

A Conversation with My Readers

I have written this book just for you: the practicing evaluator; the emerging evaluator; the “I am thinking about being an evaluator” nonevaluator; the “I do monitoring” evaluator; the “I was unceremoniously dumped into evaluation” evaluator; the “I just got out of grad school, and now I have to practice evaluation” evaluator; the “I do not do evaluations [yet]; I just need to know about evaluative thinking” nonevaluator; the researcher; and the person who needs to know how to work with evaluators.

I take each of you on a journey that demystifies evaluation, explores what it means to be an evaluator, and shares some well-kept trade secrets along the way. This is not an evaluation theory book, nor is it a research book. It is a book about being an evaluator who uses evaluation to explore, describe, explain, and eventually judge (in some way) how, and the extent to which, something does or does not work where, for whom, and why. The chapters offer various ways to engage with evaluation information and knowledge—ways that cement learning and encourage reflection. Structured guidance shows you how to untangle various evaluation situations through facilitation, negotiation, and listening. Some core aims of this book are to encourage you to experiment with different ways of thinking; consider multiple perspectives; and acknowledge and engage with the formidable roles that context, power, politics, culture, language, and values play throughout the *entire* evaluative journey.

HOW THE BOOK IS ORGANIZED

I start at the beginning, as one should when writing a book, and then build each chapter on the previous one, providing scaffolding on which to engage in any evaluative process. However, the book does not need to be read sequentially. I have written the book to support your unique evaluation adventure, which may begin in Chapter 3 or Chapter 14, then jump to the last section of Chapter 5, and then circle back to a paragraph in Chapter 2. The book

provides a structure that untangles the messiness of evaluation by informing and guiding the many choices faced, and the various discussions held, in an evaluation journey; at the same time, it allows for a realistic sense of unpredictability and differences found in each one.

The book has two parts. Part I lays the foundation for exploring any evaluative process: It breaks down and takes you through what an evaluator needs to know, whether it is an evaluative process for learning, reflection, improvement, accountability, social justice, or judgment. Part II focuses on working as an evaluator and exploring evaluation: It is aimed at fostering reflection on, and thoughtfulness and awareness of, the many kinds of evaluation that exist; discussing the many evaluative roles that can be filled; and exposing some common challenges and pitfalls often encountered in the field, but rarely described in textbooks. Fused together, Parts I and II demystify evaluation and provide a firm basis for candidly engaging with any monitoring or evaluation process, no matter what role you fulfill.

Throughout the book, I dive into the murky sea of evaluation and guide you through it all. Although I provide strategies and processes, I do not provide a “do this and then do that” model, which may work in some situations, and then not in others. Rather, I offer a way to *think* through any evaluative process so that you can comfortably engage with a peer, colleague, boss, beneficiary, or client in any evaluation situation. The book will support your work in almost any evaluation context—whether you work for a nonprofit, a community-based organization, a donor, a government, a university, an institute, or a foundation, or for some other group that aims to fix, change, influence, or in some way make something better in the social world.

THE LEARNING APPROACH

People learn in different ways and at different speeds. A concept that seems easy to grasp for you may be a stumbling block for others. This is true in most of life. Some people find it harder than others to tell time by using an analog clock, or instantaneously to know their left from their right. Often there is something in our daily lives that we find more challenging to do or understand than the people around us; often we do not talk about this. The same is true in evaluation. With this recognition, I draw on different ways of learning, including facilitated interactive activities, self-learning exercises, areas for reflection, sections for discussion, and practical applications. Furthermore, we all have lives outside of practicing evaluation. Some days I find I have a few minutes where I can quickly watch or listen to something, and some evenings I find an hour or two to read. Once in a blue moon, I find more focused time. Acknowledging that reality, I provide further ways to learn about each chapter’s topic that meet your varying needs—carefully balancing how much more you want to learn with how much time you have to learn more.

Remember Aesop’s fable about the Tortoise and the Hare? The overconfident, sleek, fancy-looking Hare bragged about how fast he could run. Tired of hearing him boast, the Tortoise challenged him to a race. The race was long and challenging, and the Hare sprinted

ahead, while the Tortoise approached the race in a precise and methodical way. All the animals in the forest gathered to watch the race, which was won by the wise Tortoise. In the book, we (you and I) are Tortoises. We do not rush to the end to assess and judge; rather, we thoughtfully engage and delight in each step of the evaluative process.

MY PERSPECTIVE

Making sense of things in the social world is not an easy task for anyone. Being asked to judge something, and in doing so to value that thing, can be daunting. There is no perfect template. An evaluator needs to engage in the academic theory and the practical side of evaluation; it is not an either–or decision. Theory informs practice, and practice informs theory. An evaluator educates when appropriate, to ensure engagement, learning, and meaning in the process and in its results. An evaluator comfortably negotiates where necessary, and acknowledges and engages with culture, language, power, politics, and values that constantly swirl in the evaluation process and influence all decisions, including her own. (See “My Use of Pronouns,” below.) Through demystifying the process, an evaluator invites people into it and enables them to join in; he actively takes that responsibility, and in doing so through evaluation makes the world a bit better off, even if just a tiny bit.

WHY I WROTE THE BOOK

The reason I am an evaluator is not that I enjoy judging things or being critical. I am an evaluator because I believe that evaluators have the potential to make the world a better place to live in. I have written this book to provide a way of thinking that supports an evaluator—not to intimidate others with knowledge, but to engage and guide them through a transparent process.

All over the world, especially where funds dedicated to social improvement are finite, thoughtful, kind, knowledgeable, and skillful evaluators can offer appropriate evaluative processes that inform decisions that affect people’s lives, animals’ lives, and the environment. Capable evaluators can fulfill an important societal role, while incompetent ones can be a detriment.

SOME GUIDING ICONS

In an evaluation report, it is very common to have a page that lists all the abbreviations used in the report, along with their full terms; this provides a quick guide that you can refer to when reading the report. In this book, I use eight icons to draw your attention to how various features support you in more easily engaging with the text. The icon list follows.

	Activities or things to do
	Advice
	Sharing a trade secret
	Suggested resources
	Clarifying questions
	
	

THE USE OF PRONOUNS

The use of pronouns is a funny thing. In older books, there is a constant use of *he*. Being a *she* myself, I find that a bit bothersome. On the other hand, when I am reading, it is not always a smooth experience to read *he/she*. For some reason, when I read *he/she*, my brain always reacts with “Wait, so which is it? A he or a she?” This makes me stop reading the sentence, decide if it is a he or she, and then reread the sentence with my chosen pronoun. So I decided to alternate between using *he* and *she* throughout the book. But this then got me thinking: “What about people who do not identify with those pronouns? Perhaps I should use the pronoun *they*?” However, the use of *they* as a singular pronoun is not yet that common, and some early reviewers of the book kept thinking I was incorrectly using the plural *they*. So a note to my readers is that I use *he* and *she* alternately throughout the book, sprinkled with a few uses of *they*, to refer to a person.

Furthermore, I often switch in the book to using the term *we* instead of *you*, when you are technically not part of the *we* conversation just yet. I often find that using the term *you* can be heard as too direct, too confrontational, or too “othering.” That is, it can be perceived as an attack that puts people on the defensive or makes them feel isolated. Of course, there are times when only *you* is appropriate, or indeed is the only option. However, using the term *we* underscores that we (you and I) are on the evaluation journey together.

CHAPTER 9

Assessing and Evaluating Progress



Two questions I am often asked when brought into an evaluative process are “How do we know whether there is progress?” and the often-related query “What should our indicators be?” Chapters 3–8 have smoothed the way for a captivating and informative discussion of what someone needs to know to make decisions on how and when to assess for progress. To understand which method of inquiry to use, read Chapter 5. To understand what to assess and when, keep reading this chapter.



