

## EVALUATIVE RESEARCH

## Key Concepts and Applications in Facilitating Change

James R. Cook and Ryan P. Kilmer

*What is program evaluation? What can it do for me, my organization, my program, my partners, and my community?*

When we talk about program evaluation or evaluative research, we need to be clear about what these terms actually mean. “Evaluation” and “research” are both words that can have some emotional loading or elicit strong reactions. The notion of being evaluated evokes among many a bit of anxiety, perhaps because of negative experiences they have had in the past or because they feel that evaluation means someone will be criticizing or diminishing what they are doing (for some, it may seem as if evaluation is spelled “e-v-i-l- ...”). In a somewhat different way, the notion of research is sometimes viewed with some hesitation. For some, it may call to mind prior training experiences in which research was equated with statistics and statistical tests, which some find aversive. Alternatively, for others, the idea of research may stimulate thoughts of being treated as guinea pigs or in a rather dehumanizing manner or, in other instances, of efforts that yield no practical utility.

When we, the authors of this chapter and the editors of this book, talk about program evaluation and evaluative research, we are talking about ways to answer questions of importance to you and the people around you. We would guess that you conduct research and evaluation in your everyday lives on a regular basis, even if you do not consider it as such. Imagine trying out a different route to work to see if you can get there sooner and comparing it to the typical way you travel. You might ask coworkers to suggest alternative routes and/or collect data about how long it took them to get to work. Or perhaps you like to bake a favorite type of cake and you want to know if reducing the oil in the cake will result in a cake that is similarly moist and delicious. In each case, you might want to take some careful measurements (e.g., of the time of day at which you leave

## Learning Objectives

1. Understand the distinction between evaluation and research.
2. Describe questions that must be answered when planning an evaluation.
3. Outline steps in developing an evaluation.
4. Understand the potential benefits in taking a partnership-based approach to evaluation.

home and arrive at work, and the amount of oil you use) that would allow you to draw conclusions about what you would want to try next time. If the cake is dry and dense, or the new route takes an extra half hour to get to work, you would likely conclude that the “experiment” was unsuccessful, and either go back to the old ways of doing things or try out a different change that you could test. In each case, you are conducting evaluative research to answer questions of importance to you that will allow you to make decisions about how you proceed in the future.

For us, evaluation is a way of “doing business” and, more generally, a way of life. We try things out, we pay careful attention to what we do and how it turns out, we evaluate how well the process played out, and then we make decisions about what we want to do in the future. We do this at work and in almost every aspect of our lives. For some things (such as the driving to work example), the “test” is rather simple, with a clear process (a specific route) and outcome (time to work), although to address our question we do need to make sure that there was not a random accident that occurred on our new route or that differences in time to work did not simply reflect that we happened to be traveling at a time when school buses are dropping children off at school and blocking traffic. In some types of research, modifying or assessing the process is simple (reducing one-third cup of oil in the recipe), but gauging the outcome may not be as straightforward—for instance, you may like your cake a bit drier, but your family may like it more moist, or the cake may be really good when fresh, but the new version becomes drier and less palatable over time. As you try to gauge what is “good,” involving others in the decision-making process can be very important, even imperative, unless you are the only one eating the cake!

**A key to any evaluation is that we are trying to answer questions of importance to our partners and to us.**

These everyday examples point to some key elements that are applicable to more complex types of evaluation efforts. Programs often have multiple components, different types of people who participate in them, changing issues of focus or varying processes over time, and a range of outcomes that can be hard to measure. However, the key aspects of any evaluation, from our perspective, are that we are trying to answer questions of importance to our partners and to us. We work together to develop the questions and find ways to obtain the answers. The knowledge we gain allows us to move forward to better accomplish our goals.

While we have used program evaluation and evaluative research somewhat interchangeably in our discussion so far, there are times when it is important to make distinctions between what is defined as “program evaluation” versus “research.” We will first outline the distinctions that are made and the reasons for those distinctions and the circumstances under which it is important to distinguish evaluation from other forms of research. We will then consider the key questions that must be answered when planning a program evaluation and discuss the ways that the answers to those questions have implications for how you would think about conducting an evaluation in the context of a partnership. Next, we will provide an overview of some major steps to take when starting to evaluate a program, policy, or practice. Lastly, we will describe some issues that you need to keep in mind to be able to successfully forge ahead with an evaluation effort while fostering and maintaining a partnership.

## DEFINITIONS: EVALUATION VERSUS RESEARCH

---

*Program evaluation* is generally defined as a set of mechanisms for collecting and using information to (a) learn about projects, policies, and programs; (b) determine their effects, both intended and unintended; and (c) understand the manner in which they are implemented (Cook, 2014). This definition might sound to many like this is conducting research, and there is clearly important overlap among methods and processes used when doing evaluation and conducting research. In many instances, program evaluation is appropriately viewed as a type of research. However, it is helpful, and sometimes essential, to distinguish program evaluation from research (see Rogers, 2014, and Small, 2012, for brief summaries of the distinctions drawn).

Perhaps the main distinction that is used to distinguish between research and evaluation is the purpose of each. Whereas the primary purpose of research is viewed as the creation of generalizable knowledge, the purpose of evaluation is more typically seen as focusing on the effects of a particular program, policy, or practice, which may or may not be generalizable to other programs, settings, or populations. This distinction has been made by the Code of Federal Regulations (U.S. Department of Health and Human Services, 45 CFR 46.102(d)) relating to the Protection of Human Subjects, which defines *research* as “a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.”

In this definition of research, evaluation can certainly be viewed as research if conducted to “contribute to generalizable knowledge.” Certainly evaluation is a “systematic investigation,” and most efforts to evaluate programs use research methods and analytic strategies that are often indistinguishable from those used when conducting research. Also, many evaluative efforts include a combination of research and evaluation (see *Combining Evaluation and Research: Evaluation of a System of Care—describing our evaluation of a local system of care—which was both*).

Why does this distinction between evaluation and research matter? For individuals who work in universities or who receive federal funding for their work, the federal regulations cited earlier require that research involving human subjects be reviewed by an Institutional Review Board (IRB) to ensure that certain standards are met for protecting human subjects who participate in research. Program evaluation, which is not conducted to contribute to generalizable knowledge, and therefore not defined as research, is exempt from review by IRBs and not subject to the same requirements as research (e.g., participants in child and family teams did not have to review a lengthy informed consent statement, sign it, and give it to the researchers at every team meeting as would typically be required for “research”). Of course, program participants providing information needed to evaluate a program must be protected from harm, and their confidentiality should be protected; the central point here is that the rules are different and review requirements are different, based on how the work is defined.

