

2

COMMISSIONING GREAT EVALUATIONS

WHO SHOULD USE THIS CHAPTER

This chapter focuses on what an evaluation commissioning manager (ECM) as defined in Chapter 1 needs to do to ensure a great basis for the evaluation. However, evaluation implementing managers (EIMs) can benefit from this chapter by understanding what the ECM should do or has done and what additional questions they may want to ask. Remember, internal evaluation managers frequently play both roles of ECM and EIM. Finally, sometimes ECMs hire evaluators to help in this process.

A senior person in a large nongovernmental organization recently confided in Jane that their organization had spent tens of millions of dollars on evaluations over the years. *Every single one* had been a disappointment, some even downright embarrassing.

This experience is far from unique. Over the years, we have heard similar stories from government agencies, philanthropic organizations, educational institutions, businesses, and many evaluators who work with them.

While some of the responsibility for evaluation gone awry must rest with the EIMs, an important theme of this book is that it takes two to tango. If a client organization does a poor job of scoping an evaluation, writing the terms of reference, or finding the right team for the job, even the most skilled and conscientious evaluation team will find it impossible to deliver what is truly needed. This is the role of the ECM in setting the stage for a great evaluation—one that matters.

A few years ago, Jane published a tongue-in-cheek blog post titled *9 Golden Rules for Commissioning a Waste-of-Money Evaluation* (Davidson, 2010) on the Genuine Evaluation blog. That post generated more engagement than anything else she and co-blogger Patricia Rogers had published. Evaluators from nearly every continent, working in virtually every sector you can imagine, chimed in to share a myriad of ways in which they had seen the investment in evaluation being wasted due to serious flaws on the commissioning side.

How prevalent is this problem? Although we are not aware of any systematic research that might answer this question, our discussions with a wide range of clients and fellow evaluators certainly suggests that it is widespread, although there is variation depending on sector, country, and so forth. Whatever the prevalence really is, there are lessons to be learned that can make a huge difference to the quality and value of evaluation.

Different language used to describe scoping:

Terms of reference (TOR) set out the purpose(s) of the evaluation and its key evaluation questions, a timeline (and possibly the available budget and existing data resources), and any direction related to staffing the evaluation; for external evaluations, this can include contractual arrangements. In some cases, the TOR includes an evaluation plan setting out the methodology.

Scope of work (SOW) is a plan for conducting an evaluation that outlines the work that is to be performed by the evaluation team. This can be included in the request for proposal.

Request for proposal (RFP), sometimes referred to as *request for application* or *request for quotation*, is a formal request for evaluators to prepare a response to a planned evaluation and is generally used to select the final evaluator for the evaluation. An RFP would include a SOW, staffing requirements, budget and contractual information, information on how evaluators will be selected, and instructions for the proposal.

In this chapter, we take a careful look at where the slippage most often happens and what ECMs can do at the front end to develop the scope, commission the kind of evaluation they truly need, and find the right team to bring it to life. We will also explore how commissioning organizations have different requirements and levels of flexibility in how they scope the evaluation and the implications for commissioning evaluations. Before we begin, let us define some of the terms used by various commissioning organizations—*terms of reference*, *scope of work*, and *request for proposals* (see above box)—and how they relate to the ultimate evaluation contract awarded.

Some commissioning organizations use the *request for proposal* (RFP) and *scope of work* (SOW) terminology while others use *terms of reference* (TOR). Figure 2.1 shows the overlaps and what is ultimately included in the evaluation contract for external evaluations. Internal evaluations will generally only need TOR since there is no competition for the evaluation. For simplicity, we will use *terms of reference* (TOR) throughout this chapter because it includes all the main points for commissioning a great evaluation: the work to be done, the requirements for an evaluation team, and the resources available.