I have a few trade secrets to share about these conversations. Well, to be perfectly honest, I have sort of a few trade secrets *and* advice combined. No one knows automatically exactly what to assess; a perfect indicator does not exist; and indicators are not always necessary. Phew. So the secrets are out. When someone asks a question regarding how to assess if an intervention is moving toward what it aims to achieve, reframe the conversation. Do not be pressured to give an automatic response, or think that you should be able to do so. To have an informed discussion about how to assess for change, influence, or progress, the program theory (including the problem statement) and the program and its logic (including intended results, how the intervention is being implemented, and the social, political, and cultural contexts) need to be clearly laid out. If this has not been done, reframe the conversations around the discussions and processes in Chapters 3–8. If the conversations raised in those chapters are clearly understood, then it is time to have the “How do we assess for progress?” discussion.

Before we have the conversation, let’s clarify the difference between measurement and assessment.

MEASUREMENT AND ASSESSMENT: THE DIFFERENCES

In Chapter 1, I have talked about how some terms have specific meanings (e.g., *causal inference*) and others have varying definitions (e.g., *evaluation*); in these latter instances, we need to negotiate their meaning. **Measurement** and **assessment** are tricky words, as they have specific meanings that are often used rather loosely in the field. Thus if someone asks you to measure something, or to assess it, stop for a moment and clarify what they mean. It is useful to distinguish between these two words, to encourage a clear understanding of what is expected in the evaluation process. Let's examine the two words a little more closely.

We use quantitative data to measure something. We measure how tall we are with a measuring tape, and how heavy we are with a scale. *Measurement* gathers data that simply quantify *what is*. So when a client asks you to measure, he is telling you that he wants information described with numbers. Or maybe not: Because the word *measure* is used commonly in everyday conversations, he could also be asking you to assess something qualitatively—so it may be worth taking the time to ask him what he wants. Measurement can be objective (e.g., determining how well someone performs on a test) or subjective (e.g., determining what people report feeling or experiencing). Measurement refers to something that is quantified (Huitt, Hummel, & Kaeck, 2001; Stake, 1995).

This leads us to the word *assessment*. When an evaluator assesses something, she is gathering evidence, which can be qualitative or quantitative. Thus assessing an intervention means that the evaluator is collecting data that will be used to understand it better. Some people use the term *assessment* to mean that there is a comparison against a norm or standard; however, that is not always the case. For example, a baseline assessment gathers evidence that is used as a reference point *to* assess for any change; the baseline establishes a “base” for comparing the situation before and after an intervention, which is then used to assess the intervention's effectiveness. (For more on baseline assessment, also called a *baseline study* or just *baseline*, see Chapter 5, pages 106–107.)

While we are here, let's toss back in the term *evaluation*. Although different people define evaluation in different ways, its core definition is agreed upon: Evaluation is a systematic, transparent process that values something, and it is political. To evaluate is to determine the worth, merit, and significance of something. (To refresh your memory, return to Chapters 1 and 2.)

So if we say we measure something, it is about numbers. I get that. Then when I do a qualitative evaluation, I am assessing something and valuing it, but I am not measuring anything. Is that right?

In qualitative inquiry, we do not measure because we do not collect numbers. We are reporting what happens in a program, what we observe, or what people say. When an evaluator is



asked a question that is suitable for qualitative inquiry, he constructs an approach to answer that question by collecting words from people through interviews, words from documents, and observations (i.e., he gathers evidence). These are not numbers; they are words in the form of descriptions, observations, and stories. To make sense of these words (which are also data), he interprets them by identifying patterns and themes; he aims to understand their meaning. He uses these data, and his understanding of the data, as evidence to assess the intervention against the evaluation question, or even against standards and norms; he just does not assess it against a set number. He could also use a mixed methods approach—gathering data to measure by means of numbers, and then using those quantified data to assess the intervention. So assessment can involve either qualitative or quantitative data.

If we had all the resources in the world, everything could be assessed. However, most likely, limited resources exist. When assessing for progress, a program manager, donor, or evaluator is interested in finding out whether an intervention is moving toward its intended results or away from them. In monitoring, the most popular way to assess if something is moving toward its intended results or not is to use an indicator. More recently on the evaluation scene are *progress markers*. Outcome mapping (which uses progress markers) and most significant change (MSC) do not use numbers to monitor for change; they assess change through gathering words and understanding their meaning. These qualitative approaches are described later in the chapter.

Again, indicators are quantitative measures. They also do not measure every single tiny part of an intervention. Indicators provide *enough* data on certain items to let us know if the intervention is moving toward its intended results, moving away from them, or standing still. An indicator can have many synonyms, such as clue, signal, or sign. That is a very brief explanation, and we talk more about indicators later in the chapter. However, before we talk about indicators in depth, we need to be clear about how to choose *what* is assessed.

WHAT TO ASSESS

Three kinds of factors come into play in evaluating progress: (1) *practical* factors, (2) *use* factors, and (3) *technical* factors. At times, these types of factors can overlap (e.g., what is practical is useful, or what is useful is technical). Although it is critical that an evaluator consider all three sets of factors when deciding what should be assessed, they rarely all have equal weight. In some cases, one practical factor may outweigh all the use and technical factors combined. Sometimes the practical and use factors “team up” together against the technical factors; at other times, technical factors become the most important. In an ideal world, all factors would inform a decision equally, but that is not how it often happens in the real world. Let’s look at how practical and use factors can influence decisions on what to assess, and then have a longer discussion of the technical ones.

Practical Factors

There are many different practical factors to consider in deciding what to assess. These factors can include *resources*. For instance, there may not be enough staff to collect all the data that are needed to assess a result. Or certain data may require a statistical software package that is too expensive to purchase, or necessitate a skill set that no current staff member possesses. Practical factors can also include *timing*. For instance, the most useful data may be secondary data captured by another organization; however, these data are collected far less frequently than is ideal for monitoring. Another set of practical factors may involve *accountability*. If a program is committed to assessing something in exchange for funding, then it needs to be assessed, even if the resulting information is not useful to inform program management decisions. Practical factors can include *politics* as well: What needs to be assessed may be what external people or groups expect to see. Finally, practical factors can include *ethics*. For instance, it might be useful for the local health clinic and educators to know how many children in the community school have HIV, so that appropriate care can be given to them; however, it would not be very ethical to test them.

Use Factors

If those are practical factors, what are the use factors? A basic premise for deciding what to assess is to assess what is useful to whoever will use the data. However, the use factor is not as straightforward as it first appears; use can be a funny thing. Managers use data to manage their programs—to identify what is going wrong, what is going right, and what needs immediate attention. However, there are other uses besides informing management decisions. Sometimes, the usefulness in collecting data is to provide information to whoever funds the program; it is for accountability, perhaps, or maybe it is just what evaluators refer to as *check box* or *tick box* use (for more on check box/tick box evaluation, see page 230) that keeps funding flowing.

Technical Factors

Last but not least, there are the technical factors that influence what to assess. The technical factors often overlap with the use factors, and sometimes overlap with the practical ones; when they all overlap, it is like winning the lottery. The technical factors are discussed after I share a melancholy trade secret.

The “sad but true” trade secret is that sometimes in the real world, when a more powerful someone tells you what and how to assess, that is often the end of the discussion—even when what is to be assessed, or how, does not make sense or is not useful to those who implement, need to manage, or benefit from the intervention. Recognizing and engaging with the fact that this happens can help keep evaluators (and those who manage a program) sane. In *monitoring*,



one solution is to collect (additional) data that are relevant and useful (if possible) to those who need to manage, or otherwise make decisions about, the intervention. In an evaluation, when not given a choice about what and how to assess, the evaluator needs to explain the decision-making process in the methodology section (more specifically in the limitations section) of the report, at a minimum. It is important to recognize that power and politics influence every part of the evaluative process, including what, how, and when to assess. (Power and politics in evaluation are covered in Part II, especially Chapter 13.)