Separate from the “legal” definition and the applicability of standards for protection of human subjects, there are other implications of making a distinction between research and evaluation, depending on the context in which the terms are used. For instance, in some academic settings, evaluation may be viewed as less important (or of lesser status) than research, meaning that applied researchers (particularly those who are more junior

## CASE EXAMPLE: COMBINING EVALUATION AND RESEARCH: EVALUATION OF A SYSTEM OF CARE

We were the lead evaluators of a project designed to transform mental health care for children and families in our county. Funded by the Substance Abuse and Mental Health Services Administration (SAMHSA), our community developed a “system of care” for children with severe emotional disturbance and their families. One requirement of the funding was participation in the “national evaluation,” a longitudinal effort designed to track changes in children and families resulting from system and community changes. Using protocols developed by a large research organization that coordinated the national evaluation, these data were then submitted to that organization and compiled across different sites throughout the country to help SAMHSA learn about the types of services and supports that contributed to child and family improvements. This, then, was generalizable research, to help understand the outcomes of these systems and what contributed to better outcomes and to guide future efforts. In addition to the “national evaluation,” each site was encouraged to conduct “local evaluations” to help guide practice within that community. At our site, we focused on implementation of key components of a system of care, particularly the degree to which system practices and processes were in line with the practice model that had been adopted. That meant we collected data about the degree to which “child and family team” meetings—in which multiple parties develop a customized plan of care for the child and family—were conducted in a manner consistent with the program guidelines. We used those data to provide feedback to teams, supervisors, organizations, and the system to help the local community implement the program with fidelity and to see if better implementation led to better child outcomes. This local evaluation was

viewed as “program evaluation,” not research, because the focus was on the ways that we could help the program improve its functioning; we could not generalize our findings to other systems.

This distinction had important implications for the evaluation efforts. The local evaluation involved asking all the people participating in a child and family team meeting, including family members, professionals, and others who had an interest in helping the family, to complete brief surveys about the focus and function of the team meeting at the end of each child and family team meeting. The data were collected anonymously and used to inform team facilitators and system administrators about how well they adhered to specific principles. The focus was on improving services at that site, not generalized knowledge, and because of this, the participants in the child and family team meetings were not required to sign informed consent statements at each meeting, which might have taken more time to complete than the surveys themselves. A brief consent statement was included on each survey to ensure that each participant knew why the data were being collected, but there was no need for each participant to read and sign a separate form and no need for the evaluators to collect and store consent forms, making the process more manageable for the evaluators and the participants.

When we submitted the protocol for the evaluation efforts to our university’s Institutional Review Board (IRB) for the Protection of Human Subjects, we made a clear distinction between the national evaluation research and the local evaluation efforts to improve service delivery. The IRB determined that the local “program evaluation” was not research, and thus did not fall under their purview.

in status) who do evaluation may be advised to describe their work as evaluative research (with the emphasis on “research”). In applied or program contexts, the term “evaluation” can have many meanings, including personnel evaluation, and may be viewed as a threat to individuals or programs. On the other hand, research is often viewed by program staff as esoteric and irrelevant.

These different views of program evaluation and research point to the importance of the context in which the work is being done. In different settings, with different funding, the same work may be (appropriately) referred to differently. This implies that we need to make sure that we use language that conveys the nature of the work in a way that avoids the negative connotations that may be present within a given context. It also underscores the need for clear, effective, and direct communication. Regardless of the label assigned to our work, we tend to emphasize the following key questions, upon which we will elaborate later:

1. *What do you/we need to know about the program and population of interest?* We tend to think of these as the “research questions” but, in some contexts, we refer to them as “evaluation questions” or merely “questions needing to be answered.”
2. *What steps would we need to take in order to obtain answers to the questions of interest?* These are the methods we need to use to give us the most unequivocal answers possible.
3. *How would obtaining answers to those questions make a difference in how the program operates?* This is the “so what?” question that, for many, distinguishes evaluation from research. If the primary focus is on the specific program, policy, or practice that is being evaluated, and not on programs, policies, or practices more generally, then this is likely to be defined as program evaluation.

This latter question is critically important. As we conveyed in the Preface to this text, in our conceptualization of evaluation, it is never just about “informing” or “knowing” (i.e., documenting outcomes or impact or reporting on program elements to funders or other stakeholders); rather, our focus is on “effecting change” (e.g., in programs, agency practices, system function) via improving practices and processes and, in turn, outcomes. In our view, the primary purpose of evaluation is to improve programs and interventions, guide program changes, and make decisions about allocation of resources.

## INITIAL STEPS IN THE DEVELOPMENT OF AN EVALUATION

---

*Identify the research questions: What do you/we need to know about the program and population of interest?*

**Our questions shape the course and scope of the evaluation effort.**

The identification of the research questions is much like specifying the destination before you begin a journey. If we do not know where we are trying to go, we may not know which route to take to get there, nor would we be able to tell if we have arrived at the right place. The questions must be clear because they necessarily shape the course and scope of the evaluation effort.

To gain a clear sense of the research questions requires first the identification of the critical people who have a stake in the success of the program (we will refer to them as *stakeholders*), including those who have the ability to act upon the findings. We need to spend time with them, understand their goals, and help them to specify what they want and need to know. This is generally an active process, whereby we attempt to put ourselves

in the position of the stakeholder and ask ourselves what we would like to know if we were in their position.

It is important to underscore that the identification of useful research questions is not a simple process. You might think that all you need to do is ask the stakeholders what they need to know, they will tell you, and you proceed from there. However, in order for this simple process to work, you would need to have stakeholders who

- have spent a significant amount of time thinking about decisions that need to be made about the program;
- understand what information is needed in order to make those decisions; and
- understand the range of possible questions that can be answered.

In our experience, many program stakeholders have not had enough experience with evaluation design and implementation to know what the possibilities are—if someone is not aware of the different options that exist, it is not possible to make an informed decision about what one wants or needs. This would be like going into a restaurant and being invited to order your food without knowing what is on the menu. Our job, as evaluators, is to provide the “menu,” explore with the stakeholders what they like (e.g., what level of spiciness do they want) or think they need, explain the trade-offs that exist in their decisions (some dishes take longer to prepare than others, some cost more than others; the decision about what to order will likely take into account how hungry they are and what their budget will allow), and help them make an informed choice about how to proceed. Making an informed decision requires that they have come to understand their needs, and can clearly articulate the options they have, based on the logistical constraints they face (or, more accurately, that we face together, because resources are always finite).