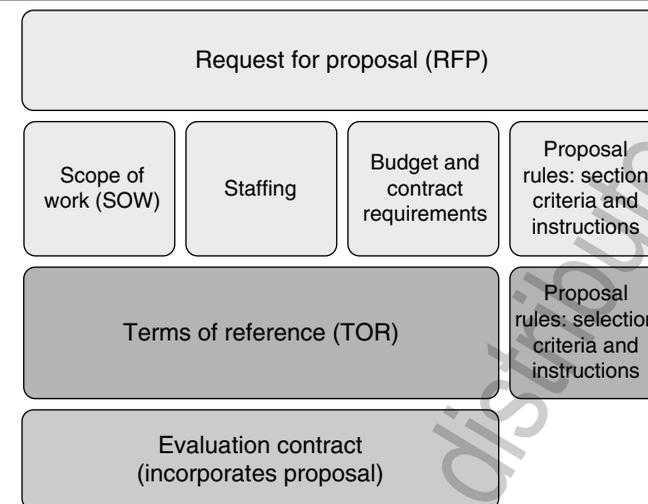
2.1 DECIDING WHETHER AND WHAT TO EVALUATE

How do we decide whether to and what to evaluate? This might sound like strange advice coming from a couple of evaluators, but one thing you should seriously consider before even getting started is whether to evaluate a particular initiative *at all*. Yes, really!

Evaluation is an investment that will require not only the funds set aside to pay evaluators but also the time and effort of the leaders and staff whose program or policy is being evaluated and the other stakeholders expected to participate in the evaluation. Before investing in evaluation, make sure it will give you a good return—that it is likely to be a “value-for-money evaluation,” as described in a 2014 blog by Dr. Caroline Heider, then director of the World Bank Independent Evaluation Group. John Wholey (1987) called this an *evaluability assessment*.

Under what circumstances might an evaluation be a bad idea? Sometimes it is blindingly obvious that an evaluation is unsafe, impossible, or highly inadvisable. Reasons

FIGURE 2.1 ■ Relationship Among Terms of Reference, Scope of Work, Request for Proposal, and the Evaluation Contract



include an unsafe context (e.g., in an active war zone) or total unavailability of a crucial type of evidence (Dershem et al., 2018). These extreme circumstances aside, it is still important to ask whether a feasible evaluation would in fact deliver insights valuable enough to justify the needed investment of time, effort, money, and other resources. “Programs with sparse data and other challenges will require a higher level of expertise and a greater intensity of effort, both of which can be costly” (Dershem et al., 2018, p. iii).

The other part of the story is to ensure that whatever is being evaluated is the right thing to evaluate. This may seem like an obvious point, but we have been surprised over the years to see many organizations commission evaluations that were only ever going to tell them things they already knew or things that were academically interesting but not particularly useful or actionable. Some of this is due to asking the wrong evaluation questions (we will explore that later), but sometimes the mistake is deciding to evaluate something that simply cannot provide the insights needed to inform thinking, design, and/or decisions. For example, Tessie found out in a kick-off meeting of an evaluation with a foundation that the key aspect of a program to be evaluated had been evaluated by a different team six months prior asking the same questions. The reason? It was in the agreement of the funding partners, and the foundation was going ahead with the plan that had been agreed upon two years prior. (You can find out what Tessie did and what happened in Chapter 4.) Who should be part of deciding what to evaluate when commissioning an evaluation? Do not keep this important conversation between an external evaluation adviser and the in-house evaluation specialists. Instead, make the most of the opportunity by involving your organizational leaders and decision makers in the conversation. Why? Because the true value of evaluation—and the numerator of its return on investment—lies in its eventual use and influence. Evaluation directors contribute most when they succeed in linking the work program of evaluation to key strategic questions, needs, and goals of the organization. There is nothing more potent for evaluation than organizational leaders waiting with bated breath for its results and then folding them into strategic discussions for the organization and its programs. Well, actually, there is one thing that can be more potent—in fact, galvanizing—for evaluators: engaged

public anticipation for an evaluation's results. Of course, many evaluations are completed without such level of engagement and anticipation, but a strong level of interest by decision makers and intended users of the evaluation is critical for value-for-money evaluation. A high-quality evaluation languishing in a filing cabinet is a major waste of resources. The only way to get use and influence right is to engage early and deeply with primary intended users of the evaluation (see Patton, 2008, and Figure 1.4 in Chapter 1 on engaging stakeholders in the evaluation process).

