



BY

Jennifer L. Brooks,  
Beth A. Boulay, and  
Rebecca A. Maynard

DATE

March 27, 2019

CASE STUDY

# Empowering Practitioners to Drive the Evidence Train: Building the Next Generation of Evidence

“

“Evaluators need to see themselves as partners with direct-service providers. Many evaluators are interested in the science of program evaluation but lack the perspective of providers looking to improve impact. Sometimes they use unrealistic instruments or models.”

Nonprofit Leader  
Project Evident Market Survey  
(Fall 2016)

# Empowering Practitioners to Drive the Evidence Train: Building the Next Generation of Evidence

## Why We Need a Next Generation of Evidence

There is no doubt that the demand for effective social programs aimed at addressing persistent and pressing problems far outweighs the supply. The predominant paradigm for increasing the effectiveness of social programs follows a linear path that begins with testing and refining a practice model, shifts to summative evaluation to demonstrate and replicate impact, and finally works towards scaling the practice. While valuable, this framework – dubbed the “pipeline paradigm” by Knox, Hill, and Berlin<sup>1</sup> – assumes the development and scaling of social programs is linear: once programs are tested, refined, and demonstrated to work, the only question is how to get more people to use them. Many current efforts to promote the use and evaluation of evidence-based practices, including tiered evidence programs like the Investing in Innovation Fund and the Social Innovation Fund, as well as the goal structure of the Institute of Education Sciences grant programs, are rooted in this assumption.

The role of evaluation and evaluators in this paradigm is clear: they serve as independent auditors of the effectiveness of practice models, engaging with nonprofit organizations primarily to understand the practice model and determine how best to evaluate its causal impacts. Because the independence of an evaluation is highly valued, nonprofit organization leaders may provide input into the primary research questions being asked, but otherwise tend to be minimally involved. As a result, evaluation often feels like something done “to” a nonprofit organization – required by or for the purpose of pleasing a current or potential funder – rather than something done “with” the organization. And rarely does anyone ask whether these nonprofit leaders find the process of evaluation – or its results – beneficial to their work.

---

<sup>1</sup> Knox, G., Hill, C., & Berlin, G. (2018). Can Evidence-based Policy Ameliorate the Nation's Social Problems? Accessed 10/27/18 at [https://www.mdrc.org/sites/default/files/2018\\_Can\\_Evidence-Based\\_Policy\\_Ameliorate\\_Final.pdf](https://www.mdrc.org/sites/default/files/2018_Can_Evidence-Based_Policy_Ameliorate_Final.pdf)

The emphasis on keeping evaluators at an arm’s length from practitioners may be appropriate when an evaluation’s primary goal is “accountability.” In those instances, the notion is that an independent and rigorous evaluation can offer a judgment on the effectiveness of a practice model, and therefore inform decisions whether to invest in or “buy” the service. Too much engagement by or with the organization “selling” the service can bias the results.

But it isn’t clear that the greatest benefit of evaluation comes from using it for accountability. Accountability is important – especially for making high-stakes decisions about how to distribute public funding. But the effectiveness of efforts built on the pipeline model is still in question. Funding approaches built from this frame – from Pay for Success to tiered-evidence initiatives – have struggled with disappointing results, with very few examples of “evidence-based” initiatives resulting in definitive positive impacts on the targeted outcomes. It is too early to declare any of these funding models as ineffective, and the pipeline model has certainly raised the bar on the rigor of evaluation research. But experience to date suggests that this model is unlikely to lead to greater scaling of evidence to more populations – at least not in the near future.

The discouraging results of these pipeline approaches is especially concerning given the resources invested in them. The cost of large-scale, summative impact evaluation is high, particularly when considered alongside the lack of clarity it typically provides. Impact studies rarely yield answers as simple as “No, program X does not – and cannot – work.” More often, these evaluations show ambiguously small effects in some circumstances but not others, effects only when the program is well implemented, or effects that fade over time. Few programs have shown unambiguous benefits for participants, and additional research seems to continuously raise doubts about the validity of those studies.<sup>2</sup> Summative impact studies are important, as is increased rigor around evaluation research. But from the perspective of a practitioner, identifying what doesn’t work isn’t sufficient. Evidence needs to help practitioners figure out what *does* work – and most importantly, what *will* work – in their own program contexts.

---

<sup>2</sup> For example, see the Coalition for Evidence-based Policy’s “Straight talk on evidence” blog: <https://www.straighttalkonevidence.org/>

While practitioners need evidence to support a wide range of decisions about program implementation and optimization, the accountability approach to evidence building provides little guidance for those decisions. This most often leaves those nonprofits not in a position to conduct summative evaluations of their own out of the conversation altogether. These leaders receive no guidance on how to systematically assess and improve their programs or evaluate whether those improvements are working.