To determine what types of questions would be useful, it is crucial to gain an understanding of the program and what it is trying to do. This helps you think about the information needs from the perspective of the program and its multiple stakeholders. In particular, the evaluator will want to develop a good understanding of the following:

1. The program’s intended beneficiaries (this term is used rather than “client” or “participant” because it reflects the fact that, often, programs are designed to benefit a range of people). In conceptualizing an evaluation and its questions, we want to be clear regarding who is expected to benefit from the program. While the expected beneficiaries are often those who directly participate in the program, in many programs there are others who are expected to benefit. For instance, the beneficiaries of a school-based social skills intervention could be the students who gain additional competencies, their classmates who have more positive interactions with the children who received the program, the classroom teacher who is able to spend more time teaching and less time intervening in disputes, and/or the parents who find their children to be better behaved and parenting to be more positive.
2. The program’s expected benefits or results. Programs are generally created to contribute to or result in certain outcomes or benefits, although it is quite

common for programs over time to effect and/or identify other unintended consequences, positive and negative. Understanding those expected and unexpected consequences of the program or policy is important for developing an evaluation.

3. The actions taken by the program to effect those benefits. This refers to what the program does to bring about change. Rarely is there a single program element that effects change. Rather, change often comes from a complex set of interactions that include the identified program components, as well as multiple other actions that, while perhaps not the intended active elements, are critically important for effecting change. For example, a program to help educate new mothers about how they can take care of their infants may be created based on a belief that new knowledge is an important impetus for improved parenting. However, if the education is provided by visiting nurses or social workers, there may be important elements of skill building or role playing that occur, as well as social support that helps the mothers become more comfortable in their role and better able to parent (separate from the knowledge gained).

Sometimes the initial research questions from stakeholders are very basic and specific. We have heard program directors indicate that they wanted to know how many people they are serving in different ways. This is what we might describe as a “monitoring” question, rather than an evaluation question, and suggests a need for improvements in the program’s internal data management capacity. If the program cannot answer a question as basic as that with its existing records, a first step before any real evaluation questions are answered would be to help them develop the capacity to reliably record and monitor their operations. Although this question is simple and not really a question of evaluation of the program, it opens the door to discover additional questions that may arise if the data were available to answer them. For example, it may be that there have been demographic changes in the program’s “catchment area” (the geographic area served by the program) and program leaders wish to understand whether “newcomers” to the area are being served and what services they are utilizing. This could have important implications for outreach efforts or changes in the nature of the programs offered. Helping the program managers or other stakeholders move beyond “monitoring” questions is critical for knowing what data to collect and how to organize the information. It is important to think beyond the immediate question (e.g., monitoring) and imagine what other questions may be relevant, even if the program has no current capacity to answer the questions. Anticipating the questions that could be answered may stimulate simple changes in the data collected on an ongoing basis, enabling answers to relevant questions in the future.

As illustrated in Box 1.1, which outlines some common evaluation questions, program stakeholders are often interested in knowing how well the program is meeting the needs of the individual participants and/or “collaterals” (e.g., those connected with or related to the program “participants” or “clients,” such as parents, spouses, neighbors, or coworkers) and/or the broader community. For example, programs may have high impact with a very small number of people, which may meet their individual needs, but if the goal is to address the needs of all (or a very significant subset) of those in the community, it would likely

## BOX 1.1 COMMON QUESTIONS ADDRESSED IN EVALUATION EFFORTS

**Monitoring:** How many people are being served? In what ways? What are the characteristics of those being served? Is this the population the program intends to serve? Has that population changed over time and, if so, why? Are there beneficiaries besides the population being directly served?

**Overall Impact Evaluation:** How well does the program effect change among the intended beneficiaries? To what degree is it having the desired effects?

**Differential Impact of Program for Different Participants:** Does the program have different effects for different types of participants? Do some respond particularly well to the program, whereas others do not?

**Fidelity of Implementation:** Is the program (or its components) being implemented as intended? Are practices consistent with the model or plan that has been adopted?

**Differential Impact of Specific Program Components:** Do specific aspects of the program lead to different

outcomes? For different types of people? Do different amounts of treatment lead to different outcomes? Are there specific doses required to have the desired impact? How does the integrity or consistency of the implementation relate to outcomes?

**Community Impact:** What are the needs of the population being served? How well is the program able to address these needs? Is there a reduction in unmet need in the community?

**Accessibility:** Are the intended beneficiaries of the program able to access the program? Are cost, distance, transportation, culture, or other factors limiting the ability of some people to benefit from the program?

**Cost Benefit or Cost Effectiveness:** What is the return on the investment made in operating the program? How does the cost of implementing the program compare with the benefits obtained, for the participants or for the community? How does the return on investment compare with that of other programs with similar goals?

be necessary to determine the needs that remain or that are unmet by the program (with implications for improved efficiency or expansion).

When we are contacted by potential partners about an evaluation or decision makers are considering funding allocations, their questions about the effects of a program are often characterized as “does the program work?” This implies that there is a “yes” or “no” answer—we generally discourage people from thinking about programs in this way because it is rare that programs have no impact (particularly if they have been in existence for a long time). Rather, a focus on how well the program works, and potential variations across different subsets of participants, helps avoid the notion that a program is either a success or a failure. There are many levels of success, and it is important to try to sort out what is working well and for whom. In fact, well beyond framing a question or evaluation around whether a program works or not, when we engage in evaluative efforts, we typically approach them

**Rather than framing an evaluation around whether a program “works” or not, we typically seek to employ data to improve program function and maximize the benefits to those enrolled.**



with the goal of employing data to improve program function and maximize the benefits to those enrolled.

This focus on conducting evaluation to make changes in the program is often referred to as a “*formative*” *evaluation* (see Chapter 16 for more information about formative, summative, and process evaluation). As the name suggests, a formative evaluation is most likely to occur when a program is new and developing, with the program using data on an ongoing basis to make corrective changes until it becomes established and stable. This is related to the notion of a “*process*” *evaluation* that focuses on how well the program is implemented in relation to its design and intent. This type of evaluation helps the program management make changes to improve its “fidelity” or consistency with the tenets of the program, practice framework, or curricular model. This focus on processes and making changes can be contrasted with the notion of a “*summative*” or “*outcome*” *evaluation*, which suggests that the purpose of the evaluation is not to improve the program, but to determine the effects of the program. This implies that the program has become stable and that the evaluation can be conducted over a sufficient time period to determine the effects of a particular, static program.