What should we ask when deciding whether and what to evaluate? Here are a few questions to guide the conversation.

1. **When we look across our entire portfolio, what do we already know and what great unanswered questions remain, even after having completed several evaluations over the years?** What do we already know and definitely do not need to ask yet again? What is already answered by our monitoring system? What are we frustrated about still not knowing? Who would use these answers if we could find them? What insights and understandings might usefully inform our thinking and action?
2. **If we genuinely want to find answers to those great unanswered questions, what would be most fruitful to evaluate?** Sometimes the best *evaluand* (what is being evaluated) is not yet another program. It may be more valuable to bring the focus up a level to conduct a more strategic evaluation of a portfolio of initiatives, policy instruments, or even the organization's entire strategy itself.
3. **Is there a more cost-effective way to get the answers we need, other than a full evaluation?** Options might include a synthesis of existing evaluations, a piece of nonevaluative research to fill a gap and answer some important nonevaluative questions, or a more in-depth assessment of the aspects of a program or policy that are the least well understood (e.g., because program evaluations never have the resources to go deep on that part). Or are we better off investing in evaluation capacity in our programs so they can undertake narrower, more locally relevant evaluations? Sometimes the best option is a mix of complementary pieces of work rather than a single evaluation or research project.
4. **How would we use the results of this evaluation?** Imagine the evaluation is complete, and it is wildly successful, *exactly* what you needed, useful, credible, and influential. What happened to achieve this result? What does it look like? How do you know it is a raging success? What made it so successful? How was it different? What was its breakthrough?

Once you identify the key primary intended users of the evaluation in the commissioning organization (and outside) and fundamental questions to guide a conversation with them, how do you engage them? A question we often hear, especially from internal evaluators, is, "It is so hard to get stakeholders to evaluation meetings. Do you have any tips?" Yes, we do! One reason why evaluators often do not get good engagement is that they may not be speaking at the level and in the language that resonates with their non-evaluator audiences, especially more senior managers. Internal evaluators in particular will find that these strategic conversations are powerful for building the visibility

and credibility of evaluation and evaluative thinking among non-evaluator colleagues and stakeholders. Another trusted strategy is to "talk money." People are frequently surprised by seeing the investment that goes into evaluations: direct expense for consultants but also time (the person hours and hourly rate) of ECMs in organizing, liaising, responding to, reviewing, and generally interacting with the evaluation. Add to that the probability times the potential cost of an evaluation going awry and having to fix political fallout and the probability times the potential opportunity cost of not using a useful evaluation, and you can develop a convincing case for focused attention to evaluation by leadership right now.

2.2 EVALUATION CRITERIA AND SCOPING

Once you have identified the most fruitful thing to evaluate, the next step is to figure out what the scope and focus of that evaluation should be.

We cover several considerations:

- drawing boundaries for what is in and what is out of the evaluation's scope
- providing enough clarity on what is needed without over-specifying the evaluation approach or methodology
- thinking expansively about the purposes and potential uses of the evaluation
- articulating the values that need to guide the evaluation

2.2.1 Evaluation Boundaries: What Is In and What Is Out of Scope

Clearly defining the boundaries of what is to be evaluated may seem like an obvious point, but there is more to it than meets the eye (Williams & Hummelbrunner, 2010). Every project, program, or policy is part of one or more larger systems, cultures, or policy environments with various contextual factors that influence them. Well-designed change will consider the most important elements of those systems, which include facilitating and enabling factors that can enhance or inhibit change. Similarly, a well-designed evaluation should consider not only how well an initiative contributes to immediate needed change but also how well it creates lasting change, which often means influencing the wider systems in which it operates. For example, efforts to address inequalities in social, educational, and/or economic outcomes should not only be accelerating progress for those who need it and repairing past harm; they should also be addressing the systemic and structural inequities that caused those unequal outcomes in the first place and that will otherwise continue to exacerbate them. Similarly, efforts to restore land, waterways, the air, and the atmosphere should ideally also address how the actions that caused the damage will be prevented in the future.