Technical factors in decisions about what to assess require that an evaluator or program manager have a thorough understanding of the problem–intervention–results links (the program logic) and the theory of change. The “what to assess” decisions focus on the intervention (what is done—the action) and the results (what happens—the result). A useful way to frame the discussion is to use a modification of the concept of so-called “key” and “killer” assumptions, taken from the logical framework approach (United States Agency for International Development [USAID], 1980). I find “killer” to be too abrasive, and the “key” part to be unnecessary, so I have slightly changed the name and the process. Briefly, because of the nature of an assumption (it is not a fact), an evaluator needs to assess whether an assumption has become a fact through the implementation of the intervention. One way to sort out what and how to assess that is to break the discussion into two categories of assumptions: assumptions around action, and assumptions around anticipated results. Deciding what to assess is connected to when to assess. An evaluator must first check the assumptions about action (i.e., make sure that whatever was supposed to be done was done), and then assess for anticipated results in their anticipated order (i.e., make sure that what was supposed to happen happened, when it was expected to do so).

Assumptions about Action

Identifying action (i.e., whether something has physically happened) is logical, although a step can easily be missed (e.g., by assuming, “Of course it happened”). Here are two key questions for addressing assumptions about action:

1. Was the intervention implemented? (Do not assume that actions happen.)
2. Was it implemented as planned? (If so, this is often called **fidelity**.)

In terms of action, knowing what to assess can be drawn from the implementation theory and program logic, which explain what needs to be done, when, and where to achieve the intended results. If the intervention or program was *not* implemented with fidelity, this does not automatically suggest a negative finding. For example, perhaps changes were made to the initial program design that ultimately contributed to the intervention achieving its results. Knowing that the intervention was not faithful to the original design, knowing what

changes were made (no matter how small), and discovering that the intervention did achieve its results are useful findings (if we think to assess them). At the same time, knowing if an intervention was implemented with fidelity is a critical part of some evaluation designs, such as randomized controlled trials (RCTs). An RCT is

a way of doing impact evaluation in which the population receiving the programme . . . is chosen at random from the eligible population, and a control group is also chosen at random from the same eligible population. It tests the extent to which specific, planned impacts are being achieved. . . . The strength of an RCT is that it . . . [helps] evaluators and programme implementers to know that what is being achieved is as a result of the intervention and not anything else. (White et al., 2014, p. 1)

For that reason, RCTs require an intervention to be implemented with fidelity, so that results from different (randomized) groups can be compared. Thus assessing the action of an intervention is critical both to managing a program (e.g., informing changes along the way) and to evaluating one.

Assumptions about Results

The second category of assumptions consists of the results at all levels. (See Chapter 7 for examples of different levels of results, and of how results link to each other.) Now, because an action happened (it is confirmed), the next decision is to choose what anticipated result(s) to assess and when. Common sense tells us that according to our program logic, one result must happen before the next one, and sometimes several results need to happen before the next one (i.e., no result should be assessed or evaluated before its time). For instance, people need to gain knowledge and skills before they can apply these. Technically speaking, all results should be assessed, yet practical factors (e.g., it is too expensive) prevent that from happening.

Action and Results Assumptions

In this section, I lump action and results assumptions together, and explain how to determine what is likely to be more important to assess than something else (from a technical perspective).

These assumptions (action and results) can be divided into two groups. I call these the “do not worry too much about me” assumptions, and the “destroyer” assumptions (i.e., the ones that have a lot of risk associated with them). The “do not worry too much about me” assumptions are (1) those in which, if something goes wrong, it is not likely to destroy the program and (2) those that are highly unlikely to be wrong. For example, it is safe to assume that people will act in a civilized manner when attending a training

on evaluation; the chances that the participants will cause a riot are very slim. Simply stated, “destroyer” assumptions are ones that are critical to the intervention, that may well be wrong, and that are more than likely to result in the intervention’s failure if they are wrong. These are the assumptions on which data should be formally collected, at set points in time.

For example, in our girls’ education program, we assume that fathers who do not support their daughters’ school attendance would be willing to attend a peer intervention activity (action assumption). That is a destroyer assumption. Why? If the fathers do not attend the peer-to-peer session, the intervention will fall apart. Directly addressing fathers is a critical part of the intervention theory and logic. Collecting data on whether fathers participate in the peer-to-peer sessions is thus critical. If the implementer knows that fathers do not attend, then there is the option to address that challenge (e.g., by meeting one-on-one with fathers). Or perhaps fathers *do* attend the peer-to-peer sessions, but the peer-to-peer intervention does not change the fathers’ perceptions. In this instance, the intervention is also likely to fail, as changing fathers’ perceptions about girls’ education forms a large part of the intervention theory. Here the theory will need to be revisited, along with a closer look at how the peer-to-peer activity was implemented (i.e., was the theory wrong, or was the peer-to-peer activity poorly implemented?). The extent to which the peer-to-peer interaction changes fathers’ perceptions is a critical assumption to be assessed. These are some examples of how, if an assumption is not assessed and is a wrong assumption, it can destroy a program. In other words, there are three items to assess: (1) whether the fathers attend, (2) what the quality of the peer-to-peer intervention is, and (3) whether the fathers’ attitudes change. All three are based on destroyer assumptions.

Applying Destroyer Assumptions to the Rest of the Evaluative Process

Before we leave the discussion on destroyer assumptions, let’s consider how such assumptions apply to *all* aspects of the evaluative process. Can the evaluators safely assume who benefits? What about assuming who values what, and how? What about assumptions made with regard to politics, culture, or the economy? Which political, cultural, social, economic, and other contextual assumptions should be monitored closely, and which ones can we pay attention to through more informal means (e.g., hallway chatter, newspaper articles, Twitter)? For all assumptions in every part of the evaluative process, there are some that will immediately present themselves as destroyer assumptions. It may seem as though *all* the assumptions are destroyers, and though that may be the case, it is more likely that several will “bubble to the top” through a facilitated discussion on them. Once more, the ultimate decision on what to assess is influenced by each intervention’s particular mix of practical reasons (which includes ethical, political, and accountability ones), choices about use, and technical decisions.



The Use of Informal Information

Informal data collection takes place almost constantly. Newspapers, radio shows, Twitter, hallway discussions, an email or a phone call, or simply observing something that suggests the intervention is not working (or is working in unexpected ways) should not be ignored. Simply because we choose not to formally assess something does not mean ignoring data that suggest (or even whisper about) a problem or an unexpected achievement. Furthermore, if we do decide to assess something, and our formal data tell us one thing and the informal data tell us another, the informal data should be considered.

The sections starting on page 177 have clarified three ways to make decisions regarding what to assess: considerations of practical factors, use factors, and technical factors, some of which may overlap. Once it is decided what to assess and when, decisions need to be made about how to assess. One of the most popular ways to assess progress is to use indicators.

HOW TO MEASURE PROGRESS

What Is an Indicator?