## **What a Next Generation Might Look Like**

Rather than viewing evaluation primarily as a tool for accountability, we believe it is time to empower nonprofit leaders to use evaluation and evaluation thinking to drive their own evidence agendas. Like staff or facilities, data are a strategic asset for nonprofits and should be a fundamental component of strategic planning. Evaluation methods are tools that nonprofit leaders can use to generate evidence for decision-making as a routine part of their practice.

To become a useful tool for nonprofit leaders, evaluation must be designed to inform the decisions that nonprofit leaders and practitioners need to make about program implementation, improvement, adaptation, and scale. In other words, evaluation should serve as the research and development component of nonprofit organizations, continually testing and improving the services delivered to achieve greater benefit for the clients served. The nonprofit leader drives these efforts, ensuring that R&D work is well-resourced with both staff and funding, and that R&D informs the strategic direction of the organization as a whole. To do this, nonprofit leaders and practitioners should partner with evaluators to develop organizational learning agendas – identifying what practices are tested, in what contexts and conditions, and with what methods.

Practitioners want digestible pieces of information that can help them build upon and improve their current practices, not evidence that tells them to throw out everything they are doing and replace it with a new model. Some researchers have discussed building evidence around grain-sized elements of practice, often referred to as “evidence-based kernels.”<sup>3</sup> Evidence at this

---

<sup>3</sup> For example, see Jones, S., Bailey, R., Brush, K., and Kahn, J. (2017). Kernels of practice for SEL: Low-cost, low burden strategies. Accessed 10/27/18: <https://www.wallacefoundation.org/knowledge-center/Documents/Kernels-of-Practice-for-SEL.pdf>

altitude may be more readily implemented since it can be introduced in a sequenced manner, tweaking and refining current practices rather than lurching repeatedly from one model to another over time. As one practitioner noted, “Many funders want to fund programs that are evidence-based or to fund the type of research that would lead to a third-party rigorous evaluation that ‘proves’ program impact. However, just as important is funding the steps leading to this type of rigorous evaluation work.”<sup>4</sup> If resources are also directed towards supporting this earlier stage practitioner-led work, there is an opportunity to build a larger pipeline of programs that produce meaningful outcomes.



“Researchers and their funders need to do a better job meeting the needs of practitioners and policy-makers.”

Robert Granger, Chair of Edna McConnell Clark Foundation Evaluation Advisory Committee

## The Role of Researchers and Evaluators in the Next Generation

Under this new paradigm, evaluators become partners with nonprofits in supporting program adaptation and improvements that increase program impact, rather than simply serving as independent auditors of overall program impact. Evaluators work side-by-side with practitioners to figure out what they need to know in order to better understand who they are serving, how well they are serving them, and what impact they are having. The questions practitioners need to answer can vary from basic information about how well they are reaching their target population through outreach activities, to how long their clients stay in their programs, to which modes of training staff or providing services are most effective.

The increased empowerment of nonprofit leaders in no way makes the role of the evaluator less important. As Dr. Ruth Neild, former Director of the Institute of Education Sciences, put it, “Evaluation is not a DIY endeavor.”<sup>5</sup> Nonprofits will rarely have the staff capacity or resources to run complex evaluation studies. Moreover, evaluators can provide a helpful external lens on a nonprofit’s services, processes, and theories of change, serving as a critical friend in helping the nonprofit question assumptions, identify challenges, and test solutions.

Note that for the evaluator, this new generation of evidence means a shift in *role*, not a shift in *method*. To date, too much of the discussion about evidence-based practice among evaluators has focused on in-fighting

---

<sup>4</sup> Survey data from the Innovation Network; Voices from nonprofit leaders from surveys conducted by Context Partners in early 2018 and Project Evident before its launch in the fall of 2016.

<sup>5</sup> Personal communication with one of the authors, February, 2016.

about methods. Concern about a lack of rigor in determining the causal effect of programs on outcomes led to an important movement to highlight the “gold standard” of experimental designs in impact evaluations. This movement drew attention to the fact that many evaluations use less than ideal designs for testing the effectiveness of programs, paying little attention to whether the findings of those studies capture the causal role of the intervention in affecting the outcome of interest. But this movement also led to a backlash from others, who noted that experiments are not always appropriate or feasible and that they, too, can suffer from limited rigor.