It is, of course, a delicate balance. Draw the boundaries for an evaluation too wide and the job becomes unwieldy and unfeasible within the budgetary and timeline constraints. Draw the boundaries too narrowly and the evaluation will be myopic and not sufficiently insightful. Add to that the politics of including elements that the initiative might not have been intending to influence but that may be important in hindsight, and you can see we have our work cut out for us! Do go ahead and bite the bullet, though; it is

important to have these conversations within the commissioning organization (or with your funder) *before* locking down an evaluation budget and timeline, if possible. This will help avoid getting boxed into a set of constraints that could severely limit the value of the evaluation before it even goes out for bid.

Frequently, in cases where an organization has defined the general budget available for evaluation, the question is one of setting priorities within these budgetary constraints. Is it best to centralize resources in one or a few larger evaluations, or is it best to have a mix of a few larger and a few smaller evaluations that are decided and controlled locally (i.e., at lower levels of the organization or system)?

2.2.2 Clarity—But Not Overspecification—About What Kind of Evaluation Is Needed

An important balance when scoping an evaluation is providing enough clarity about what you are looking for without over-specifying the evaluation approach and methodology. Over the years, we have seen countless RFPs where the required evaluation methodology is spelled out to a level of detail that no longer requires any serious thought by prospective bidding evaluators or organizations about what would be the best fit. For example, we have seen RFPs specifying that the methodology to be used will be surveys and interviews of a specific group of people, using a set of prespecified instruments, and sometimes even specifying the analysis strategies to be used. As experienced evaluators, when we see RFPs like this, our first thought is that this is a paint-by-numbers evaluation job that requires no deep expertise and could be easily conducted by a team of master's students. Pass! That is boring. But in all seriousness, if an organization already knows that this is exactly what it needs, maybe the right team is one of less-experienced evaluators working under an internal evaluator.

It is also very possible to be far too vague about what kind of evaluation you need. It is common to see RFPs asking for an evaluation to be done without much thought given to the kinds of evaluation approaches that might be the best fit. That is often not on purpose; it sometimes means that someone with, say, a policy or research background has been tasked with writing the TOR, and they may not be aware of the range of possible evaluation approaches. Non-evaluators are often surprised that our discipline is so deep and rich, with many different approaches that can be applied depending on the purpose and intended uses of the evaluation.

Many organizations that commission external evaluations or use internal evaluations rely on guidance and documents for how their organization manages evaluation. This is, of course, a good practice and part of building internal evaluation capacity, creating transparency, and enabling more people to become ECMs. Sometimes these rules, however, become too narrow. For example, in the World Food Programme's Decentralized Evaluation Quality Assurance System (DEQAS) that guides country-managed evaluations, data collection in the country is allocated three weeks, which is barely sufficient for secondary data analysis but woefully inadequate for participatory methods, gender analysis, or even engaging seriously in answering more strategic evaluation questions. In fact, the allocation of time for these country-level, country-managed evaluations places emphasis in planning (9 weeks), inception for more planning along with evaluators (7 weeks), and reporting (11 weeks). There may be good institutional reasons and history why this is the case, but it is worth reviewing the implications of

such evaluation resource allocation and the expectation that decentralized evaluations will rely mostly on secondary data analysis.

Here is a suggestion: Put someone in charge of writing the RFP who has attended at least half a dozen major evaluation conferences. Don't have one in house? Maybe it is time to invest in a couple of evaluation conferences for your writers of evaluation TOR—or at least send them on a good course! In fact, Tessie's evaluation management course usually includes proposal writers.

