

PROGRAM EVALUATION

Promises and Pitfalls

Ryan P. Kilmer and James R. Cook

Evaluation can have many roles and purposes and can yield meaningful benefits. In Chapter 1, we outlined some basic parameters for evaluation (and what constitutes evaluation) and discussed some foundational content for thinking about this work, including some common questions addressed via evaluation. As we have underscored, for us, it is critical to use evaluation not just “to know” (i.e., documenting a targeted outcome or an unintended impact) but “to do” (i.e., effecting change, such as improving a program or intervention).

In this chapter, we will consider some of those roles and benefits in more depth. However, we would be doing a disservice to you, our reader, if we do not also engage some of the very real challenges faced by evaluators, ones that reduce the likelihood that a sound evaluation will be implemented or supported—or that impact the degree to which the evaluation and its findings even get used. We will start off by considering briefly some of the important roles and benefits of evaluation, including the rationale for conducting evaluation.

ROLES AND BENEFITS OF EVALUATION: WHAT IS THE RATIONALE FOR CONDUCTING EVALUATIONS?

This section outlines some of the major reasons for conducting evaluation—what it can “buy” an organization or program, how it can be used, and what its benefits might include. Throughout, we will discuss brief examples to help illustrate our key points.

Purpose of evaluation: Assessing program impact. Did it work? In recent years, there has been an increasing emphasis on “accountability” for public sector systems (such as our schools, child welfare agencies, and the juvenile justice system) as well as nonprofit organizations. That is, their funders and their stakeholders (boards of directors, community members, supporters, staff, and the intended

Learning Objectives

1. Learn the role(s) and benefit(s) of, and rationale(s) for, program evaluation.
2. Describe approaches, ideas, and alternatives to evaluation that reduce the likelihood that sound evaluation will occur.
3. Understand factors and conditions that can impact the degree to which evaluation data are used.

beneficiaries of programming or services) want to know if the program or intervention “worked”; did it yield the impact, effects, or benefits that were desired or intended? Schorr (1997, p. 117) describes how the “public wants proof of results” and quotes Alice Rivlin (1971): “Effective functioning of the system depends on measures of achievement ... To do better, we must have a way of distinguishing better from worse.” This “did it work?” question is addressed via *outcome evaluation*, and this objective likely reflects what most people think of when they consider “evaluation.” However, we need to not only think about those intended (positive) results or outcomes, but also consider and assess possible unintended, negative consequences, known as *iatrogenic effects*. That said, it is difficult at best and impossible at worst to know if a program, intervention, or initiative “worked” without truly knowing what it involved and what they did in actual practice.

For example, the national evaluation of systems of care, also known as the National Longitudinal Study, examined a host of outcomes associated with efforts designed to revamp the system of services and supports provided to children with severe mental health difficulties and their families (e.g., Center for Mental Health Services, 2003, 2004; Holden et al., 2003). Over a period of years, various annual reports summarized results indicating a modest to moderate level of success, such as reduced need for special education, fewer suspensions from school, and decreased involvement in the juvenile justice system (e.g., Center for Mental Health Services, 2003, 2004; Cook & Kilmer, 2004; Holden et al., 2003). However, this outcome-focused work did not account for the variability across sites, from differential interpretation of key principles to very substantial differences in implementation (Cook & Kilmer, 2004, 2012; Kilmer & Cook, 2012). Funded communities were required to take part in this national evaluation, but they were not required to assess their *implementation fidelity*, including the degree to which they were putting key principles for the philosophy into actual practice (Kilmer & Cook, 2012; Kilmer et al., 2010). Data reported via various mechanisms, including reports to the U.S. Congress, would effectively summarize diverse outcome data—and, on a national scale, the outcomes from sites with strong implementation would be combined with those with middling to poor implementation, with the overall results suggesting modest to moderate positive changes. Site visitors conducting system-level assessments (Brashears et al., 2012) noted uneven implementation across key principles for these federally funded systems as well as meaningful variability in sites’ quality improvement processes; however, that information was not integrated in the evaluation products describing outcomes nationally. For instance, it was not possible to examine effects separately for high- versus low-fidelity communities.

In our own community, we conducted a multicomponent evaluation in which we focused on what care coordinators and planning teams were doing (particularly in regards to key elements of the practice model they were to be implementing) and what children and families received and experienced through the system, as well as the required outcome-focused work. We found that, when youth were served via planning teams that engaged in high levels of implementation of the identified model, they improved more and at a greater rate than those whose planning teams did not (e.g., Haber et al., 2010; Hemphill et al., 2010; Johnson et al., 2011). Without taking into account what sites did—specifically, the

We cannot say “it worked” if we do not really know what “it” was. Process data are crucial for contextualizing outcomes. Without such information, the understanding or meaning we can make from the outcome data is more limited.

degree to which their practices and processes aligned with the system of care philosophy and key practice models—it is impossible to truly judge the effectiveness of the approach or how well the program initiative “worked,” since there were different versions of the approach being implemented in different locations.

We learn such information via a *process evaluation*, which focuses on monitoring the program and its implementation, assessing if it is being conducted as planned or in a manner that is consistent with a particular model. This degree to which a program evidences faithful replication or alignment to an intended model or program (e.g., Durlak & DuPre, 2008) reflects *implementation fidelity*. Quite simply, we need to know what the program or initiative is doing, and how those actions occur, in order to try to gauge its outcomes; without the process assessment, we may falsely conclude that the program as planned had particular effects, when we are really evaluating what may be a very different program (the version that was actually implemented). This brings us to our next major role for evaluation.