These distinctions between formative and summative, or process versus outcome, evaluations often are based on assumptions that we rarely see in practice. Most programs are not static, even after operating over many years. In fact, the longer a program has been operating, the greater the likelihood that the program has experienced some “drift” in its goals and/or operations; that is, the program’s function or objectives have changed from what they were initially. Program operations may also change more deliberately, due to changes in resources, client needs, or staff capabilities. With the exception of the rare “controlled trial,” in which the program is intentionally kept static while it is being evaluated, most programs undergo changes over time, which may not always be obvious to program management. Please see Case Example: Needs Assessment with a YMCA, which illustrates how broader community and contextual factors can be salient contributors to program evolution and change.

In addition, we often need to help the program’s stakeholders get past an erroneous assumption that if the program is designed to do something, then it certainly must be doing that. Clearly, many (perhaps most) programs are not implemented as planned, and sometimes there is a major gap between the intended and the actual implementation of the program. Thus, the degree to which the program is implemented as designed and the relation between implementation and outcomes are often very important questions to answer, for both new and well-established programs. These answers can be crucial for making program-related changes, particularly given the growing evidence that program implementation is often highly related to the effectiveness of the program (Durlak & DuPre, 2008). Similarly, because programs rarely have all the resources they need, the efficiency of programs (i.e., their return on investment) is often important to assess (see Chapter 12 for a discussion of cost-benefit and cost-effectiveness types of evaluations).

Because program stakeholders often have not had the training or experience to readily articulate the range of questions that an evaluation might answer, it is useful for the evaluator to “put on the hat” of the program management and attempt to understand what questions might be of interest to them. In addition to the questions already mentioned, program leadership or other stakeholders may be concerned with the degree to which the intended beneficiaries of the program can access the program. This can mean that there

## CASE EXAMPLE: NEEDS ASSESSMENT WITH A YMCA

We were once asked to help a YMCA undertake a needs assessment to determine what services the Y could provide to the surrounding community. The Y was originally created to serve the surrounding neighborhoods, and a main feature of the Y was its Olympic-sized swimming pool and strong competitive swim team. However, over the prior decade, the neighborhood had undergone significant changes, such that the largely white, working-class neighborhood had changed, with many families moving to other neighborhoods in more affluent suburban areas. Now the neighborhood surrounding the Y was largely populated by poor, minority families. Many of the families who had moved away continued to take advantage of the swimming program, and others from those more distant neighborhoods became

members and used the Y. New leadership at the Y recognized that it was no longer meeting its mission of serving the local community, and they wished to learn from the neighborhood residents how the Y could best serve them so the Y could shift their programs to address their needs. Note that while the actual activities of the Y had stayed fairly constant over time, changes in the surrounding population resulted in it no longer achieving its goals. The needs assessment identified neighbors' needs and the potential barriers (e.g., financial, transportation) the neighbors would experience, and the Y made a major shift in their programs and fee structure to address these issues. They made a change to better serve their new neighbors and meet the needs of their changing context.

are concerns about distance and transportation, which may impede the intended beneficiaries' abilities to get to and use the program. However, access may also be limited by the cost of the program (families may not be able to afford the program), the hours of program operation (if it is only open from 8 to 5 on weekdays, working people may have difficulty getting off work to participate), and cultural appropriateness (the staff may not speak the language or may not be sensitive to the cultures of the intended beneficiaries). If a program is not accessible or acceptable to the intended beneficiaries, many will not use the program. In short, there are many questions that can arise in an evaluation of a program, and it is useful to explore the options with the program's stakeholders before proceeding too far.

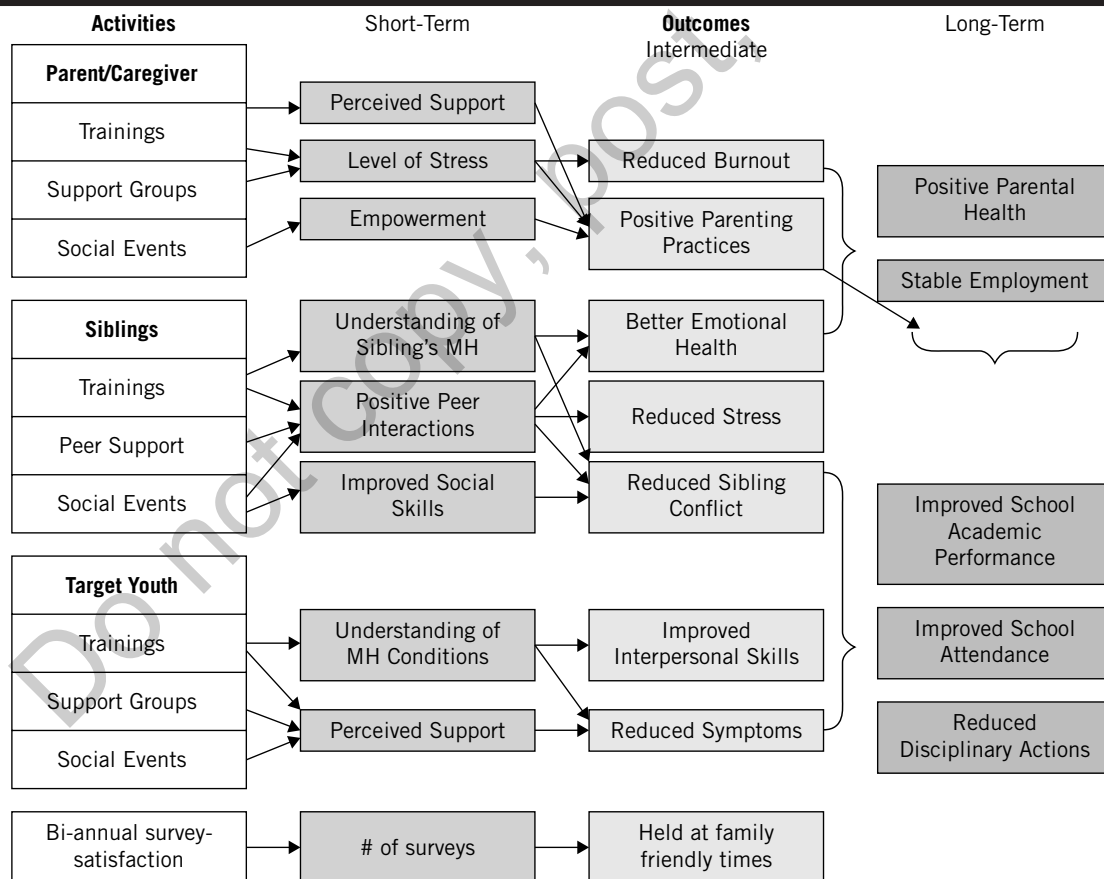
As the questions evolve, it is useful to continue to ask what the stakeholders already know about how and how well the program benefits the population of interest and whether they see it as benefiting some more than others. These questions then generally lend themselves to asking about how they might know if the intended program benefits occurred. This then starts the journey into understanding what types of measurement and methods might be most useful for answering the research questions of interest. Before we propose to collect any new data, we want to know and understand the existing data collection efforts and how we can use the program's ongoing data management structures to evaluate the program.