While those who *commission* evaluations do not necessarily need to have the kind of rich and deep expertise of those who *conduct* evaluations, they do need enough working knowledge to be informed and discerning purchasers and users of evaluation services. The more you know about the different options you might ask for, the more you can effectively request the kind of evaluation that is going to be most useful to you and your organization—or at least have more productive conversations with evaluators about the pros and cons of what they are proposing.

Here are some helpful tips for getting enough clarity about what kind of evaluation you need:

- **Stop short of building your own preferred design or approach into the TOR.** Instead, lay out some principles for what is going to align best with your goals. Focus on how you would like the evaluation to be used and how it will be influential: by whom, about what, and to what end.
- **Research is *not* evaluation** (or why some of the world's most useless evaluations were conducted by researchers instead of evaluators). If the RFP for the evaluation looks and feels like an RFP for an applied research project, it is probably missing the mark.
- **Do not be limited to formative and summative uses of evaluation.** There are so many more options out there: Developmental evaluation, which walks alongside an emerging policy or project and helps you define and develop it as you go, can be highly actionable in ways that traditional evaluation is not (Patton, 2011; see more on types of evaluation use in Chapter 9).
- **What is the evaluation really for?** Answering this question (as addressed in the next section) will also help clarify the kind of evaluation you need.

2.2.3 Purpose and Use: Identifying the Purposes and Uses of the Evaluation

Clarity about the purposes and uses of the evaluation can make or break any evaluation. Who and what is the evaluation for? If this is not crystal clear, you are in trouble.

Let us start with the primary intended users. Who is the evaluation for? Who *should* it be for? This is a powerful question that can drive strategy considerations (How do organizational leaders wish to engage colleagues and partners in developing or revising a vision and strategy?), decisions (How can program managers use the evaluation to inform their decisions?), advocacy (Who do we want to influence?), equity considerations (Are we effectively including everyone who should have a say in what gets evaluated and the key questions?), and organizational culture (Is this evaluation emphasizing collaboration, efficiency, audit, support for human rights, etc.?).

Evaluation often needs to consider the needs of a wide range of potential evaluation users:

- the overseeing agency or funding organization(s)
- cabinet, elected, or appointed officials in local, regional, or central government
- organizational providers or implementers
- participants/those impacted and their families
- communities, tribal organizations, and advocacy groups
- employers
- taxpayers and the public

This task—identifying the potential users (and “influencers”) of an evaluation—is sometimes confused with the task of identifying all who have a stake in the program. The groups overlap but are not the same. The primary intended users of an evaluation are those who are likely to see or hear the evaluation findings and apply its insights in some way—to inform decisions (e.g., to develop or improve the policy or program) or to feed into their thinking and actions more broadly.

Next, think through what these audiences will be using the evaluation for and what will be most useful for them:

- What do they need to know? For what decisions or purposes?
- When do they need the information?
- What kinds of evidence will be convincing to them? What level of certainty makes sense in this context?
- What need is there for evaluation capacity building (i.e., what skills do they need to be able to make best use of the evaluation)?

By clarifying how the evaluation is likely to be used and by whom, you can scope the evaluation to ensure it delivers, within budget, what the commissioning organization and various stakeholders need, when they need it, and in a form that is credible, convincing, and actionable. In fact, when you do a good job of engaging key stakeholders in responding to the questions above, you will frequently find yourself having to fit the excitement and expectations of these stakeholders within a finite budget. (This is a good problem to have.)

Determining the purpose and intended uses of evaluation should not be a desk exercise or a guessing game. Even a modest investment in consulting with potential users will help the evaluation manager shape a more compelling and useful evaluation as well as create anticipation among key stakeholders. In fact, when we have dug deeper on the question of purpose for an evaluation, a frequent goal is to solidify shared understanding among key stakeholders. This does not mean that everyone shares the same goals but that everyone understands each other’s interests and goals and respects them. Engaging those stakeholders from the beginning sets up an evaluation process that is more likely to develop shared understanding and working together toward a shared vision. When an evaluation feels like it is working well and people are engaged, curious, and open, it is frequently because the engagement process has been well managed, constructive,

and fruitful, even when people do not agree on everything. Michael Patton calls this “evaluation process use,” and it begins here with identifying the purpose and scope of the evaluation (Patton, 2008).