These evaluation debates lose sight of the real challenge in connecting research to practice and policy. Applied research and evaluation studies often focus on the impacts of fully developed and complex program models to inform decisions about adoption and funding, but most practitioners need far more basic information to support ongoing implementation and improvement, which is often not included in rigorous summative evaluations. It is not uncommon, in fielding impact evaluations, for evaluators to learn that nonprofit leaders lack good data for tracking enrollment or implementation. From a practical standpoint, it makes far more sense for program administrators to assess how well their existing practice aligns with research evidence than to throw out what they are currently doing and replace it with something new.

For evaluation to be meaningful for practitioners, *the decision that needs to be made takes priority over the method being used*, with rigor relating to the ability of the evaluation design to adequately address the decision maker’s questions. So, for instance, if a nonprofit leader wants to understand attendance patterns among their enrollees, rigor would relate to the measurement of attendance, the sampling of enrollees, and the ability to follow enrollees’ participation over time. In contrast, if the nonprofit leader wants to test different strategies for enrolling and retaining clients, a rigorous study would need to isolate the causal effect of those different strategies.

## **Defining “Rigor” in the Context of Continuous Improvement**

Rigor is important in all evidence building, whether its purpose is to inform public spending or to inform continuous improvement. In addition, rigor applies to many different aspects of evaluation design and implementation – measurement, sampling, causal inference – and the aspects of rigor that matter most vary depending upon the purpose of the study.

In the new generation being posited here, there is a shift from the nearly exclusive emphasis on causal inference in discussions about evidence-based practice. Causal inference still matters when testing the effects of a practice, or changes to a practice, on key outcomes. But the questions of nonprofit leaders will not always be ones of impact, and – as noted above – other features of their evidence may be more central to some of these questions.

In addition, the cumulative nature of evidence-building in this new framework means that there will likely be multiple “studies” informing the understanding of any given phenomenon, lowering the stakes of getting any single analysis wrong. So, for instance, the demand for rigor when testing a new practice in one community center will be less exacting than that needed to make that practice standard in all community centers nationwide.

Randomization – both randomized control trials and randomized tests of systematic variation – are equally relevant in the contexts of continuous improvement and accountability. There is nothing about continuous improvements that makes randomization irrelevant or impossible. But randomization in the context of continuous improvement is likely to test smaller units of change and systematic variation in practices, to provide more actionable information to practitioners.

---

## **Putting Practitioners in the Driver’s Seat of Evidence-Building: An Example from Project Evident**

PowerMyLearning partnered with Project Evident to develop a strategy for building evidence as it rolled out Family Playlists, an initiative where students teach parents or family members recently learned skills, who then





**“This (evidence building) work is core to who we are and who we want to be.”**

Elisabeth Stock, PML

provide feedback on the lesson to the student’s teacher. These interactive homework assignments leverage family engagement to strengthen the school-family-student triangle that is critical to boosting student achievement. PowerMyLearning adapted Family Playlists from a model with rigorous evidence of effectiveness and had conducted a case study in a single school, but was eager to develop an evaluation strategy tied directly to the improvement, adaptation, and scale of the initiative.

Project Evident staff worked with PowerMyLearning staff to develop a Strategic Evidence Plan that outlined a learning agenda mapped directly onto the theory of change behind Family Playlists. PowerMyLearning articulated evaluation questions at every stage of the theory of change, from whether and how often teachers assigned playlists to their students, to how well the family-student interactions went during the assignments, to impacts of the model on student social-emotional and academic outcomes. The plan embeds rigorous measurement and evaluation methods in the learning at all stages, so that PowerMyLearning benefits from causal evidence for decisions about implementation in the same way that external stakeholders will benefit from causal evidence of impacts on student outcomes.

With support from Project Evident staff, PowerMyLearning is launching an experimental study on the impacts of different approaches to supporting teachers to assign more playlists. PowerMyLearning staff are randomizing teachers to receive different supports to determine those that best promote assigning multiple playlists. Successful completion of this study will serve as a model for how PowerMyLearning will answer a host of questions it has about how to best implement the Family Playlists program and will be a source of continuous learning for the organization.

Even as PowerMyLearning has increased capacity to engage in generating its own rigorous evidence, it understands the value of partnering with evaluation experts as its learning agenda evolves to include questions about impacts on students. Its Strategic Evidence Plan includes a strategy for continually evolving the learning agenda and for assessing when to turn to evaluation experts for evidence generation. PowerMyLearning remains in the driver’s seat, and plans to bring the best evaluation resources to bear when addressing its learning agenda.



