the table due to many experiences I had that differed from my classmates.

As I noted before, I had gone to Princeton for undergrad so had a strong academic preparation, but I sit at the crossroad of many life experiences. As a Black woman who grew up in the 70s and 80s, I floated between predominately white and predominately Black environments; mixing life in the suburbs with extended times in urban and rural communities. I have interacted with both the rich and the poor, as well as everyone in between. Additionally, I had a very different family history than most as I could trace my great grandmother's emancipation from being enslaved to the Jim Crow era in the South for my grandparents and parents. I had grown up hearing oral histories about desegregation and the Civil Rights movement that differed from what was written in most books, especially thirty years ago.

My different perspective and life experiences manifested in how I participated in class. Years later, a faculty member from graduate school shared with me that I the kinds of questions I was asking as a student were very different than the ones she heard from others, and I can see why. Having spent

time in lots of different environments and interacting with people from all walks of life made me naturally question simple assumptions about how people think and what they value. If the simple model being presented did not ring true with what I had seen in the world, I questioned it. Even in my work focused on education, I would often question the assumptions of models and the interpretation of results if they did not comport with what I learned from my mother about the inner workings of schools based on her many years as a high school teacher.

Ultimately, as I made in point one, that questioning of our field's assumptions and approaches is incredibly healthy. The most impactful work uncovers new insights, and that often comes from improving our models to better reflect the diversity and nuances of our complex world. Today, some of the best research takes note of how communities actually work, and making sure that our questions, hypotheses, and models are in tune with the reality of the world, rather than a narrow set of outdated assumptions, is how we advance. Some of the best research in any field comes from the introduction of new perspectives and opinions.

### ***Listening To and Engaging Practitioners and Communities***

For my final point, let me emphasize the wealth of knowledge that resides in the communities we seek to study. Let me give you some examples where this has been essential in my own work.

When overseeing an evaluation of a program designed to help families save from college, I had the opportunity to observe the meetings and talk with the staff. I quickly learned that the most important part of the intervention was the mother-to-mother exchanges about the importance of the program. In seeking to improve the model, we could have designed all sorts of fancy information sheets and protocols, but the true answer focused on participant-to-participant exchanges and networks.

For another program, when I was designing an intervention I planned to evaluate, I had to work closely with schools to get parents to participate. Quickly, I discovered that the schools knew little about how to get parents to come to workshops. It was from talking to parents during one of our pilot sessions that we confirmed some things we already knew—such as the difficulty of parents trying to balance school meetings with multiple jobs. But we also learned about how the school choice system in the district results in families living far away from their children's schools and the fact that public transportation did not make it feasible for them to attend evening events. As a result, we moved the workshops to public libraries and community centers in the neighborhoods of the families we wanted to target rather than having the workshops be school based.

And returning to the series of studies I did on remedial and developmental education, it was from talking to the staff at public colleges in Ohio who run these programs that Eric and I discovered the key to our estimation strategy. The assignment policy for each college differed considerably, and time and time again, staff reported that the test and GPA cutoffs used to determine course placement were a function of long-ago administrators who had set them based on opinion rather than evidence, especially due to the fact that years ago there was not an extensive body of research on this topic. The different placement cutoffs approximated something almost random, and so we were able to use quasi-experimental methods to estimate the effects of remedial and developmental programs.

The punchline is: listen carefully to the communities you seek to study and incorporate their insights. In some of my work, their perspectives influenced not only my research design but also informed some of the design decisions that had to be made about the intervention or program I hoped to evaluate.

And this brings us back to Dr. Rossi, who wrote in his “Iron Law” of program evaluation that “the main reasons for a program’s failure to have effects as misunderstanding the problem, designing the wrong intervention, or poorly implementing a good intervention.” The way we address these problems Dr. Rossi has highlighted is by getting close and listening to the communities we seek to study. Are we attempting to solve the right problem? Are we looking at the right outcomes and set of participants? Do our existing programs really work the way we think? It is through conversation and careful listening that we can open up a whole set of possibilities.

And on this point, I am incredibly appreciative of qualitative researchers, especially the many who have influenced my thinking and helped me to craft better hypotheses and questions in my work. I also recognize it takes a deep level of listening and observation and patience that I find incredibly difficult to do in a systematic way, and so I have great respect and admiration for my colleagues who utilize qualitative methods.

The importance of the perspectives of practitioners is also the reason why we need to democratize the understanding and use of evidence. At my institution, the Harvard Graduate School of Education, we just finished redesigning our master’s in education degree program, and a central innovation is the addition of four Foundation courses. One of them is Evidence. It is required of all master’s students, and we ask them: In a landscape where studies emerge each day, how do we become critical consumers of new information and distinguish myth from fact? How do we use existing research to make decisions that are likely to generate the best possible outcomes for our learners? Understanding the answers to these questions is important for all of us, but for education professionals, we are also hopeful that this effort will improve the chance of research use while also bringing practitioners into research conversations about the nuances that truly make a difference.

By broadening the conversation to include voices from practice, policy, and communities, we will be prompted to question our assumptions and build better the models that will serve as the basis for improved program design and evaluation.

### ***Closing***

In closing, good program evaluation, and impact analysis more generally, uses clear, rigorous methods—but excellent impactful program evaluation, the kind that pushes the frontier, addresses the realities of the world, is firmly footed in deep inquiry, and takes in diverse perspectives.

It asks good questions, including ones that challenge the assumptions and models that dominate the field. We must entertain with wonder all the many things we don’t know, and not ascribe human behavior to simple models or make assumptions about what we are observing rather than taking the opportunity to be curious when we find puzzling patterns.

It values new and underrepresented perspectives to improve our questions, models, research design, and interpretation. It does this both by listening closely to the communities we seek to study and the practitioners who work there.

These elements have deep implications for how we could advance not only the field but also the state of knowledge and the program and policy tools at our disposal. Thank you again for this award.



## WORKS CITED

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment." *Quarterly Journal of Economics* 127: 1205-1242.

Card, David and Alan Krueger. (1994) "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, vol. 84(4): 772-793.

Hoxby, Caroline M. (2000a) "Does Competition among Public Schools Benefit Students and Taxpayers?," *American Economic Review*, vol. 90(5): 1209-1238.

Hoxby, Caroline M. (2000b) "The Effects of Class Size on Student Achievements: New Evidence from Population Variation," *Quarterly Journal of Economics*, vol. 115(4): 1239-1285.

Rossi, Peter. (1987) "The Iron Law of Evaluation and Other Metallic Rules." *Research in Social Problems and Public Policy*, vol. 4: 3-20.

Rossi, Peter. (2003) "The 'Iron Law of Evaluation' Reconsidered." Presented at 2003 AAPAM Research Conference, Washington, DC.  
[https://welfareacademy.umd.edu/rossi/Rossi\\_Remarks\\_Iron\\_Law\\_Reconsidered.pdf](https://welfareacademy.umd.edu/rossi/Rossi_Remarks_Iron_Law_Reconsidered.pdf)