2.2.4 Values

An important contribution of evaluation is that it clarifies the values that are important to the commissioning organization and stakeholders and then applies them systematically to assess programs and organizations; for example, equitable and shared growth, women’s equality, human rights, sustainability, antiracism, and so on. When we articulate the values that are important in an evaluation, we enable evaluators to operationalize them and use them as part of the evaluation criteria. Once articulated, these values will inform the evaluation questions and be enacted in the methodology and highlighted in the evaluation report. These values will ultimately drive evaluation design: Who collects data from whom? What data sources and perspectives are included? Such values might include equity, sustainable development, human rights, and transparency. If the commissioning organization is committed to certain values—for example, if you want your evaluators to look like those who the program serves, if you want localization and engagement, if you believe that climate or equity are key—then be ready to articulate these in your TOR.

2.3 DETERMINING THE BEST WAY TO CONTRACT AN EXTERNAL EVALUATION

An important decision in commissioning an external evaluation is determining what kind of contract to use. Some contracts are easier for the commissioning organization, and some are more advantageous for certain kinds of EIMs but not others. The main types of evaluation contracts include the following:

- **Fixed-price contract:** This means the evaluation implementer is paid one negotiated price for tasks and deliverables outlined in the TOR upon completion of work or on a schedule of fixed-price amounts tied to specific deliverables or a group of deliverables. The financial risk lies with the evaluator—if the costs are actually lower, they earn more; if the costs are higher, they still need to produce the deliverables. Fixed-price contracts are simpler for invoicing and payments because they do not require justification of specific expenses. Note that sometimes the fixed portion is only for labor, and travel and other costs are reimbursed at cost.
- **Time and materials contracts:** The evaluator is paid for labor using fixed (negotiated) labor rates and material costs up to a ceiling. This has less risk for both parties because payments are based on what is used, although the commissioner holds the risk of not having enough funds because of decisions made along the way.

Fixed-price contracts are the most common for shorter duration evaluations, while time and material contracts offer advantages for evaluation services that span a longer period or where there is a need for more flexibility in the TOR over time.

2.4 GETTING PRACTICAL: THE PROCESS OF WRITING AN EVALUATION TERMS OF REFERENCE

In addition to the strategic discussion in earlier sections, an ECM will ultimately have to put pen to paper and establish the evaluation parameters in a way that evaluators can understand so they can submit strong proposals to conduct the evaluation. Every commissioning organization has its own requirements, processes, and norms for writing the TOR. The discussion below is designed to sit alongside those requirements rather than replace them. It may also spark some ideas for improving the TOR writing process in your evaluation commissioning organizations.

2.4.1 Content of the Evaluation Terms of Reference

Guidance on what kinds of content a good TOR should include is generally consistent across funders, as can be seen in the sample of evaluation funder guidance shown in Table 2.1.

As you can see, the TOR usually specify the following:

- purpose of the evaluation (including intended use)
- background and context

TABLE 2.1 ■ Examples of Guidance on Evaluation Terms of Reference Content from Three Evaluation Commissioning Organizations

General Topics (Comprehensive List)	United States Agency for International Development (USAID)	World Food Programme (WFP)	World Bank
Purpose	Purpose of the evaluation and how it will be used	Background	Background information and rationale
Background	Background and history of the projects or activities being evaluated	Reasons for the evaluation	Specific objectives of the evaluation and evaluation questions
Evaluation questions	Questions that must be answered and how they might be answered	Subject for the evaluation	Scope of the evaluation
Evaluation methodology	Expected deliverables	Evaluation approach, methodology, and ethical considerations	Approach and methodology
Expected deliverables	Expertise requirements	Organization of the evaluation	Governance and accountability
Expertise requirements	Time frame and budget		Guiding principles and values
Time frame and budget	Expertise needed to do the job		Professional qualifications
Requirements for submitting the proposal	Time frame and budget available to support the task		Deliverables and schedule
			Budget and payment
			Structure of the proposal and submission guidelines
			Additional references or resources