Purpose of evaluation: Understanding outcomes. Do we know why it worked or did not work? Process-oriented information is critical for contextualizing outcome-related data—and helping to make sense of findings. If outcome data do not reflect the positive outcomes for which program leadership hoped, we need process data to understand why. Was it a problem of program theory? That is, were our ideas regarding how the program’s actions would lead to the desired change (the theory of change) incorrect? Alternatively, was it a problem of implementation? That is, while a model may have evidence indicating its effectiveness, did our program staff not (or inconsistently) implement it with fidelity? Such linkages between process and outcome data will be discussed in detail in Chapter 6, but we will address these challenges briefly here.

As others have noted (e.g., Kloos et al., 2012), issues and errors in program theory led to some prominent programs (such as the Drug Abuse Resistance Education, or DARE, substance abuse prevention program) not having their intended effects. Another frequently cited example is Scared Straight. This program, intended to deter youth from engaging in delinquent or criminal behaviors, involved bringing groups of youth to prisons and having them see first-hand the consequences of these negative behaviors (e.g., Gilna, 2016). The program approach included incarcerated individuals relaying their stories, often including grim details, while yelling at or confronting the youth. The program was shaped by the idea that seeing this potential reality, and having the incarcerated individuals yell threats at the youth and try to frighten them, would shake the youth out of their rule-breaking ways (e.g., Gilna, 2016). This notion was not supported by research or evaluation; the program theory was incorrect. In fact, multiple studies found that program participants actually appeared to be at increased probability for committing crimes, and evaluation found that the program increased participants’ risk for arrest, by unintentionally reinforcing attitudes and behaviors associated with breaking the law (e.g., Petrosino et al., 2005; 2013; also see Kloos et al., 2012). This is a clear example of an iatrogenic effect.

In other instances, the program theory is sound, and there may even be evidence to support its effectiveness, but it does not seem to “work” or yield the intended effects in a given setting. The issue in such cases may reflect problems of implementation—more specifically, the program’s implementation may not be of the quality or dosage of the original or it may not align adequately with a model that has documented desired

effects (e.g., Durlak & DuPre, 2008). We have seen many instances of how low-fidelity implementation—or not implementing the program or intervention in a manner that is consistent with the evidence-supported model—does not yield the desired effects. Issues of implementation have been discussed at length by many authors (see, e.g., Durlak & DuPre, 2008; Fixsen et al., 2005). In their review of five meta-analyses and 59 additional studies, Durlak and DuPre (2008) concluded that findings strongly supported the notion that “effective implementation is associated with better outcomes.” Going further, these authors wrote:

A major implication emanating from these findings is that the assessment of implementation is an absolute necessity in program evaluations. Evaluations that lack carefully collected information on implementation are flawed and incomplete. Without data on implementation, research cannot document precisely what program was conducted, or how outcome data should be interpreted.

Purpose of evaluation: Comparing models and assisting with model selection. When high-quality process and outcome data are collected, it is possible to compare the outcomes obtained via one program or intervention model versus another and inform the selection of a program model to utilize going forward. This could include comparing curricular models for a school-based program, deciding upon a framework for a community-based mentoring program, or assessing different home visiting approaches for families. The bottom line is that a careful analysis of the characteristics of programs as implemented can enable evaluation to support data-guided choices.

Purpose of evaluation: Informing changes. Should we do something differently? Providing information to improve programming or inform decision making about the program or initiative is a prime role of evaluation. This interest in using evaluation to improve programs is closely related to the objective of understanding outcomes; process data are absolutely necessary for guiding any kind of improvement or program refinement. Data can shed light on areas of strength and those that appear to carry weight in contributing to desired outcomes, as well as elements that warrant additional training or professional development, program components that do not “add value” in their current form, or gaps in a program model.

Sometimes these changes can be quite straightforward. As part of one evaluation, we provided early process feedback about how the pilot of a particular mental health, cross-system practice model was being implemented. We pointed to several specific areas in which implementation was going well and less well. After the meeting during which we shared these “early returns” with administrators and staff involved in the particular initiative, we had staff approach us and quietly disclose that they did not know what a particular intended program element was (we had noted the regular absence of this element in our discussion with the group). This suggested a need for additional training as well as supervision specific to the key practice model components. Without the process data—and a mechanism for providing feedback—this gap in implementation would have likely persisted. In this particular instance, the finding pointed directly to the need for resources dedicated to professional development and support.

In other instances, the circumstances are not as clear-cut or readily addressed. In another evaluation (in fact, in multiple related evaluations over multiple years), the data indicated that practices were not reflecting the cross-agency and cross-system collaboration thought to be crucial for the effort's success. Our data helped shape some strategies for addressing these issues, and some partners were interested in and willing to modify their actions. Unfortunately, all of the players did not share that sentiment, despite our use of diverse strategies to engage key leadership and staff, demonstrate how we could help support needed changes, and share a subset of results illustrating the potential benefits of the work when this collaboration occurred and the broader model was implemented effectively. These issues, in short, led to initiatives that underperformed and did not realize their potential—they did not have the intended benefits for the children and families served.