## What Will it Take to Shift the Paradigm?

Many applied social scientists are passionate about their work benefiting society, and may be enthusiastic about playing more of a partnership role with nonprofits. This commitment already drives many researchers and evaluators to partner with nonprofits and governments, and to offer input and advice to practitioners on an ad hoc basis at little to no cost. There are likely many researchers who would be willing to rethink the way they work with programs and practitioners to better support program improvement. Others have always worked with practitioners as partners.

At the same time, the incentive structures for researchers do not always support increased engagement with practitioners. Academic researchers are rewarded for developing a research agenda, receiving prestigious research grants to support that agenda, and publishing work in peer-reviewed journals. Partnering with nonprofits in the way described takes more time and requires researchers to give up some control over the research agenda or evaluation that is conducted. It is also likely to include more mundane analyses, such as descriptive data on program participants, that are critical but may have little to no value to academic audiences. Likewise, the incentives for researchers in evaluation firms promote involvement in longer-term, larger-scale research studies that apply rigorous methods, not necessarily the types of work that benefit practitioners.

In addition to misalignment with incentive structures, this new approach to research and evaluation may require researchers to develop new skills for working with nonprofit organizations and practitioners. To give practitioners a meaningful voice in developing a learning agenda, researchers need to listen to and understand practitioners needs. This may be challenging for some, since few researchers have played a senior role in a nonprofit context. Moreover, certain behaviors that are tolerated and even rewarded in scientific communities may cause friction and reduce trust between researchers and practitioners. For instance, most researchers are taught to identify logical fallacies and to openly criticize each other's products in order to increase accuracy and quality. These behaviors can be off-putting to non-researchers and may create barriers to practitioners taking ownership of their learning agenda. Instead, they may look at the evaluator as the "expert" to whom they must defer.

As always, funding can be a major driver of behavior. If government, philanthropy, and other private funders provide resources for practitioners and evaluators to collaborate on a learning agenda for program improvement, support increased data capacity within programs, and provide clear signals to evaluators that they expect and reward partnership, the evidence-building field will shift.

Of course, the field will always need independent summative impact evaluations. But in the next generation of evidence, those evaluations would be one part of a longer-term and more strategic learning agenda oriented toward improving a nonprofit's capacity to drive impact. Summative evaluation would stand alongside rigorous, more rapid tests of program improvements, better descriptive analysis, and any number of other tests to improve program implementation and design.



**"It has been amazing to watch even strong, long-established organizations learn and improve as they work to craft a strategic evidence plan."**

**Dr. Rebecca Maynard,  
Professor of Education and  
Social Policy Emeritus at the  
University of Pennsylvania  
Graduate School of  
Education**

## Summary

The movement toward greater use of data and evidence in social programs is encouraging, and there is a widely shared commitment from evaluators, policymakers, practitioners, and philanthropists to getting better outcomes for the vulnerable segments of society served by nonprofits. At the same time, the current paradigm for using evaluation to get those outcomes has faced challenges. It is time to revisit that paradigm and flip evidence building on its head, engaging nonprofit leaders in developing learning agendas that support continuous improvement toward better outcomes for their constituents.

This paradigm shift will alter the role of evaluators in generating evidence: instead of serving as independent auditors, they become partners in program improvement, working alongside the nonprofit leaders. For their part, nonprofit leaders get to be a conductor on the evidence train, developing learning agendas that will help them answer their most immediate questions about improving the implementation of their programs.

This new paradigm requires evaluators to complement their methodological skills with stronger relational skills. Evaluators must learn how to work with practitioners as partners, understanding their needs and demands and valuing the experience and perspective they bring to the work. Of course, evaluators are influenced most by the incentives they face

– from funders, from their institutions and from their peers. Thus, funders and research institutions (academic and non-academic) must consider how their current incentive structures reward – or discourage – behaviors required for evaluators to partner with nonprofits around continuous improvement.

This change is possible; not the least because many applied social scientists have a strong commitment to having their research make a difference for society. For younger generations of evaluators, advances in technology and big data have made data-driven decision making a key component of everyday life. This generation may be particularly attracted to the potential for rapid innovation in measurement and learning that comes from partnership with nonprofits. Efforts like Project Evident are harnessing this passion, helping to build partnerships between researchers and practitioners in order to develop and implement strategic learning agendas.

"It has been amazing to watch even strong, long-established organizations learn and improve as they work to craft a strategic evidence plan," said Dr. Rebecca Maynard, Professor of Education and Social Policy Emeritus at the University of Pennsylvania's Graduate School of Education. "Most 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." Together, we can expand on successes and lessons learned in the field of evidence building, and foster a new generation of evaluation.