This process of asking questions and working to understand what the program is trying to accomplish is the start of what we refer to as an articulation of the program's "theory of change." A theory of change describes the processes that the program has put in place to accomplish its goals. Specified within a theory of change are the program activities that are intended to effect certain short-term changes; these short-term changes are then prerequisites for later and/or larger changes that subsequently lead to the longer-term goals of the program. For example, the leadership and staff of a program designed to

provide support to families of children with severe emotional problems recognized that they needed to be able to document the impact of their services and supports in order to sustain their funding. The program provided a range of services to the parents, the children with emotional problems, and their siblings, in order to strengthen the family and ultimately improve the mental health of the diagnosed child. To help clarify how these different activities were expected to lead to benefits for the parents or caregivers as well as improvements in the diagnosed child, a *logic model* was created that laid out, in a diagram, the program's theory of change. The creation of a logic model, to graphically illustrate how a program is expected to work, can be a critical tool for informing organizational and evaluation planning. The logic model can help frame an evaluation's focus and objectives, from its data sources to its key questions and indicators of outcome. A logic model is also very important for helping the different stakeholders, including the evaluator, be clear about the manner in which the different program components and goals logically fit together.

As seen in Figure 1.1, a de-identified and simplified logic model for this family support program, there were multiple ways that the different program components were expected to lead to improvements in children. Parent support groups were expected to lead to lower

**Figure 1.1 A Logic Model of a Family Support Program**



levels of stress among the parents, which would lead to more positive parenting and better child outcomes. Similarly, the activities provided for the siblings helped them to maintain a positive peer group and to understand the needs of their diagnosed sibling, leading to lower levels of conflict in the household and helping the sibling maintain emotional health. The activities for the child with the diagnosis (i.e., the target) were expected to help provide some respite for the parents and help develop greater social and interpersonal skills in the child. The collective impact of these different shorter-term outcomes was then expected to help reduce the symptomatology of the diagnosed child and increase his/her positive functioning at home and school. The logic model helped clarify the intermediate steps that contribute to better outcomes for the diagnosed child, and therefore helped identify the measures that needed to be used at different points in time to determine whether the program was having the desired results.

The logic model helps “connect the dots” between different program elements and describes how they contribute to different types of change in the (sometimes multiple) intended beneficiaries of the program. As a logic model is created and shared, the research questions can and should be refined, clarified, and prioritized as more is learned about the program and its operation. It is important to understand also that it is rare for a program to have the resources available to answer every question of interest in an evaluation, which means that the prioritization of the questions becomes increasingly important with more complex logic models.

*Outline the methods: What steps are needed in order to obtain answers to the questions of interest?*

The measures and methods need to follow the research questions. While this may seem to be an obvious statement, there are many times when people affiliated with programs identify a way of measuring something about the program, and then see it as a useful way to evaluate the program. Starting with the measure and allowing that to determine the nature of the evaluation is like the tail wagging the dog; the measurement used should always follow from discussion and understanding about what we want and need to know. The question(s) of interest should always guide the methods and measures of an evaluation. Remember that “not everything that can be counted counts, and not everything that counts can be counted” (Cameron, 1963, p. 13). It is critically important to have a clear understanding of what the program needs to know, and then focus the methods on obtaining answers to their important questions (please see Case Example: Ensuring Alignment Between Measures and Program Goals as an example).

**According to Cameron (1963, p. 13): “not everything that can be counted counts, and not everything that counts can be counted.”**

Once we have reasonable clarity about what the program is trying to accomplish, its theory of change, and research questions of importance, decisions about the methods to use for answering the questions require some additional information.

What are the *sources of information* for answers to the questions? Program participants are often an important source, but there may be other people, such as parents, spouses, teachers, or supervisors, who may be able to shed light on whether the participants are changing in the ways intended by the program. The answers to questions may come from existing records kept for other purposes (e.g., attendance records or test scores in schools; crime reports; emergency room visits), and data may come not only from the people

## CASE EXAMPLE: ENSURING ALIGNMENT BETWEEN MEASURES AND PROGRAM GOALS

We were once working with a senior center to help them evaluate the effects of an exercise program on the seniors who participated. While the program was primarily focused on improving or maintaining the seniors' quality of life, the county funders insisted that program staff assess weight loss. When the program leaders protested that many of the participants were fairly frail and should not lose weight, the county insisted that they have the seniors weigh themselves regularly and that the program track and report these weights. Since weighing the participants was not a costly endeavor and not viewed by the participants as a problem, the program

leadership decided that they could easily satisfy the county's demand, but that they would still want to develop strategies to answer questions of importance to the program (e.g., how well participants could engage in activities of daily living). However, it would have been tragic if the county made its funding contingent on weight loss, when that was not a primary goal of the program and could potentially be hazardous to the participants. Alternatively, it may have required that the program recruit overweight seniors to "successfully" participate in the exercise program.

expected to benefit from the program, but also from others presumably not affected (a comparison group). The type of sampling is also a decision to make. For small programs, the "sample of the whole" may be appropriate, but for larger programs, a subset may be selected, and it is important to make sure that the sample is representative. As the sources of data become clear, the evaluator must almost certainly revise the questions to make them more specific and precise. Some questions may also be abandoned or revised based on the conclusion that there are no feasible mechanisms for answering the questions. For example, if an appropriate comparison group is available, a research question may shift from "how much do participants gain?" to "how much more do participants gain than the comparison group?" Please see Case Example: Evaluation of an Early Childhood Program as an example.