Purpose of evaluation: Informing resource allocation. Should we continue to fund it? How should we use the resources we have? Attesting to a program's effectiveness or impact can aid in deliberations about the use of public resources or the resources of a nonprofit or other organization. This is a central emphasis of much evaluation work and can relate directly to the degree to which a program, intervention, or initiative receives continuing funding, or funding to support its expansion. As we note in Chapter 3, we co-directed an evaluation of an early childhood education program that had sustained dramatic cuts to its funding shortly before we were engaged in work with the school system. The program and school system were hamstrung by the fact that they did not have rigorous data to support the program's effectiveness. During a time of resource strain and needed budget cuts, that lack of data did not serve the program well, and the program was targeted for major cuts. More broadly, the presence of evaluation data, with substantive information about program functioning and impact, can aid in deliberations about the use of public resources.

Such resource decisions are not limited to whether a program should be funded (or expanded) or not. Rather, evaluation data can be used to inform more fine-grained decisions, including about the use of professional development or training funds, the need for additional person power, the benefits of enhancing or augmenting a particular program element, and other decisions regarding program components. Evaluation can assess the "value add" of specific program components—or of innovations—and findings can also support the need for resources dedicated to enhance the implementation of those program elements. This may include additional support for consultants, coordinators, or coaches to help oversee or provide feedback regarding such elements as team planning meetings, group activities, program or classroom practices, or the informational support provided to parents. Alternatively, evaluation can help identify program components that are not adding to the overall value of the program, and may thus be either strengthened or eliminated. Chapter 12 provides an in-depth description of cost effectiveness and cost-benefit analyses.

Purpose of evaluation: Support grant applications. In a similar vein, evaluation data—even a robust evaluation plan—can help an organization, program, or initiative to be competitive for grant funding. Rigorous evaluation data are viewed positively by funding entities; preliminary data can help demonstrate promise; and a well-designed evaluation plan can convey the value placed on data-guided decision making, accountability, and a culture of learning and improvement. Years ago, the leadership of a small local partner that provided

family support services shared with us that our multiyear relationship (e.g., Cook et al., 2008; Kilmer et al., 2009), including our collaborative effort to develop a logic model, modify their data collection strategies, and build their evaluation capacity, was explicitly identified by a local funder as a key element in their positive funding decision. By her account, the funder saw those steps and the local partner's ongoing relationship with us as characteristics that helped set them apart in a competitive funding field.

Unfortunately, many applications submitted for program or initiative funding tend to frame evaluation as an afterthought, with underdeveloped plans or underfunded evaluations. In an optimal circumstance, the plan for evaluation would be developed hand-in-hand with the program's plan. Doing so—with an understanding of program objectives and how specific components are designed or intended to support program goals—would ensure targeted collection of process and outcome data and, more generally, access to the information necessary to answer questions of interest to the program or initiative leadership as well as key stakeholders or partners. The development of a rigorous evaluation plan can help establish trust with stakeholders, demonstrating an investment in quality programming, thorough understanding, and fiscal stewardship. It can also help improve the program planning from the outset, by encouraging program planners to think in some detail about the ways they would detect changes, and the theory of change that would be expected to lead to those changes.

WHAT ARE BARRIERS TO EVALUATION? WHAT ABOUT WHEN THIS DOES NOT GO WELL?

Up to this point, we have described the manifold benefits of evaluation. Most critically for us, evaluation is necessary to improve, and optimize, the programs, services, and supports provided to individuals, children, or families who are their intended beneficiaries. In addition, the groundwork for the evaluation should be linked inextricably to the development, planning, and implementation of the program. Given the clear benefits of evaluation, it would seem as if every program, initiative, or organization should be doing this, right? Unfortunately, however, “real” rarely aligns with “ideal,” and many agencies, organizations, and programs do not have the capacity to support well-designed evaluation; they dedicate insufficient resources to support evaluation; or they operate as if evaluation is an “add on” or a “have to” that allows them to meet some reporting need (as if they are just checking a box). This section considers some factors and dynamics at play when strong evaluations are not sought or used—or their data are discounted.

A focus on outputs. In the place of an evaluation, many organizations track and report lists of the programs they have run, the number of people they have served, the presentations given, and other activities. We see this reliance on outputs or activities with some regularity—it often reflects the lack of dedicated resources to, or capacity for, evaluation. We have seen groups or organizations report that they conducted x number of programs in the community (sometimes including the number of people in attendance); disseminated x thousands of pamphlets or informational brochures; coordinated x number of program/support groups; or provided their program or intervention to x number of people. However, simply reporting such outputs does not provide any information regarding their effectiveness or impact. For example, we do not know if or how someone used the

information provided, if they followed through on the referral made after the health risk screening, or if they attended to the discussion and used the material from the coping skills group. Based on these outputs alone, we cannot gauge if our actions constituted an adequate dose of the program or intervention, if we had the intended effects (or if there were unintended effects), or if the program was delivered with fidelity. We hope it is clear that this approach to “program evaluation,” which would be more accurately referred to as program monitoring, is extremely limited. Tracking and reporting this information only permits the conclusion that the program or initiative did *something*, not that it made a difference.