Now that we have decided what to assess to see if the intervention is making progress, and when to assess it, let's talk about a common concept: the illustrious **indicator**. Indicators are sometimes called *performance indicators* (as they measure performance), are sometimes labeled *key performance indicators*, and are sometimes given other names with slightly nuanced definitions. Whatever label is used, indicators are one of several ways to measure an intervention's progress (or lack thereof). So what is an indicator? Indicators are just what the word suggests: indications, signs, signals, ideas, pointers, or gauges that an intervention is moving toward or away from its intended results, or simply standing still in its quest to achieve these results. An indicator is not a result; it is an indication of a result. Think about some common indicators in your own life. When your stomach growls, it is an indication, sign, or signal that you are hungry. Or maybe not: It might be indicating that you ate bad food, or that you feel sick. However, a growling stomach does not definitively mean that you are hungry, or sick, or ate bad food. It merely suggests what could be happening. However, an indicator at the lower level is often very much like a result, but not quite the result itself (it is a closely related measure of achievement); the indicator makes the result measurable. Table 9.1 shows three illustrations drawn from the girls' education example.

There are several other labels that often have the same meaning as *indicator* (in addition to *sign*, *signal*, *gauge*, and *clue*), such as *measure* or *metric*. At this point in the book, multiple labels for the same concept should come as no surprise. However, sometimes a *metric* is only used when it refers to a specific number, which is also called a *target*. (A later section of this chapter provides a further description of targets and benchmarks.) The often-repeated advice in this book applies here as well: Clarify terms, labels, and definitions before embark-

TABLE 9.1. The Indicator: Making Results Measurable

	The very first thing that happens	What happens next?	What happens after that?
Result	Fathers who do not want their daughters to attend school come to the peer-to-peer session to meet with fathers who support girls' education.	Fathers learn reasons for educating girls.	Fathers send their girls to school.
Indicator	Number of fathers who attend X number of peer-to-peer sessions.	Number of fathers who can articulate the importance of educating their girl child.	Number of girls previously kept from school by their fathers who are registered to attend school.

ing on a discussion about how to implement the concept. (See Chapter 1 for a discussion on the importance of language in evaluation.)



Indicators are quantitative? My donor is asking for qualitative indicators; what are those?

Ah, the old qualitative indicator. It's a bit of a contradiction, really. I do engage with qualitative indicators in this chapter because I know that donors and governments, among others, ask for them, and they are used in the real world. An example could be "quality of health care as perceived by teenagers." Essentially, a qualitative indicator is one for which data are collected via qualitative inquiry approaches, and then the data are quantified. If a donor or manager asks for qualitative indicators, ask him for what purpose these data should be collected. If it is to provide deeper insight, introduce the MSC or outcome mapping's *progress markers* (both discussed later in this chapter) as optional choices.

Here is an example of what you might find in the field. A health organization wants to measure the quality of its service at local health clinics. To measure quality, the evaluator develops measurement items, such as collecting data on how long a patient waits to see a doctor, and how patients rate the friendliness of the nurses on a scale of 1–5. These quantifiable data then provide indicators of the service quality at the health clinic. However, the evaluator could also collect stories from the patients and assess those stories to better understand the quality of care at the clinic. Since you asked, let's delve a bit deeper into the distinction between qualitative and quantitative indicators.

A Discussion of Qualitative and Quantitative Indicators

Quantitative indicators are very straightforward: The data are collected with a closed-ended approach and often provide a simple number, percentage, or ratio. One exception would be with policies, where moving from a *green paper* (initial thoughts) to a *white paper* (final paper) would be an indicator that a bill is moving toward being approved. Qualitative indicators—

well, I would say there are no such things, but as I just noted, I have seen them with my own eyes. When these are requested, providing them can be challenging because qualitative data need to be quantified as suggested on the previous page. Look at the following example of indicators that draw from qualitative data:

- Number of people who report feeling good.
- Number of people who report feeling bad.

An open-ended question can be used to gather these data. For example, “How are you feeling today?” would be an open-ended question to gather data for these two indicators. The responses could include “good,” “great,” “awesome,” “very good,” “super,” “not so well,” “terrible,” and “awful.” These data are then analyzed (see Chapter 5 on qualitative data) into themes of “good” and “bad.” The indicator data are as follows: Five people reported feeling good, and three people reported feeling bad. In rare cases, qualitative indicator data stay in words and are not quantified; as such, they do not fit the true definition of an indicator, but they exist. Indicators are intended to measure something and provide a quick glance that shows if an intervention is on track or not.

A final, general note on indicators: Indicators are often used in the politics of making certain groups visible or making them disappear. *Disaggregating* data (meaning separating it) by sex, age, marital status, and other characteristics that make a difference is extremely important. It is important because if data are not separated, their aggregation can mask trends that are different. (See more on these kinds of discussions in the Resources on page 196.)

BOX 9.1. Progress Markers

A **progress marker** is a bit different from an indicator. Used in outcome mapping (discussed in Chapter 15; see also the Resources feature at the end of this chapter), progress markers are qualitative data collected to mark progress and show a result. A progress marker must be descriptive and therefore qualitative in nature. An example for the result “More participation by disabled persons at community meetings” would be collecting data by observing disabled people attending, speaking at, and being listened to at the meetings. In contrast, an indicator could be “Number of disabled people at the community meeting,” which would *signal* that there is more participation by disabled people at the community meetings. Progress markers are always qualitative, and show results through describing what happens through progression (Earl et al., 2001). With some programs, it is sometimes useful to combine the two approaches (indicators and progress markers) in a hybrid approach, and perhaps even to add a third, MSC, for a three-way hybrid. Again, MSC is discussed in more detail later in this chapter.

A Good Indicator Is a Useful Indicator

A useful indicator is one that provides credible data (for more on credible data, see Chapter 5) to inform a management decision. Based on the data collected, the decision might be “Everything looks great; let’s keep moving,” or “Something is not right here; let’s check it out.” Different organizations use different ways of examining indicators to see if they are “good” indicators, so it is important to find out what those criteria are. However, general criteria for indicators include that they are *reliable*, *accurate*, *useful*, *feasible*, and *timely*. Here are some examples of how these words are used:

- *Reliable*: If different people were to use the data collection tool, would they get the same result? *Reliable* doesn’t mean *right* (e.g., a broken scale could consistently tell you that you weigh 200 pounds when you actually weigh 150).
- *Accurate*: Does the indicator give the true value (e.g., you do indeed weigh 150 pounds and the scale shows that you do weigh 150 pounds)?
- *Useful*: Does the information gleaned from the indicator help in managing the program, or is it collected and just put to the side?
- *Feasible*: Can the data be collected with the available human and financial resources, or is the collection process too complicated or too expensive?
- *Timely*: Can the data be collected when they are needed? *Timely* also refers to having the data when decisions need to get made.

You may hear the acronym SMART when people talk about critiquing indicators. The acronym stands for Specific, Measurable, Assignable, Realistic, and Time-bound, though sometimes the letters in the acronym are used differently. SMART is often associated with Peter Drucker (1995, 2007), an organizational theorist who popularized it in connection with his “management by objectives” approach. It is a somewhat tired acronym. Here is a more basic and quite sensible way to determine how “good” an indicator is. Test it and, after gathering the data, ask these three questions:

1. Were the data collected with a reasonable amount of resources?
2. Did the indicator measure what it was supposed to measure?
3. Were the data used?

Common Challenges in Using Indicators

Here are nine challenges evaluators often encounter when using indicators to measure progress, and some advice on how to engage with each challenge.

- *Challenge 1.* The program only collects indicator data. The larger picture, and the other program components not measured through indicators, are often forgotten. This reliance on only a few indicators leaves big gaps in understanding what is happening, and creates the potential for things to be missed that may influence the achievement of results.

Engage with all available information. Consider introducing multiple approaches, such as progress markers or MSC. Formative evaluations by external evaluators who can bring a different view to the program might provide some additional insights. If formal changes are not acceptable, consider gathering data informally via a less structured and more holistic approach, such as at meetings or workshops.



- *Challenge 2.* Indicator data and other data sources provide contradictory information. For example, indicator data suggest that the program is moving along in the right direction; however, other data suggest multiple problems. The other data are ignored.