It is also necessary to determine the timeline under which the evaluation will be working. *When are results needed* and from what time period will the evaluation data be obtained? Sometimes answers to research questions are needed quickly, for example, when budgetary decisions about program funding are being made. In other instances, programs may have the luxury of collecting data over time to provide answers. The methodology of choice and, to some extent, the questions that can be answered are determined in part by the timeframe that is required. For example, if answers are needed quickly, then the evaluation may need to rely on records that are already available (known as *archival data*). Of course, without good records going back in time, it may be difficult to understand the longer-term effects of the program quickly; instead, it would likely be necessary to follow up or track the participants over time. That said, even if you do have data available over a period of time, the data may not capture the information that is wanted or needed. This can then require further revision of the questions to be answered to account for the constraints of data and time. When this type of shift occurs and the questions deemed important cannot be adequately answered, the evaluator and the program management

## CASE EXAMPLE: EVALUATION OF AN EARLY CHILDHOOD PROGRAM

In our work with a large school system to evaluate an early childhood intervention program, we wanted to know whether children evidenced growth in their social-emotional skills as a result of the program. We identified measures for teachers to assess their students' social-emotional development and planned to have all teachers assess their students. However, the school administrators felt it was important to provide feedback to the teachers about their students' performance on the measure. We, as evaluators, noted that providing feedback would interfere with our ability to evaluate the program because this measure had not been used before and there were no plans in place to continue to use that measure. As such, the addition of feedback to teachers was a change in the program that would occur only during the evaluation and could potentially improve teachers' ability to educate their students, that is, the version of the program we were to evaluate might have better outcomes than the version of the program that had been in place

(and would continue after the evaluation). In sum, we would not really know how well the program "as is" would do in the absence of the new "intervention" (teacher feedback). However, because the program was large with several thousand children, we proposed that we randomly select half of the teachers for a "feedback group," allowing us to determine the impact of the feedback on the students' development over the course of the year in the program compared to those students of teachers who did not receive feedback. Thus, we were able to respond to the questions and concerns of the program leaders by changing the research questions and the methods to include a good comparison group to answer questions of importance to the program. In addition, because the students in the "feedback" condition gained more than the students in the "no feedback" condition, we learned about a simple, inexpensive way to improve student learning (see Cook et al., 2014; Gadaire et al., 2020, for more).

must carefully determine whether the "new questions" are sufficiently important to warrant the time and energy that the evaluation would entail. If not, one option may be to retain the original evaluation questions and set up the data collection and management systems to allow answers at a future time. When using existing data, you also have to be careful to determine whether the data are accurate because they may not have been collected and recorded as consistently and carefully as needed. When you have the ability (and resources) to determine what data will be collected and how, you clearly have more latitude in answering the questions of greatest interest.

The timing of data collection is also important as you consider, in your theory of change, when you might expect to see program effects. If you collect data shortly after participants complete a program, you will only be able to assess the short-term effects of the program. If you anticipate that the effects take a longer time to be realized (or if longer-term effects would be most important, useful, or compelling to assess), then you will want to determine when you would best be able to measure those and collect data at an appropriate time (or at multiple times). Sufficient detail in your logic model/theory of change is important for determining when you need to collect data. A focus on longer-term changes will also generally require an investment in a longer-term evaluation in order to track and collect information over time. Without that commitment, the types of questions that can

## CASE EXAMPLE: IMPORTANCE OF LONGITUDINAL EVALUATION METHODS

In the evaluation of a prekindergarten education program targeting 4-year-old children who were judged to be educationally “at risk,” the school board wanted to know the degree to which the program resulted in educational gains for participants. In particular, it was hoped that the program would help the children become “ready” for kindergarten, and subsequently more successful in school over time. Outcome criteria of interest to the school system were the third and fourth grade end of grade tests. We designed an evaluation that would determine how these children improved in their verbal ability and in their social-emotional development over the course of the year and found that children in the program improved in their verbal and preacademic skills as much as children in a similar program that did not select children who were educationally at risk (the same social-emotional data were not available from children in the other program). Although the children made gains over their year in the program, the school system only funded the evaluation effort for 1 year, with no support to follow the children until third or fourth grade. Consequently, without follow-up, we could not really tell if any longer-term gains resulted from the program. We did, however, examine archival data from children who had been in the program in prior years and found that the gains made in the program year appeared to “wash out” by the time they took the third grade end of grade tests. We also looked at the quality of the

schools they entered and found that those in lower performing schools were less likely to sustain their gains over time. Without the ability to follow up the children over time, we could not identify which aspects of the program (including implementation of the program’s curriculum, which we assessed) were related to longer-term outcomes (nor explore the impact of social-emotional gains or whether they sustained, because the school system did not assess children’s social-emotional functioning as a matter of course). While the use of the archival data provided information about the importance of the schools that the pre-K students entered, follow-up of the students would have provided a much more complete picture of the effects of the program.

On a different note, we do want to point out that a major concern with this evaluation was an attempt to evaluate a 9-month, pre-K program on its impact on third and fourth grade testing, when there are 4–5 years of classroom instruction and other life events that happen between the end of the program and these outcomes. Those test scores are quite distal outcomes! Since test scores are affected by many factors (e.g., teacher quality and school climate, parental education and involvement, traumatic experiences of children, test anxiety), it is probably not wise to evaluate a program serving 4-year-olds on these test scores without accounting for these other factors.

be successfully answered will be limited. (Please see Case Example: Importance of Longitudinal Evaluation Methods as an example.)

*Identify how the evaluation (in particular, the answers to the research questions) will make a difference in program operation. As we indicated before, this is the key “so what?” question that, for us, defines evaluation (as opposed to research more generally).*

The question of how the evaluation will be used to effect change in the program is perhaps the most important question of the three. We urge our partners to consider this question carefully because, if we cannot clearly identify how the results would translate into improvements in the program, it is hard for us to justify spending the time (and using the resources) necessary to do the work. Ideally, this question is being raised at every step along the way, helping ensure that you will not spend time working to answer questions

## CASE EXAMPLE: REVISING RESEARCH QUESTIONS TO BE MORE “ACTIONABLE”

In a state-funded education program, program administrators wanted to know about the effects of a program. We learned that the state mandated the curriculum and dictated the processes for implementation. When we asked how different results would result in changes in the program, the administrators indicated that their hands were tied because they had no power to change it (although they could argue to the state to change its requirements). Upon further discussion, we asked if all program staff implemented the program in the same ways (it would be highly unlikely that they did) and, if they

did not, did the differences in implementation result in different outcomes. Since they believed that there was variability in implementation and that this variability would likely affect outcomes, we revised the research questions, focusing on a careful evaluation of the variability of implementation across different sites, staff, and aspects of the program. This focus had clear implications for the administrators in that it would allow them to identify strategies for increasing quality control through changes in training and the supervision of their staff.

about which stakeholders do not really care or that would be “merely interesting.” If we do not have clarity about how results would be used to effect change, we would suggest that the research questions need to be reformulated or the methods need to be revised to obtain “actionable” results (please see Case Example: Revising Research Questions to Be More “Actionable” for an example).