The weight of political forces or reputation. Sometimes factors seemingly unrelated to the project or program come into play and influence the likelihood that evaluation findings will contribute to program improvements or that an evaluation will even occur. In some cases, when an effort or program has a powerful, well-known, or well-liked champion, founder, or director, that individual’s political weight, popularity, or reputation can override evaluation findings or even lead to a conclusion that there is no real need for an evaluation. Sometimes systems or funders continue to fund efforts without requiring a critical evaluation of the work, believing that the work must be of value because it is associated with a particular champion or program director, even when there is no clear evidence that the program is having the desired effects. In other cases, programs are viewed positively because the work reflects values or goals that align with prominent or popular views (or, in some cases, social policy), or because the program is viewed as an “evidence-based program.” However, being guided by admirable values and goals, having a well-known champion or leader, and using a program model found to be successful at another place and time do not necessarily translate into program success; evaluation is necessary to gauge the effort’s effectiveness and identify strategies for its ongoing refinement.

We have encountered this issue in different forms. In one example—we will be intentionally vague here—we were brought in to evaluate an initiative that included a diverse range of stakeholders. We were invited to join the leadership and planning team, and we worked with the various players to understand the facets of the effort. We sought to partner and collaborate, but the nature of the relationships that emerged (and the collaboration that followed) varied substantially among the stakeholder groups and their organizations. Early on, the initiative hit some bumps and the road was rocky—the nature of the work was new and different for the vast majority of the individuals being asked to implement the intervention and, notably, they regularly seemed to face challenges to their participation in this effort from their own agencies. We worked to provide balanced process-related information as well as specific, data-guided recommendations, with the goal of informing changes and shedding light on difficult areas that warranted additional attention or resources.

The initiative’s leader had been active in the community for some time, was well connected, and, overall, had a positive reputation. Our experiences with this initiative leader during this specific effort led to a much different impression. Over time, as it became clear that there were major discrepancies between the actual initiative as implemented and the public description of the program, the initiative leader began to create roadblocks for the evaluation, from shifting established communication responsibilities and procedures to limiting access to needed data. She questioned every element of the evaluation and attempted to sow doubt about its methods, despite having been part of the discussions and

processes through which the evaluation plan was developed. This individual met directly with the prime funder regularly and made sure the initiative had ongoing support; in that context (and others), we learned that she also made unfounded statements about us, our practices, and the evaluation. In the end, we tried multiple strategies to foster collaboration, to strengthen the broader effort via direct work with the partners who were interested, and to convey the initiative's challenges, the issues we were encountering, and some specific ways in which they could respond. Unfortunately, these attempts never gained sufficient traction to help the initiative turn a corner or to overcome the efforts by the initiative's leader to undermine the evaluation. Our final report documented the numerous and assorted challenges to the evaluation and drew on the data available to provide recommendations for modified practice. We also reported on outcomes—given that the practices did not align with the plan or the identified model for care, it is not surprising that the indicators of outcome suggested no benefit (and, in some cases, deterioration) for the children and families involved. Our contract was not renewed (much to our relief), and they identified and worked with multiple different evaluators over the years. The initiative's "champion" was powerful and well thought of enough that she was able to secure funding to sustain the initiative. Even now, years later, it has not achieved its potential. We will say more about this evaluation experience in the section below discussing "Organizational Factors" for why evaluation data do not get used.

Personal experience and testimonials. Many program staff and intervention providers will share stories regarding program successes and specific positive outcomes. When asked, they frame evaluation as unnecessary, given their experiences doing the work day to day. In fact, we have had some program staff or leadership literally say: "We just know it works." While some anecdotes of individual and family success can surely be compelling, and such testimonials are regularly used and can be helpful in fundraising, it is hard to know how representative they may be of program participants' experiences. In addition, we do not know what other resources, services, and supports may have contributed to the positive outcome or change in circumstance the staff witnessed. In brief: these experiences and stories do not take the place of an evaluation.

WHY DON'T EVALUATION DATA GET USED?

Sometimes an evaluation has been conducted, but the data or findings do not get used. A diverse range of factors can contribute to that outcome, some attributable to evaluators, some to organizations, and others to a variety of external influences.

Evaluator-related factors. There are times when the evaluator or evaluation team is responsible for their work not being used. Sometimes evaluation results are not provided in a timely manner, reducing their potential impact, because they are delivered outside a useful time window, or other program or contextual changes have made them less representative of the program's current function. While we were not involved in the effort, we know of one such evaluation (we will be intentionally cryptic here): Despite describing the work and its design in various forums and settings, the evaluator missed deadlines for the project's report and other deliverables and did not provide findings to the program leadership in time to inform some key decisions about the program's future. In the end, very significant decisions were made, without empirical grounding, and the program and its

delivery were changed meaningfully. The evaluator's lack of follow-through had repercussions for the program (and its intended beneficiaries) as well as the evaluator's reputation in the community.