Again, engage with all available information. Imagine that an airplane is spiraling out of control and heading toward the ground, even though all the indicator instruments on the control panel say everything is fine. Indicators are just that—indications of something happening or not happening—and sometimes they are wrong. If other signs or signals tell you a different story from the one told by your indicator data, consider them carefully.



- *Challenge 3.* There is no agreement on what the indicators should be. Arguments ensue about identifying indicators, with little knowledge of the program logic or program theory.

Find common ground. Avoid discussions of indicators until the program logic and underpinning theory are clear and agreed upon. Even after the theory and logic are clear, arguments may still ensue. Consider using a framework (e.g., SMART, the organization's own criteria, or ones the group agrees on) to assess the indicators, so that the discussion is organized and clear. If there is still no agreement, consider bringing in one or more experts in the field.



- *Challenge 4.* The indicator is useful to the donor, government, foundation, or other funding sources, but not to the implementer. Often funding to an organization or government department comes with indicators attached. Sometimes these indicators do not adequately (or sometimes even closely) measure what the organization or department is doing. For example, an organization accepts money based on an agreement on the grand problem (e.g., HIV/AIDS, global climate change, human rights). Later, the implementing organization realizes that the mandatory indicator(s) do not even remotely measure the organization's intervention, which addresses vastly different pocket problems.

Explore the options. In this situation, there are at least four options. The first is to give back the money or stop accepting the funds (painful choices, but options just the same). The second is to negotiate a change in the indicators through a facilitated discussion between the implementer



and the funder, which explores the intervention, the pocket problems it addresses, and its intended results. The third option is to change the intervention. The fourth option is for the organization to meet its contractual obligation (providing data on a required indicator) and then collect additional data to manage the program.

- *Challenge 5.* Indicators state, “Change in . . . ,” but there is no baseline. Whenever there is an indicator that begins with “Increase in . . . ,” or “Decrease in . . . ,” or “Change in . . . ,” baseline data are needed. To measure any change, an evaluator needs to know this: A change from what?



Again, explore the options. There are several strategies. First, consider that there might be baseline data existing within, or outside, the intervention. See Chapter 5 for how to identify an existing baseline. If there are no documented baseline data, consider collecting retrospective data, as also described in Chapter 5. A third option is to remove the indicator and replace it with one that is feasible to measure.

- *Challenge 6.* A qualitative indicator is mandatory. It is not clear why qualitative indicators are needed; however, it is mandatory that they be collected.



Consider who wants qualitative indicators, and for what reason. It is possible to quantify qualitative data (as discussed earlier in this chapter). Find out why these data are being requested. Understanding why they are requested will bring insight into what the donor, government official, or whoever mandated collecting them wants to know. Once it is clear what this person or persons want to know, alternatives might be possible. For example, MSC could provide a viable option.

- *Challenge 7.* The intervention has started and is well underway, and then a new person to whom the program manager is accountable (e.g., the donor, a government department official) changes the indicators. Sometimes, when new people with decision-making authority join an organization or take on a new position, they change the indicators, or refine their definitions.



Identify the reasons for the change. There could be very good reasons (e.g., the new person realized that different indicators, or refined definitions, would be more useful, or the funding stream has shifted and necessitated the reporting of different data for accountability reasons). On the other hand, perhaps the new person did not like the old indicators for some reason, and just decided to change or refine them. Regardless of the reason, it is important to take stock and understand the impact that changing one or several indicators (or their definitions) will have on data collection and other resources. Furthermore, it is critical to understand the extent to which these new indicators are useful to program management, and to consider different options if they are not useful. For instance, is it possible to continue collecting data on the old indicators (i.e., are there enough resources)? While the new indicators may not be as useful as the old ones (to the program manager), are they useful “enough” to enable the implementers to continue managing the program effectively?

- *Challenge 8.* The intervention changes, and the current indicators are no longer appropriate.

We live in a rapidly changing world. Problems shift, interventions shift, and intended results may shift. Whenever something shifts, it is critical to reexamine indicators to ensure that these signs or signals are still relevant to the intervention. Sometimes indicators cannot be changed, even when they are no longer relevant. For instance, an intervention's funding may be based on certain accountability measures; the indicators are sealed in stone. In such a situation, document that the indicators are no longer relevant, and give clear reasons (e.g., the intervention is doing something different because the problem being addressed has changed). Provide these reasons to the person(s) to whom you are accountable, and encourage them to meet with you to discuss changing the indicators. Recognize, however, that sometimes indicators cannot be changed for any reason during a funding cycle; in that case, roll your eyes upward (make sure no one is watching) and collect those data.



- *Challenge 9.* There are 394 indicators (or some other huge number) for one intervention. Unless it is an extremely large and enormously well-resourced intervention, it is unlikely that collecting data on all of these indicators will be useful to manage an intervention.

It is highly unlikely that one person, or even one group of people, can realistically monitor 394 indicators. Those are a lot of signals or signposts to keep an eye on. Thus consider the extent to which each one is necessary. To do so, consider the following process. Identify which indicators are mandatory. The mandatory ones are likely to be related to funding (which could be related to political reasons) and need to be set aside in the “keep” column. Then review the ones that are left. Group them as, for example, quality indicators, process indicators, and results indicators. You could even break down the results indicators into output, outcome, and impact indicators. Then decide how many in each group are “enough” that are feasible to collect and provide adequate signals that something is going right (or going wrong). How many are enough? It may take a few trial-and-error efforts to identify the “magic” number, but experience is often the best guide. As noted earlier in this chapter, consider what it means not to collect certain data, or not to disaggregate it; if this is not done, what will remain hidden or not known? Find someone who has knowledge, experience, and familiarity with the intervention (or a similar one), and ask for advice.



If that is an indicator, what is a target?

A *target* is something an intervention aims to achieve, or (to keep the target metaphor) something to shoot for and aim to hit. It is almost always a number. The reason to keep discussions about targets and indicators separate is that they are truly different discussions. Imagine the following discussion. One person says, “I think the intervention should measure if 75 girls attend school in the fall semester.” Someone else says, “Well, I do not agree with that.” My question would be this: “You do not agree with what? The indicator [number of girls attending school] or the timing [in the fall semester] or the target [75]?” Discussing the indicator, the timing, and the target will call for entirely different conversations. Table 9.2 shows four examples of indicators and their targets.



TABLE 9.2. Different Kinds of Indicators and Their Targets

Indicator	Target
Number of fathers who attend peer-to-peer sessions and stay longer than 30 minutes.	By the end of the first year of the intervention, 75 fathers attend peer-to-peer sessions and stay longer than 30 minutes.
Percentage of girls between the ages of 14 and 16 in the community who attend school.	75% of girls between the ages of 14 and 16 in the community attend school.
Percentage of girls who achieve at least 95% attendance in school.	90% of girls maintain at least 95% attendance in school in the first term of the year following the intervention.
Percent change in the number of fathers who regard girls' attending school as important.	25% more of the fathers regard girls' attending school as important.

Keeping conversations streamlined will support a clear discussion, help everyone to reach an understanding, and eventually (as needed) foster agreement. Furthermore, targets can change while the indicator remains the same. An interesting demonstration of how they relate to each other and to the entire evaluative process is provided in Chapter 10 through an interactive game (which is also an activity and a process).



If that is a target, what's a benchmark?

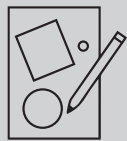
These words are sometimes used interchangeably. The best advice is to ask whomever you are speaking with how they use this term. In general, a **target** is a specific number. We could aim to get 75 women (a target) to attend the university's agriculture course in the fall semester, and then 50 women (another target) in the spring, and so on. **Benchmarks** are normally taken from the literature or from another similar intervention in a different site, or even from the same intervention in a previous year; they provide numbers or statements with which to compare the intervention's progress. Or, in some interventions in health or economic interventions, for example, some benchmarks are calculated according to formulas. Some of you may be reading my example and thinking, "Wait, that is not how my organization defines those terms." Heavy sigh. Be clear on how the terms *target* and *benchmark* are being applied in your situation.