This example suggests the need to examine all three questions in a way that recognizes how the answers to each can affect the others. The specification of research questions should be made with a clear consideration of their implications for program improvement. Development of the methods to answer the questions often leads to revisions of the questions to match what is feasible, because we do not want to have research questions that cannot be answered. Then, changes in research questions must be re-evaluated to determine if they have important program implications and can inform action. This cycle of adaptation of one set of questions in light of the answers obtained for the other sets reflects part of the dynamic nature of evaluation that, at its core, involves capacity building among partners. That is, the different parties in the partnership must learn from one another, bringing their respective strengths to the table. Ultimately, through this iterative process of working through the different facets of the evaluation process, a set of methods is developed that can answer research questions that have important implications for improving the program.

We recognize that, when planning and conducting evaluations, the goal of having important research questions answered in the most unequivocal manner possible is often subject to compromise as practical considerations (e.g., availability of resources, such as funding and time) exert control over the process. Yet, at the same time, if the goal is to improve the program and its outcomes for participants, finding the right combination of salience and pragmatism is critically important. There is often a trade-off in designing evaluations between answering important questions with less robust methods versus answering less important questions with stronger methods. We believe that this trade-off should always lean toward answering important questions, as long as there is sufficient



rigor of design to gain knowledge that can guide practice. There is certainly little to be gained by using strong (and often costly) methods to answer trivial questions about which nobody cares.

Throughout our discussion of the early steps in conducting an evaluation, the interactive nature of the process of making decisions should become obvious. As evaluators, we ask questions, sort through possible answers, revisit the questions, and continue to try to formulate a plan that will answer important questions and make a difference. As indicated by this book's title and its emphasis, this is work that is often conducted as part of a partnership, and we might argue is *best* conducted as part of a partnership. This focus on partnerships reflects a particular value orientation, one that is consistent with our training and perspective as community psychologists.

### **The Importance of Transparent Values in Evaluation**

Evaluation is an endeavor that is steeped in values, which often become apparent in understanding the dynamic contextual and political processes that lead to the development of programs, their changes over time, and the decisions to evaluate them. Programs are created as reflections of peoples' needs to effect change. The goal or mission of the program is a reflection of what the founders, and hopefully the current leaders of the program, view as important. Similarly, the strategies used to accomplish the mission also reflect a set of values. It is important that we, as evaluators, see these program values as consistent with our own. Because the primary goal of evaluation is to help improve the program to better accomplish its mission, the mission of the program must be something we can actively support. Even if the current processes used to accomplish the mission are less than optimal, the evaluation should help the organization or program create processes that advance the mission as well as possible. For example, we have worked with the local school system to help it evaluate and improve its prekindergarten program (we discuss this partnership in more detail in Chapter 3). We made a point from the beginning that we were strong advocates for "high-quality pre-K programs" because the literature is clear that these programs can have important positive effects on children's growth and development. The goal of the evaluation, then, was to help determine the degree to which the program was implemented in a way consistent with what research would suggest is "high quality," and work to increase the likelihood that every child experienced a "high-quality" program. As seen in this example, we can easily evaluate the program while we advocate for the mission of the program, and work to help the program be successful, with the emphasis on ensuring that every child gets the best program possible. In a sense, when we align the mission of the program with part of our mission as evaluators, then the "client" of the evaluation is the "client" of the program. In the example of the pre-K program, we were always working for the best interests of the children and families being served and, because that was the espoused mission of the program, we could always focus on that when making decisions about how to proceed.

Thus, in our minds, evaluation is not a "value-free" practice or endeavor. Our approach to evaluation is also informed by our values. For instance, facilitating the participation of people in processes that affect them reflects a key part of our value orientation. This participatory orientation is part of a "capacity-building" focus of our work—we want people to participate in evaluation processes so that they can develop their abilities to critically examine what they do and, as a result, make efforts to effect positive changes (see Hogan et al., 2017,

for background on how we train students to use a capacity-building approach in community work). As such, much of what we do in evaluation fits a *community-based participatory research* model (e.g., Viswanathan et al., 2004; also see Chapter 3), in which we involve the stakeholders in determining the questions to be asked, establishing the methods used to answer them, interpreting findings, and developing recommendations for action. Because our goal is to evaluate in a way that leads to action, and evaluators rarely have the power to change the program, it is critical to work together in partnership with program leadership to ensure that the evaluation is answering questions of importance to them and that they can see how answers point to specific changes they might make to improve the program. The buy-in of program leadership and stakeholders is critical for the action to occur because ultimately they are the ones with the power to effect change in the program.

Throughout the next chapters, we will build upon these themes and focus on the ways that we can build partnerships to conduct effective evaluation. While partnerships are not necessary for conducting evaluations (and many evaluators do so without developing partnerships), we find that partnerships are certainly helpful and make the work more rewarding. When conducting evaluation in the context of a partnership, it is much easier to bring together stakeholders, since you have built a relationship with them. We also find that stakeholders are more likely to be honest and forthcoming with information that can be useful for conducting the evaluation when they see the evaluator as an ally working toward a common set of goals. Some might argue that a partnership approach to evaluation runs the risk of evaluators becoming biased and losing their ability to report findings objectively and accurately and make difficult recommendations. However, if we are guilty of bias, it is bias toward making sure that the program is maximizing its benefit to the purported beneficiaries. We work to maintain our role as a critical friend (i.e., “a trusted person who ... is an advocate for the success of that work,” Costa & Kallick, 1993, p. 49) and work with our multiple stakeholders to understand the program and develop strategies for improving it. That is what we see as our role in partnership-oriented evaluations.