- evaluation questions
- evaluation approach or methodology
- expected deliverables
- expertise requirements
- time frame and budget
- requirements for submitting the proposal

While the general categories in an evaluation scope are similar, there is significant variation in the specificity required by different types of stakeholders. Some commissioners (e.g., United States [U.S.] federal government organizations) have strict requirements because of procurement rules, while others (foundations and nongovernmental organizations) have more flexibility. There will be more discussion about this later in this chapter.

There is ample guidance developed and shared openly on the internet on the mechanics of what to include and how to develop a good TOR for evaluation. In our experience, the art of a strong TOR within these categories hinges on avoiding a few pitfalls and engaging in some good practices.

2.4.2 Pitfalls to Avoid in Evaluation Terms of Reference

Hopefully, the guidance below will go a long way toward avoiding some of the problems we have all encountered in the commissioning of evaluations. You will see that most of the issues with a TOR are not because of bad writing but because of lack of preparation and consultation.

Pitfall #1: The Boilerplate Problem

Have you ever written an evaluation proposal—or waded through a pile of proposals—where the proposals seemed to be little more than large amounts of copy-and-paste boilerplate? Proposal writers and reviewers alike find this exasperating because it is both uninformative and boring! What might be behind that?

Using boilerplate language and copying from another evaluation’s TOR may mean that the author does not know what is needed, does not have access to people who can help clarify it, or is not interested in spending the time to figure it out, possibly because the bid seems like a long shot. A common result of boilerplate TOR is poor proposals that sound generic. To the proposal reviewer, this can feel as if the evaluation is not that important to the bidders. If the evaluation is important and worth submitting a proposal for, it is worth investing time to figure out what is needed, and if it is not important, it is best not to waste time and money on it.

If you are an ECM and this happens to you a lot, ask yourself—or ask some of your evaluation contractors—whether you are giving them sufficient information to be able to write an intelligent proposal. Some procurement processes forbid any interaction between bidders and the commissioner beyond a timed submission of questions and responses going to all bidders, leaving proposal writers flying completely blind on many of the essential details. In the absence of a nuanced understanding of what you need, they fill the white space with their “tried-and-true” formula and hope it looks impressive enough to get them on the shortlist so they can actually talk with you.

Pitfall #2: The Internal Contradiction Problem

Imagine how confusing it is being asked to implement a participatory approach to an evaluation with only two weeks for data collection or imagine being asked for a culturally responsive evaluation that prescribes a methodology of only a survey using an existing survey instrument. Such contradictions typically show up in a TOR when a draft is circulated, different people provide their edits, and those edits are accepted without any follow-up discussion among editors or consideration of the implications.

Pitfall #3: Being Vague About the Budget

Sometimes a TOR, especially for newer commissioners of evaluation, does not provide cost guidance and budget expectations. The motivation is frequently to ensure cost competition and pay the lowest price they can for the evaluation. This silence on the budget might also be motivated by shifts toward evaluation procurement that appear to be designed to mimic the purchasing of products. The broad intent behind the competitive bidding approach does make sense, given the need to ensure that taxpayers' or donors' money is spent wisely. However, it is often highly problematic in practice.

In the early 2000s, when a U.S. federal health agency first expanded its international programs that included evaluations, it repeatedly canceled contract competitions because it consistently underestimated the cost required to implement the scope of the work. After the sixth such cancellation under its contractual mechanism Global AIDS Program: Technical Support Services, all six eligible contractors wrote a joint letter that they would cease bidding and wasting their proposal funds on competitions that went nowhere. Eventually, the commissioning health agency increased its costing capacity and established realistic expectations, and the bidding resumed.