Choosing to use qualitative or quantitative indicators, progress markers, or some other form of assessment that gauges progress depends on what types of data will provide the most useful, feasible, accurate, cost-effective information for learning, improving and/or judging the program, and making management decisions. The needs for these types of data are then balanced against cost, time, and the skill sets and knowledge required to collect and use the data. Mixed into those decisions are considerations of what data are most credible and needed by those to whom the evaluator (or program manager) is held accountable. It is most

often the evaluator's role to facilitate a discussion and help the program manager (or the person who needs to engage with the decision) make the best decision possible, with what is known at that point in time.

Critiquing Indicators

No indicator is perfect, and most indicators can be, for lack of a better description, ripped to shreds. An evaluator needs to be comfortable interrogating indicators and, at times, supporting a process that interrogates indicators. Facilitating the Party Game (Activity 9.1) introduces the concept of an indicator and the different kinds of indicators, and it shows how to develop indicators and how to tear them apart; in so doing, it demonstrates that no indicator is perfect. The Party Game, with a graphic for the invitation, is available on this book's companion website (www.guilford.com/podematerials). Let's have a party!



ACTIVITY 9.1. The Party Game: Developing Indicators and Spotting Their Weaknesses

Purpose: To become comfortable with indicator development, use, and refinement through learning how to develop, write, and critique one.

Materials: You will need one or two copies of the invitation in Figure 9.1 for each group, printed on one piece of paper apiece. (See the invitation on the companion website.) Each group needs blank paper (it can be scrap paper) to write their answers, and a pen or marker with which to write. As the facilitator, you will need a flip chart and marker to note critical points.

Time: Groups need approximately 10–15 minutes to construct their answers. The sharing of the indicators, and the learning process, take approximately 20–25 minutes. The total approximate time is 30–40 minutes, depending on the number of groups, the interaction, and the discussion of each question.

The exercise: To play the game, divide the participants into equal groups of no more than four people per group. If the same people have been sitting together in previous activities, mix them up. Provide

each group with one to two copies of the party invitation and blank paper on which to write their responses. Read the introduction on the party invitation, and ask each group to provide one indicator for each question.

There are two rules. First, an indicator cannot be repeated. For example, a group cannot have the same indicator for the third and fourth questions. Second, each answer must be written like an indicator.

Some groups might need an explanation of a proxy indicator. A *proxy indicator* is an indirect indicator; it is used when a result cannot be directly measured. For example, to know how many people come to the party, a person cannot count the number of people (this is a direct indicator). A proxy indicator could be the number of coats hanging in the closet or the number of shoes at the door.

Give the groups about 10–15 minutes to develop one indicator per question. When they are done, they should have developed six indicators. Have the members of one group read out their indicator for the first question only. Ask if anyone has the same or a similar answer. Then allow the other groups to critique that indicator. Use critical facilitation questions such as these:

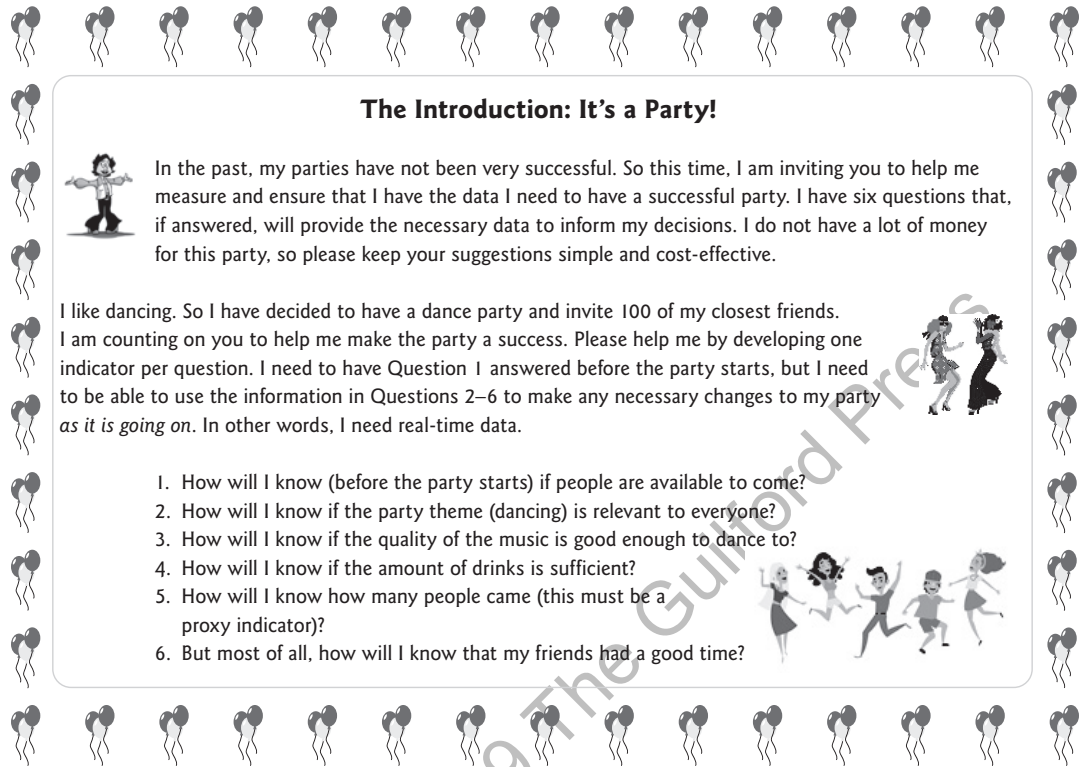


FIGURE 9.1. Party invitation for use in the Party Game (Activity 9.1).

- What else might that indicator be measuring?
- How might that be difficult to measure?
- How realistic/feasible is that?
- What could go wrong with that indicator?

Continue until all six indicators are discussed.

The game requires extremely good facilitation skills, for two reasons. First, although it can be lots of fun, it is your job to listen carefully and draw lessons. While I have listed the main lessons, many more lessons will emerge from playing this game. Second, you need to be able to pick up the mistakes, correct them, and offer solutions, all while keeping it fun; it is a party, after all. Table 9.3 (pp. 192–193) lists each question, some potential responses to each question, and the key lesson and specific facilitation tips for each question.

Critical learning points: An indicator is an indication, sign, or signal of a phenomenon, and most can be torn to shreds. There is no such thing as a “perfect” indicator. The game has four aims. Participants should:

1. Be comfortable discussing indicators.
2. Understand that indicators have advantages and challenges.
3. Understand the role of an indicator.
4. Understand that there are different types of indicators.

The game enables participants to develop a basic knowledge and understanding of indicators, as well as of the role played by indicators in the evaluative journey.

Is there anything else we need when using an indicator, or just the indicator itself?



Sometimes an indicator seems simple—it is one sentence. However, there is often a whole lot more going on behind that indicator and its one sentence, which is needed to ensure that the indicator is useful. For instance, a program manager or evaluator needs to understand what data should be collected, when and how the data should be collected, how the indicator value will be calculated, who will analyze the data, when it needs to be reported, and so on. Also, sometimes indicators include suitcase words, which are often used for brevity (see the discussion of suitcase words in Chapter 7). The attached indicator sheet should define each such word so that it is measurable (e.g., *empowerment*, *capacity built*, *pass rate*). The indicator's details need to be clear to ensure clarity, transparency, and credible data.