In other instances, evaluator-related issues reflect challenges with communication. For example, the evaluator(s) may not have adequately conveyed the relevance of the findings to decision makers or other stakeholders (Kilmer, 2015). This can occur when reports or presentations are overly technical, they focus too heavily on statistics and effects, and their descriptions are jargon-heavy. Evaluators are better served when they can develop evaluation products that are tailored to the background and understanding of their audiences. Put frankly, evaluation findings will not be used if they are not communicated in a user-friendly manner that speaks to the concerns of the stakeholders (Kilmer, 2015). The technical elements (and reporting appropriate statistical approaches and their results) are important to be sure; it is also necessary to ensure that we as evaluators communicate "the story" of the evaluation and help readers understand the meaning of the results and their implications. In turn, in addition to the full evaluation report (perhaps with its most technical aspects in an appendix), we virtually always find it helpful to develop an executive summary or, in some instances, a one- or two-page brief of the evaluation's key findings and recommendations, to help ensure that the main points are more likely to be digested by critical stakeholders.

Organizational factors. Sometimes organizational leadership keeps evaluation findings at an arm's length because the results suggest that changes are needed. This behavior may reflect a fear of or discomfort with change or, more simply, organizational inertia. That is, it is easier for many in leadership just to keep programs and their procedures in place, to follow the path of least resistance and do nothing differently, regardless of the issues identified via evaluation. Some may say, regardless of findings, that the results confirm their expectations or "what we already know"; others, when learning of recommendations, say "that is what we already do." We have faced such situations, and in these circumstances it is difficult to gain momentum in working for change; we often find that the leadership did not seek evaluation of its own volition, but instead was strongly encouraged to do so by funders or others. In other instances, program leadership and staff may worry about negative findings and program problems coming to light, which may have the potential to impact negatively their funding or fundraising. Alternatively, the leadership may be concerned about the effects of any negative findings on the work or on staff morale. This aversion to any risk can undermine efforts to learn about the program and identify strategies for improvement.

Unfortunately, another phenomenon that we have encountered in the context of more than one evaluation is the desire to "look good" rather than actually "be good." This has occurred as part of higher profile efforts, sometimes with significant funding behind them, that have been undertaken with considerable fanfare. We will remain vague here for reasons we hope are evident. In the two most prominent examples of this issue that we have faced, the data, collected via multiple sources and using multiple methods, suggested that, at the core, those involved with the effort were not doing what they planned or said; the reality of implementation did not align at all with the model or plan. It likely goes without saying that outcomes mirrored those issues of process; that is, the obtained outcomes did not approximate the desired or hoped for outcomes nor those reported by similar initiatives. We provided process-related data for each of these initiatives, regularly reporting

areas that could be addressed to help “right the ship.” Despite our efforts to be balanced, to report areas of strength and those processes that were being implemented with fidelity as well as areas that needed attention, leadership for each initiative pushed back in multiple ways. It is not surprising, in such circumstances, to have program managers criticize the data, argue that the results are not representative of the program as a whole, assert that steps have already been taken to improve the processes in question, or say that those findings do not really matter anyway. We have learned that it is useful to have a fairly thick skin in evaluation work and to make sure that the evaluation data and processes are sound, to help counter such arguments.

This brings us back to the example we noted earlier in which the well-connected and well-liked initiative leader took multiple steps to impede the evaluation. In view of the issues at play, that experience warrants mention here. In fact, in a particularly striking incident, in a meeting preceding the release of an anticipated interim report about one of the efforts, we were asked “not to include” any content that could be perceived as “negative or that would raise questions” (the question was not asked by the initiative leader but by one of that individual’s colleagues, and the leader was present). This request cuts to the core of an evaluator’s professional ethics and integrity (see Chapter 5), and we conveyed directly that the data were the data and they “tell the story.” We emphasized that, at that time, some indicators would be viewed as encouraging and others would be seen as pointing to clear issues, underscoring that we needed to gauge the quality or representativeness of the data, the degree to which caveats and qualifiers would be necessary (or if it was “too early” to report a given finding), and the like—but that those decisions would be our call. Given the context, it was necessary to assert ourselves and draw some clear lines regarding the integrity of the process. In that example, over time, the attempted partnership deteriorated; the initiative’s leader chose to question the data and our methods and impugn our professionalism, rather than engage the findings thoughtfully or consider how to implement some of the strategies that we (and others) had communicated. Our contract was not renewed after the 1-year pilot, and the available data indicated that the initiative did not benefit.

This specific challenge, that is, when an organizational or initiative leader seems to care more about perception than about doing right by the work’s intended beneficiaries, constitutes the most significant challenge that we face in our evaluations. In these instances, the leaders have not responded to clear and open communication nor to data- or logic-based rationales for needed modifications. Instead of framing the results as pointing to opportunities for change and improvement, they dug in and talked of the positive work being done. Of particular note, when this has occurred, the individuals heading the initiatives have evidenced a moral compass and sense of professional integrity that vary dramatically from our own. They have woven falsehoods about their own work as well as our evaluations (and us) and, in the end, they have done a disservice to the initiative’s potential beneficiaries. These two efforts were opportunities lost.

External factors. Non-data-related factors can keep a program alive or hurt it and have substantive implications for evaluations. There are different perspectives, agendas, and stakes at play for the varied parties involved in the evaluation of a program or initiative. It can, at times, be challenging to navigate waters that may be experienced as politically treacherous. These waters can reflect resource strains or needed budget cuts as well as shifting policy interests and needs, especially during difficult budget years.