## FURTHER READING

Association for the Study and Development of Community. (2001). *Principles for evaluating comprehensive community initiatives*. Author. <http://www.racialequity-tools.org/resourcefiles/CVP062001.pdf>

Community Tool Box. *Developing a logic model or theory of change*. <https://ctb.ku.edu/en/table-of-contents/overview/models-for-community-health-and-development/logic-model-development/main>

Cook, J. R. (2014). Using evaluation to effect social change: Looking through a community psychology lens. *American Journal of Evaluation*, 36, 107–117 <https://doi.org/doi:10.1177/1098214014558504>

Hogan, K. S., Tynan, J. M., Covill, V. J., Kilmer, R. P., & Cook, J. R. (2017). A capacity building framework for community-university partnerships. *Collaborations: A Journal of Community-Based Research and Practice*, 1(1), 1–28. <http://scholarlyrepository.miami.edu/collaborations/vol1/iss1/1>

Logic Model Workbook, Innovation Network. <http://www.nebhands.nebraska.edu/files/logic%20model%20workbook%20innovation%20network.pdf>

The Pell Institute and Pathways to College Network. *Evaluation toolkit: Using a logic model*. <http://toolkit.pellinstitute.org/evaluation-guide/plan-budget/using-a-logic-model/>

## KEY CONCEPTS

---

*Archival data:* Information and records that have already been collected or are available, prior to the initiation of an evaluation or a research effort.

*Community-based participatory research (CBPR):* “A collaborative research approach that is designed to ensure and establish structures for participation by communities affected by the issue being studied, representatives of organizations, and researchers in all aspects of the research process to improve health and well-being through taking action, including social change” (see Viswanathan et al., 2004).

*Formative evaluation:* A type of evaluation focused on making changes and refinements to the program. Most likely to occur when a program is new and developing, this evaluation uses data to make corrections or refinements until the program becomes established and stable.

*Logic model:* A graphic illustration of a program’s theory of change or how the program is expected to work, used to help frame an evaluation’s focus and objectives, from its data sources to its key questions and indicators of outcome. Also used to clarify the manner in which the different program components and goals logically fit together.

*Outcome evaluation:* Also referred to as a “summative” evaluation, the purpose of this type of evaluation is to determine a program’s or initiative’s effects.

*Process evaluation:* A type of evaluation that focuses on how well the program is implemented in relation to its design and intent. For instance, is the program being implemented with fidelity or in a way that is consistent with the methods specified as important for that program?

*Program evaluation:* A set of mechanisms for collecting and using information to (a) learn about projects, policies, and programs; (b) determine their effects, both intended and unintended; and (c) understand the manner in which they are implemented (Cook, 2014).

*Research:* “A systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge,” as defined by the Code of Federal Regulations (U.S. Department of Health and Human Services, 45 CFR 46.102(d)) relating to the Protection of Human Subjects.

*Theory of change:* A description of what the program is trying to accomplish and the processes that are in place to accomplish these goals. This includes specification of the program activities that are intended to effect certain short-term changes; these short-term changes are then prerequisites for later and/or larger changes that subsequently lead to the program’s longer-term goals.

## QUESTIONS FOR REFLECTION

---

1. Consider the key questions and steps outlined in this chapter, that is, identifying the research questions, data sources, methods, timeline, and possible implications of the work. How are these steps interrelated? To what degree does the answer to one influence another?
2. As you consider a program or initiative with which you have familiarity, imagine yourself as a manager or director of the effort. What questions would drive your interest in evaluation? What kinds of indicators would be crucial to track?
3. What are the potential benefits of a partnership-based approach to evaluation? What might be some challenges inherent in this participatory, collaborative approach? How might a new

- or novice evaluator get started with such an approach?
4. We note that, for us, evaluation is not “value free.” As you think about your own work or programs you know, consider the alignment of your values with the effort or program. How did you navigate that relationship? If a program’s mission did not align with your values, would you agree to serve as its evaluator? How would you approach that situation?

## REFERENCES

- Cameron, W. B. (1963). *Informal sociology: A casual introduction to sociological thinking*. New York: Random House.
- Cook, J. R. (2014). Using evaluation to effect social change: Looking through a community psychology lens. *The American Journal of Evaluation*, 36, 107–117. doi:<https://doi.org/10.1177/1098214014558504>
- Cook, J. R., Kilmer, R. P., with Tynan, J. M., Larson, J. C., Hogan, K. S., Scott, J. T. B., & Mihalski, J. L. (Contributors). (2014). Evaluation of bright beginnings: processes and outcomes: final report. Community psychology research lab, University of North Carolina at Charlotte. <https://www.cms.k12.nc.us/cmsdepartments/accountability/REA/Documents/Evaluation%20of%20Bright%20Beginnings%20Final%20Report%2011.2014.pdf>.
- Costa, A., & Kallick, B. (1993). Through the lens of a critical friend. *Educational Leadership*, 51, 49–51.
- 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
- Gadaire, A. P., Armstrong, L. M., Cook, J. R., Kilmer, R. P., Larson, J. C., Simmons, C. J., Messinger, L. G., Thiery, T., & Babb, J. (2020). Enhancing social–emotional development in pre-k: Using data to address students’ needs [Manuscript under review].
- Hogan, K. S., Tynan, J. M., Covill, V. J., Kilmer, R. P., & Cook, J. R. (2017). A capacity building framework for community–university partnerships. *Collaborations: A Journal of Community-Based Research and Practice*, 1(1), 1–28. doi:10.33596/coll.10
- Rogers, P. (2014). Week 19: Ways of framing the difference between research and evaluation. *Better evaluation: Sharing information to improve evaluation*. [http://www.betterevaluation.org/en/blog/framing\\_the\\_difference\\_between\\_research\\_and\\_evaluation](http://www.betterevaluation.org/en/blog/framing_the_difference_between_research_and_evaluation)
- Small, P. (2012). Four differences between research and program evaluation. *Blog: Nonprofit capacity building*, retrieved from <https://managementhelp.org/blogs/nonprofit-capacity-building/2012/01/08/four-differences-between-research-and-program-evaluation/> on March 26, 2018.
- U.S. Department of Health and Human Services. *Code of Federal Regulations*. 45 CFR 46.102(d), relating to the Protection of Human Subjects.
- Viswanathan, M., Ammerman, A., Eng, E., Gartlehner, G., Lohr, K. N., Griffith, D., Rhodes, D., Samuel-Hodge, C., Maty, S., Lux, L., Webb, L., Sutton, S. F., Swinson, T., Jackman, A., & Whitener, L. (2004). Community-based participatory research: Assessing the evidence. Evidence Report/Technology assessment No. 99 (prepared by RTI-University of North Carolina evidence-based practice center under contract No. 290-02-0016). AHRQ publication 04-E022-2: Agency for Healthcare Research and Quality.