Finally, government departments, donors, foundations, and nonprofits often have their own format for clarifying indicators (sometimes called an *indicator specification sheet* or an *indicator reference sheet*). This format needs to be used to ensure that the indicator meets the organization's needs. Ask the organization that you work with, or for, to provide you with its template. See the next question to understand further what happens when only the indicator statement (i.e., one short sentence) is written.

Our organization has gone through long discussions to select what are likely to be good indicators. Once the indicators are selected, what else can go wrong in measuring progress?



Quite frankly, lots can go wrong, despite having well-thought-out indicators (and the same holds true for progress markers). Let's talk through some possibilities. Knowing what can go wrong prepares an evaluator to prevent it from happening. Let's pretend that it is your job to provide evaluation support to seven environmental nonprofit organizations, each located in a different country (or county, state, or province). You are hired 3 months before the nonprofit organizations need to provide the first report to their donor (funder); this report focuses on providing data from one mandatory indicator. Let's see what could go wrong.

You decide to bring representatives of all seven organizations together in a nice conference hotel to host a 1-day facilitated dialogue with them. Each organization has the same program logic and the same donor-mandated indicator. You review the one indicator with them, which is "Number of national parks and protected areas with adequate management." The donor has not provided any guidance on the indicator or its data collection. You note that *adequate* is a suitcase word (again, see Chapter 7). So you ask the representative of each organization how that group defines *adequate*. Five of the seven different nonprofits have defined this word so that it has meaning to them and is evaluable. The representatives of the two groups who do not have a definition ask to move to another room and work together to define the suitcase word. As you are pressed for time, you agree.

You now have five nonprofits represented in the room. It is not clear how the data are being collected, or where the data are coming from, so you ask how this is happening. All

TABLE 9.3. Questions, Potential Responses, Key Lessons, and Facilitation Tips for the Party Game (Activity 9.1)

Question ^a	Some potential responses (there are many)	Key lesson and facilitation tips
1. How will I know (before the party starts) if people are available to come?	<p>Number of people who respond, <i>disaggregated</i> (i.e., broken down) by yes or no responses.</p> <p>Number of people who respond positively.</p>	<p>People want to state the data, such as “yes or no,” or “98 people.” These are not indicators; these are <i>data</i> for an indicator. It is a good learning point to distinguish data from indicators.</p> <p>People want to put in <i>methods</i> here, and will respond with “Twitter/Facebook responses,” “email responses,” or “phone calls.” These are ways to collect data on the indicator, not the indicator itself.</p> <p>People want to add in <i>targets</i>. Targets are generally numbers, as discussed in a question and answer in the chapter text (p. 187). We are not setting targets in this discussion, or at least not yet. If someone says, “75 people respond,” they have jumped one step ahead. Remove the target (actual number).</p>
2. How will I know if the party theme (dancing) is relevant to everyone?	<p>Number of people dancing.</p> <p>Number of people who say they like dancing.</p>	<p>Here the question concerns “relevance.” I want to know if the theme is important to the matter at hand (which is my party).</p> <p>So let’s look at two ways to challenge these common answers:</p> <ul style="list-style-type: none"> • Why might people <i>not</i> be dancing? • Why might people say they like dancing, when they really do not? <p>There are always alternative explanations available (often funny ones!).</p>
3. How will I know if the quality of the music is good enough to dance to?	<p>Number of people dancing.</p> <p>(The group may have already provided this answer to Question 2; if so, it is a repeat indicator, and the group needs to choose another.)</p>	<p>Here the question involves “quality.”</p> <p>Remember that each group has its own set of answers, and that the group members cannot repeat their <i>own</i> indicator. One indicator cannot measure two phenomena. Otherwise, we will not know which question is being answered (or which result being measured is being achieved).</p>
4. How will I know if the amounts of drinks are sufficient?	<p>Number of people holding each type of drink.</p>	<p>Here the question concerns “sufficiency”: Do the amounts of drinks meet everyone’s needs?</p> <p>The lesson often drawn from this discussion is about time: At what point in time is this measured? It is different</p>

	<p>Number of each type of drink available.</p> <p>Number of drinks per person.</p>	<p>if people complain that there are not enough drinks when the party is almost over, as opposed to an hour after the party has started. This then leads to the question “Hey, wait, how long is this party?”, which begins another discussion.</p> <p>When questions arise about indicators, they also begin to arise about the intervention (in this case, the party). Thus lots of conversations start out with “What is the indicator?” or “How will we measure this?” and evolve into discussions of program logic, program theory, or program action. Do not discuss the theory and logic for the party; make the point that indicators often lead into these kinds of discussions.</p>
<p>5. How will I know how many people came (this must be a proxy indicator)?</p>	<p>Here answers are often the most creative, and the easiest to attack. Some responses may be the number of coats (if it is cold), the number of umbrellas at the door (if it is raining), or the number of plates used.</p>	<p>This question requires a <i>proxy indicator</i>, which is a way to estimate when you cannot directly measure what you want to measure. So, for example, a response <i>cannot</i> be “Number of people who came through the door.” That would be a direct measure.</p> <p>How might any of the answers in the middle column not give a good indication of how many people were at the party? For example, some people may not wear a coat, or they may not take it off. Or someone may use one plate more than once, while another person will take five plates.</p> <p>The key lesson here is that we do tend to rely on proxies, yet look how easy they are to pick to apart!</p>
<p>6. But most of all, how will I know that my friends had a good time?</p>	<p>Number of people who stay after the official time of the party is over.</p> <p>Number of people who report having a good time.</p>	<p>Why might some people stay after the party has ended? Some reasons might be that they are too tired to go home; they have missed their bus; or their Uber driver did not arrive.</p> <p>The key lesson in this discussion, however, is that having one indicator to measure a very large and critical result is often not enough to fully understand the result (in this case, whether and to what extent people enjoyed my party). This is probably a question that is better answered with an evaluation (in-depth questions) than with having just one indicator. While the other indicators would help me make changes (buy more drinks, change the music) to improve my party, the final question probably requires more information. For example, I do not want to have an indication that everyone had a good time; the party is important to me. I want to know <i>why</i> they had a good time (or if they did not have a good time) and what made them enjoy themselves. A simple “Most respondents reported having a good time” would leave my information needs unsatisfied. Thus indicators are great for monitoring, but not so much for some important aspects of evaluation (e.g., valuing and judgment).</p>
<p>²Remember that Question 1 needs to be answered before the party starts, and that Questions 2–6 are to be answered after the party starts, to provide real-time data.</p>		

five representatives respond that The Nature Conservancy, a large organization, collects these data in each country. You breathe a silent sigh of relief and, as the day is over, end the meeting. Several weeks later, you follow up with the representatives of the two nonprofits who did not define the word *adequate*. They report that they never did reach agreement and they will get back to you, but they do not.

It is critical that no suitcase words exist in the indicator *unless* they are unpacked in another document (e.g., indicator specification sheet). If suitcase words do exist, define them. While defining the word can seem like an easy task, at times it can be formidable. Here, an evaluator would draw on her facilitation skills to support groups to define such words so that each word or concept is assessable and meaningful (or bring in a knowledgeable other to support the process; see Chapter 1).

A month later, you contact the representatives of the five nonprofits that defined their indicator to check on their data collection and see if they need any support. Only three of the nonprofits have the information they need. “What happened?” you ask the representatives of the other two. One representative tells you that the Nature Conservancy in her group’s country is extremely large and the group did not have the name of the person who captures these data, so she was just passed around and around, and was unable to get the data the group needed. The other representative had the name of the person who was responsible for the data that were needed, but when he called her, she said they only collected the data annually (the data are needed semi-annually). Now there are only three nonprofit organizations that have access to the data they need. But wait: Now the representative of one of these organizations calls you back and says, “I have a slight problem. The data we needed are available, but many other organizations have identified the secondary data as inaccurate and untrustworthy.” Oh, dear.