Shifting policy interests can sometimes reflect what some of our colleagues have deemed the “shiny penny” phenomenon, in which a new (or sometimes revisited) topic or initiative can capture the attention (and dollars) of local leaders or funders. In the short term, this can lead to interest in and support for evaluation of efforts or programs that function in this new area. However, the other side of that coin is that such shifting priorities can also lead directly to funds being diverted from other evaluation or program-enhancement efforts. Those functioning in areas that are no longer high priority can see their evaluation findings or recommendations fail to garner interest or, in some scenarios, can even experience the loss of expected funding. As we note in Chapter 3, we encountered this in our work with our local school system’s early childhood program. In brief, even though the work had clearly been conceptualized as multiyear, system representatives conveyed in the latter part of year one that there would not be system resources to support the ongoing evaluation. This required the partners to seek other funds and led to a disruption of our work (also see Tynan et al., 2015).

WHAT YOU CAN DO

As we will discuss in the next chapter, there are steps evaluators can take to increase the likelihood that evaluation results will be used and have an impact, ideally by informing improvements in the program or initiative. A key is working with program leadership and staff, determining jointly the questions that will guide the evaluation.

Adopt a partnership-based approach to evaluation. This approach (see Chapter 3 for a detailed discussion) is foundational to this text. Creating an evaluation context centered on collaboration and learning (as opposed to “outsiders” coming in to judge, offer critical or punitive feedback, or put a program at risk) is crucial. Doing so can facilitate buy-in and participation across an organization, program, or initiative, which can yield better access to and quality of data. Working in partnership also increases the probability that the evaluation is experienced less as a “have to” or something “done to” the program and its staff, and more of an opportunity for us to learn about and maximize the impact of the program, thereby enhancing the likelihood that the evaluation (and its findings and recommendations) will be used.

Discuss concerns openly. Evaluators can set the stage for partnership by fostering open and direct communication—this is necessary for all relationships but holds particular salience when developing new ones. We can raise the possibility of and ask about risks or costs that may be eliciting concern. For instance, program leadership and staff may have concern about the evaluation’s findings or the results’ potential impact on their bottom line. It can be useful to name the issues explicitly and then have the opportunity to address them head on. Sometimes it can be useful to acknowledge and appreciate the very real risks about which a program or organizational head may be concerned. Some concerns can also readily be reframed. For example, negative findings are a concern for potential funders, board members, staff, and other stakeholders if the organization chooses not to respond; however, when an organization recognizes that results do not align with the intended effects and demonstrates that they are seeking to improve the quality of their program, that focus on improvement can actually be of benefit.

Explore—and demonstrate—possible benefits. We recommend, early on, inquiring about questions of interest to program or initiative leadership. What do they want to know (but have not had the capacity to answer themselves)? How could the evaluation yield useful information that would help them improve their program? (See Chapter 1 for questions that programs might consider addressing in an evaluation.) When possible, it can also be illuminating to hear from program staff, ask about their questions, and learn their observations. In such discussions, it is helpful to underscore the importance of the evaluation for the intended beneficiaries of the effort—we try to emphasize the goal of optimizing the potential benefits for that targeted population (e.g., we want to maximize the strength and quality of the program for all of the enrolled children). Staff in social programs are generally drawn to that work because of their commitment to improving the well-being of the children, families, or other intended beneficiaries of the program. As such, they are often very receptive to evaluation efforts that can lead to improvements in their ability to benefit their clientele.

In addition, early on in an evaluation, there is utility in providing brief summaries of findings, in digestible formats. Compelling visuals or preliminary findings regarding possible trends can help those associated with the program have a more clear sense of what evaluation results could look like and how the data may be useful to them. For instance, summaries of the needs and challenges faced by intended program beneficiaries can help an organization's leadership and staff to define the scope of the problem(s) they are trying to address, and this information can be used in their grant applications and fundraising materials. It can also lead to program staff identifying additional questions that the data might answer, further investing the program staff and leadership in the evaluation efforts. A key is using early data to demonstrate the potential yield and usefulness, the promise, of the evaluation.

We have been able to see the fruits of demonstrating such promise or helping a program toward “small wins” repeatedly over the years. As one example, we have partnered (in various roles) for over a decade with a local agency that runs multiple different programs in our community. The director of the program that had the least information about its function and impact, and the least capacity to track and report data, consistently shared that she and her staff were not interested in an evaluation; they did not need it, they knew the program worked, and they did not want to subject their staff or program participants to evaluation processes. Over a period of years, our doctoral students (under our supervision) worked with the agency's other two programs, helping them develop or clarify logic models (see Chapter 1) and build capacity for tracking and reporting data that would be of interest to them. Agency leadership and the program directors shared their enthusiasm for their changes and the newfound capacity. In one meeting, with multiple diverse stakeholders present, the agency's executive director shared how, because of one of our student's efforts with one program and the new data to which this executive director now had access, he was making some changes in how he was allocating resources to the program. The data shed important new light on some person power needs and how they could make a difference. In fact, after this work with the agency's other two programs, the director of the third program evidenced a shift in her perspective—she became very interested in having one of our students work with her program, enhancing their capacity to track, use, and report data.

CLOSING THOUGHTS

While there are many challenges in doing community-based evaluation work, those challenges also yield opportunities for fruitful partnerships—and the chance to make a difference by facilitating needed change for social good. Evaluation offers some clear promise for having this kind of impact; however, all potential partners do not necessarily share this view, and all partnerships are not necessarily successful.