There are three critical lessons here. First, when a program manager of a nonprofit (or any organization) is depending on another organization to collect necessary data, the program manager needs to ensure that there is a clearly identified contact person who will provide the data (or access to the data). Second, although secondary data can be very useful, as such data can save a nonprofit (or any organization) time and money in data collection, these data must be accessible when needed. Third, the secondary data must be credible.

BOX 9.2. Primary and Secondary Data

Primary data are data for which you have control over how to collect the data, where to get the data, and from whom. Secondary data are collected by someone else who has control over how it is collected and from whom. *If you have no control over the data’s credibility, then they are secondary data.*

There are now only two nonprofits writing their reports using their indicator data. But now a representative of one of these nonprofits calls you and tells you that a big argument

has ensued about whose responsibility it is to analyze the data and turn it into a report. The argument seems to be rooted in two overlapping problems: The organization has no staff who are confident to interpret the data, and no one will take responsibility.

The critical lesson here is to ensure that the person responsible for reporting on the data understands how to analyze the data and turn it into a report—or that the nonprofit has the resources to access a team or an individual who can support the group in doing so.

Now only one nonprofit remains that has no complications. This nonprofit has several staff people who understand and can analyze the data; facilitates the use of the data by presenting it to the management team for reflection; and then provides the findings to a staff member who writes a well-constructed final report.

Sigh. One out of seven nonprofits has submitted a report including the mandatory information. Do not despair; it is through mistakes that we learn. This example demonstrates how little pitfalls can turn into gigantic, gaping holes in the evaluative process. In working with any organization, client, or colleague, thinking through these kinds of details (long after choosing or being given an indicator, progress marker, or other way of assessing) is critical to having a useful evaluative process.

In the chapter you talk about other ways to assess progress or change besides the indicator, such as progress markers and MSC. Can you talk a bit more about MSC?



The indicator is the most popular method of measuring progress. The progress marker has its own small following. The *most significant change* (MSC) approach has a core group of supporters, including me. Sometimes called the *story* approach, it is an approach that does not use indicators; rather, data are collected throughout the intervention's implementation through storytelling, which then provides data on results, usually at the higher level (i.e., not the activity or output level). So, for example, stories are not collected to indicate that an intervention took place, but rather to describe what happened because of that intervention, with a specific focus on providing contextual details and specifics (such as who did what, when, and why). A key part of the inquiry is to learn *why* a person is telling that story because the answer to this question makes it clear how that story is important, and thus what makes it an example of "significant change" (Davies & Dart, 2005).

I am working with a program that needs to develop indicators. We have a clear theory of change, and that is helpful to determine what to assess. However, is guidance available on the typical or necessary indicators in certain fields (e.g., education, health, environment)?



There are global initiatives such as the UN Human Development Index, Education for All, the Convention on International Trade in Endangered Species of Wild Flora and Fauna, to name a few, and even the UN's Sustainable Development Goals can provide guidance. Some

of these indicators can be mandatory for your intervention, or they can provide guidance on what kinds of indicator data can be useful. Furthermore, government departments and large organizations often have data indicator sets; be sure to ask whomever you work for or with, if any such guidance exists.



INDICATORS, PROGRESS MARKERS, AND MSC

- **Have a few minutes?** Check out a page on the Centers for Disease Control and Prevention (CDC) website for a quick description of indicators (www.cdc.gov/std/Program/pupestd/Developing%20Evaluation%20Indicators.pdf).
- **Have 30 minutes or more?** Glance through more information about MSC on a document available on the Overseas Development Institute (ODI) website (www.odi.org/publications/5211-msc-most-significant-change-monitoring-evaluation).
- **Have some more time?** For indicators, read the Performance Monitoring and Evaluation Tips provided by USAID (https://pdf.usaid.gov/pdf_docs/pnadw106.pdf). For outcome mapping and progress markers, look at www.outcomemapping.ca for an explanation, a free book, training materials, and a community of followers. For MSC, Rick Davies manages a website named Monitoring and Evaluation News (www.mande.co.uk), which provides a wealth of information on all kinds of M&E topics; the information on MSC includes links to papers, descriptions, blogs, databases, and trainers (www.mande.co.uk/special-issues/most-significant-change-msc).

WRAPPING UP

An evaluator needs to prevent people from jumping straight into discussions of indicators, or of how to assess an intervention, without first giving attention to everything else that influences these decisions. Having a discussion of what and when to assess is intricately intertwined with the program logic and theory of change, which are in turn affected by three types of factors: practical factors, use factors, and technical factors. After deciding what and when to assess, an evaluator must then consider how to assess progress; indicators constitute the most common quantitative approach, and progress markers and MSC are slowly gaining ground on the qualitative side. A conversation on how, when, and what to assess should not be mingled with a discussion on targets and benchmarks; as in any evaluative discussion, it is important to keep the conversations focused on one topic at a time. In this chapter, we have also played the Party Game and learned in this activity about developing and critiquing indicators. We have also filled the role of an M&E expert and learned several critical lessons on things that can go wrong in monitoring a program, and ways to mitigate them. Before you leave this chapter, however, I invite you to have a conversation with me.

Our Conversation: Between You and Me



Here I describe one situation that may be encountered with regard to indicators, and I share a trade secret. Practicing with this common scenario will help you as you encounter these kinds of situations in the real world.

At the beginning of an intervention, indicator sets are mandated by the donor or government agency. These indicators are not helpful in managing the program, and they cannot be changed. The program manager does not mind collecting the mandated data, but she still needs data that will inform her management decisions. How should an evaluator advise her client?

I am going to provide a more in-depth response than the one I have provided earlier in the chapter, and in doing so, I share a trade secret. When an indicator is required (e.g., a manager or board of directors mandates it for accountability purposes) that is clearly not useful for managing a program, it is acceptable to create additional ones that do—and thus to have two sets of indicators or other ways to assess progress. I am *not* suggesting that as evaluators, we should act like dishonest accountants who keep two sets of books, one of which is undisclosed. We are not hiding data; rather, data are collected and used for different purposes and for different users. It is acceptable to collect data that are not used by the donor or others to whom a program is accountable. Indeed, I recommend letting the donor or others know that you are collecting additional data. Should anyone want to see these data, then by all means share them.



To be a bit more specific, there are different audiences for different data. More simply put, different people want or need to know different stuff, and sometimes at different times. Sometimes broad accountability measurements are needed by the funder on a yearly basis. However, the program manager needs monthly data that are very specific and frequent. For example, knowing the number of people (disaggregated by gender) who attended a training could be “enough” information for your donor. However, to manage the intervention, the program manager needs data on which geographic areas the attendees came from, and how well the topic was presented.

So that was a monologue, not a dialogue. Here is the conversation: What happens when there are no additional resources to collect more data, and the data being collected are not useful for day-to-day management of the program? Now we (the evaluators) are invited into the discussion. How should we support the program manager? First, we commiserate with the group or person. Second, we identify the reason(s) why the indicator is irrelevant. Third, we identify new indicators, or determine what else can be done to assess change over time. Fourth, we present and negotiate a change in indicators or in how progress is assessed. These four steps may not result in a change, in which case we revert to the first step. *How would you start a conversation where someone approached you about an indicator that did not provide useful data?*

Copyright © 2019 The Guilford Press.

No part of this text may be reproduced, translated, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the publisher.

Purchase this book now: www.guilford.com/p/podems

Guilford Publications
370 Seventh Avenue
New York, NY 10001
212-431-9800
800-365-7006
www.guilford.com