In this chapter, we discuss some of the common pitfalls that we and other evaluators encounter. Some of the issues can be readily addressed. For instance, an agency that focuses on outputs—and that has leadership and staff that want to learn and get better—is a well-suited candidate to expand their focus to questions about outcomes and, more broadly, for building evaluation capacity (see Chapters 7 and 13). Similarly, those who direct and run programming for which evaluation has been focused solely on outcomes will often have interest in how they can improve, how to connect the dots, and how to capture the specific impact of their efforts. Those collaborations can be very positive.

It can be much harder to partner when critical leaders or staff are not interested or willing participants in the effort. There are multiple stakeholders and influences at play in any evaluation and any potential partnership. Every organization or program head does not really want to know or learn about what has gone well versus less well. As evaluators, we can communicate directly, work with stakeholders to design and conduct a strong evaluation, and consider data-guided actions to which the results point. However, as we have conveyed in this chapter, we do not always have control over (or access to needed) data or the degree to which our evaluation results and recommendations are used. Sometimes, difficult dynamics can create tensions (e.g., relationally, ethically) or raise concerns about the political or relational implications of our work and our own decisions and actions. In such circumstances, we need to stay true to our own values and moral code. Evaluation projects will come and go, but our integrity is the bedrock on which many relationships, and our careers, are constructed.

While every attempted partnership does not work and does not result in a successful collaboration, and those issues can impact our evaluations, we believe that it is important to keep trying. The opportunity to effect change and to influence the impact and reach of programs, organizations, and systems is a significant motivator, and evaluation is a prime mechanism for our communities to know how to allocate resources to address important issues. We have found that a partnership-based approach is particularly useful in facilitating evaluation-focused collaborations. The next chapter engages this topic in some detail.

FURTHER READING

Access Alliance Community-Based Research Team, in collaboration with Switzer, S., & Adams, M. (2012). *Community-based research toolkit: Resources and tools for doing research with community for social change*. <http://communityresearchcanada.ca/wp-content/>

[uploads/2018/05/Community-Based-Participatory-Research-CBPR.pdf](https://www.cdc.gov/nczod/oddsat/docs/default-source/evaluations/2018-05-Community-Based-Participatory-Research-CBPR.pdf)

Centers for Disease Control and Prevention. *Introduction to program evaluation for public health programs: A self-study guide*. <https://www.cdc.gov/eval/guide/introduction/index.htm>

Heider, C. (2016). *What happens after the evaluation is completed?* Independent Evaluation Group/World Bank Group. <https://ieg.worldbankgroup.org/blog/what-comes-after-evaluation-completed>

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community-based research: Assessing

partnership approaches to improve public health. *Annual Review of Public Health, 19*, 173–202.

Macfarlan, A. (2014). *Lessons learnt*. Post at BetterEvaluation.org. https://www.betterevaluation.org/en/evaluation-options/lessons_learnt

Also see the Community Toolbox, administered via the University of Kansas: <https://ctb.ku.edu/en>

KEY CONCEPTS

Latrogenic effects: Possible unintended, harmful, or negative consequences.

Implementation fidelity: The degree to which the program or initiative is being conducted in a manner that is consistent with the model, intervention, or framework selected for the effort. When practices and processes align strongly with the planned model, the effort is said to have high implementation fidelity.

Outcome evaluation: An approach to evaluation that focuses on the program or initiative's effects, including potential benefits as well as unexpected or negative effects.

Partnership: Working relationship in which the parties involved identify and work toward common interests,

build their efforts on the partners' complementary skills and perspectives, and develop processes and establish outcomes that yield mutual benefit.

Partnership approach to evaluation: An approach to evaluation in which the involved parties engage in open and direct communication, employ participatory practices, and ensure that the effort yields mutual benefits. In our view, this approach also uses evaluation to benefit programs, program participants/planned beneficiaries, and organizations.

Process evaluation: An approach to evaluation that focuses on how the program or initiative is operating. It involves monitoring and assessments of implementation. A particular emphasis is on whether the program is being conducted as planned.

QUESTIONS FOR REFLECTION

1. As you reflect on organizations or programs with which you have experience, consider the roles and potential benefits of evaluation for their programming. To what degree were they able to answer questions of interest to their stakeholders? In what ways would evaluation have been helpful?
2. What pitfalls or barriers to evaluation have you encountered in your own experiences or do you know of in your community? What dynamics and influences were at play? Why was evaluation not conducted or, if evaluation occurred, why were its findings not used?
3. Consider how the issues described in this chapter have arisen in your own work. Identify an experience or involvement that faced one or more of these challenges, and describe the strategies you employed—or could have employed—in trying to ensure the usefulness of your efforts.
4. More broadly, consider the nonnegotiable elements in your own approach. That is, what values, principles, or ethical codes guide your work? How do those relate to both the promise and potential benefits of evaluation as well as the possible pitfalls?

REFERENCES

- Brashears, F., Davis, C., & Katz-Leavy, J. (2012). Systems of care: The story behind the numbers. *American Journal of Community Psychology, 49*(3-4), 494–502. doi:10.1007/s10464-011-9452-z
- Center for Mental Health Services. (2003). *CMHS national evaluation: Aggregate data profile report—Communities funded from 1997 to 2000*. ORC Macro.
- Center for Mental Health Services. (2004). *Comprehensive community mental health services program for children and their families evaluation findings: Annual report to Congress 2004*. Macro International.
- Cook, J. R., & Kilmer, R. P. (2004). Evaluating systems of care: Missing links in children's mental health research. *Journal of Community Psychology, 32*(6), 655–674. doi:10.1002/jcop.20024
- Cook, J. R., & Kilmer, R. P. (2012). Systems of care: New partnerships for community psychology. *American Journal of Community Psychology, 49*(3-4), 393–403. doi:10.1007/s10464-012-9516-8
- Cook, J. R., Kothandapany, N., Weber, L. J., Wall-Hill, S., & Kilmer, R. P. (2008). Evaluation of a family organization: Integrating local and national evaluation data. In C. Newman, C. Liberton, K. Kutash, & R. Friedman (Eds.), *Agenda and Proceedings: 21st Annual Research Conference: A System of Care for Children's Mental Health: Expanding the Research Base* (pp. 185–186). University of South Florida, Louis de la Parte Florida Mental Health Institute, Research and Training Center for Children's Mental Health.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3-4), 327–350. doi:10.1007/s10464-008-9165-0
- Fixsen, D., Naoom, S., Blase, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network.
- Gilna, D. (2016). "Scared Straight" programs are counterproductive. *Prison Legal News*. Retrieved from <https://www.prisonlegalnews.org/news/2016/jun/3/scared-straight-programs-are-counterproductive/>
- Haber, M., Cook, J. R., Kilmer, R. P., & Hemphill, B. (2010, March). *Relationships of child and family team meeting processes to initial and long-term changes in youth functioning*. Paper presented at the 23rd Annual Mental Health Research and Policy Conference, Tampa, FL.
- Hemphill, B., Cook, J. R., & Kilmer, R. P. (2010, March). *Child and family team processes as predictors of changes in youth functioning*. Paper presented at the 23rd Annual Mental Health Research and Policy Conference, Tampa, FL.
- Holden, E. W., Santiago, R. L., Manteuffel, B. A., Stephens, R. L., Brannan, A. M., & Soler, R. (2003). Systems of care demonstration projects: Innovation, evaluation, and sustainability. In A. J. Pumariega & N. C. Winters (Eds.), *The handbook of child and adolescent systems of care: The new community psychiatry* (pp. 432–458). Jossey-Bass.
- Johnson, V. A., Bishop, J. T., Malo, A., Lott, F. M., Wilson, C. A., Kilmer, R. P., & Cook, J. R. (2011, March). *Assessing the impact of family support programming in a system of care: Associations between type and intensity of support provided and family outcomes*. Poster presented at the 24th Annual Mental Health Research and Policy Conference, Tampa, FL.
- Kilmer, R. P. (November, 2015). *Promises and pitfalls in program evaluation: Applied community research and ... inaction*. Invited colloquium—Department of Psychology Excite and Engage Speaker Series, University of North Carolina at Charlotte.
- Kilmer, R. P., & Cook, J. R. (2012). Moving forward with systems of care: Needs and new directions. *American Journal of Community Psychology, 49*(3-4), 580–587. doi:10.1007/s10464-012-9513-y
- Kilmer, R. P., Cook, J. R., & Palamaro Munsell, E. (2010). Moving from principles to practice: Recommended policy changes to promote family-centered care. *American Journal of Community Psychology, 46*(3-4), 332–341. doi:10.1007/s10464-010-9350-9
- Kilmer, R. P., Wall-Hill, S., Cook, J. R., Kothandapany, N., & Weber, L. J. (2009). Evaluating ParentVOICE: Building family organization sustainability through family-university collaboration. In S. Swart, B. Friesen, A. Holman, & N. Aue (Eds.), *Building on family strengths: Research and services in support of children and their families* (pp. 164–168). Portland, OR: Proceedings of the Building on Family

Strengths Annual Conference and State of the Science Report. Portland State University, Research and Training Center on Family Support and Children's Mental Health.

Kloos, B., Hill, J., Thomas, E., Wandersman, A., Elias, M. J., & Dalton, J. H. (2012). *Community psychology: Linking individuals and communities* (3rd ed.). Belmont, CA: Cengage Learning.

Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2005). "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. *Campbell Systematic Reviews*, 1(1), 1–62. doi:10.4073/csr.2004.2

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M. E., & Lavenberg, J. G. (2013). Scared straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review. *Campbell Systematic Reviews*, 9(1), 1–55. doi:10.4073/csr.2013.5

Rivlin, A. M. (1971). *Systematic thinking for social action*. Washington, DC: Brookings Institution.

Schorr, L. B. (1997). *Common purpose: Strengthening families and neighborhoods to rebuild America*. New York: Anchor Books, Doubleday.

Tynan, J. M., Larson, J. C., Mihalski, J. L., Bishop Scott, J. T., Hogan, K. S., Cook, J. R., & Kilmer, R. P. (2015, June). Translating research into action: Successes and barriers enacting recommendations within a school system. In J. C. Larson (Chair), *A partnership approach to evaluating a publicly funded pre-kindergarten program: Key findings and action steps*. Symposium conducted at the 15th Biennial Conference of the Society for Community Research and Action, Lowell, MA.