Clear budget and time parameters need to drive the number and scope of evaluation questions and evaluation requirements. When there is not even a ballpark budget given, it is extremely difficult for prospective evaluators and evaluation teams to figure out what the ECM is seeking. Evaluation questions can be answered in various ways, starting with a lean approach (possibly relying primarily on secondary data already collected) all the way to an extremely thorough, high-quality, and robust evaluation with site visits and extensive primary data collection or something in between. When the budget is vague, the evaluation bidding process becomes a “guess the number in my head” exercise. It is of no use to anyone if many proposals are way off base because they scope and design an evaluation that is either methodologically overkill or far too light.

Pitfall #4: Procurement Inflexibility

Many of the evaluators and clients we speak to feel very constrained by procurement processes. These have been put in place with the best of intentions: to ensure the best possible use of funder or taxpayer money via a competitive bidding process and to ensure fairness and equitable access to all bidders. Makes sense, right?

Here is the problem. Many procurement processes were originally developed for the purchase of products (such as computers) or simple services (such as catering) where the requirements are predictable and can be specified in detail. Suppliers simply need to demonstrate that their product or service will meet all required specifications within the required time frame. The task for the procurement team is to assess track record and qualifications to ensure those claims are plausible and then choose the most cost-competitive offer.

Evaluation, in contrast, is much more complicated and complex. Even with clearly laid out evaluation questions and requirements, there are infinite possibilities for how to approach the job. In addition, it is almost certain that by the time an evaluation is awarded, the requirements will need to change. In some cases, such as for the U.S. federal government, procurement is governed by very strict regulations that leave almost no leeway. For example, some RFPs require an evaluation plan to be submitted two weeks after contact award, which makes any consultation almost impossible.

Best practice here is to include an inception phase in all evaluations, something that is thankfully becoming increasingly common practice. For example, the United Nations' contractual requirements for evaluations include an early product called “the inception report,” which includes any revisions to the evaluation questions, requirements, and methodology mutually agreed to by the evaluators and the commissioner to update the evaluation plan based on the current organizational reality and relevant changes to the broader context. Another practice to consider is not including specific evaluation questions but articulating some lines of inquiry and using the inception phase to develop the evaluation questions. This allows additional flexibility to shift or tweak the evaluation questions as you are finalizing the design and getting the input from your evaluators on what questions can be reasonably answered within the evaluation constraints (time, budget).

We have also seen evaluation procurements that include an evaluability assessment/design phase and a separate phase for evaluation implementation based on the design. This lowers the risk for both the commissioning organization and the evaluator. Some foundations conduct a request for qualifications and then invite one or more organizations (who are compensated for their proposal preparation) or require the entire evaluation plan to be developed during the inception phase. Not only does this better serve evaluation commissioners by ensuring the evaluation team has a deep understanding of requirements before the evaluation begins; it is also fairer to the evaluators. Developing an evaluation plan is a significant amount of work and is not something we should be asking people to do for free.

Building a strong inception phase into your TOR allows the important dialogue and collaboration between the ECM and EIM to get the evaluation design right (rather than in the proposal submission phase; see more in Chapter 4).

Pitfall #5: The Methodology Directive Trap

Sometimes a TOR goes overboard and not only specifies requirements but also spells out exactly how to fulfill them. When that is the case, there is no room for evaluators to showcase their thinking and demonstrate their expertise. There may be cases when commissioners do know exactly what they want, in which case, it might be more cost-effective to hire evaluation interns or students with university professor supervision rather than invest in a professional evaluator. Or the TOR writer might have fallen into the “similar to me” trap, lacking imagination for what they do not already know, thus seriously limiting the methodology options.

We can appreciate cases where an organization is devolving the authority to commission evaluation to different parts of the organization, and to compensate for possible lack of capacity, it provides narrow guidance for TOR development and cost. An example of such a heavily directive approach could be seen when the World Food Programme (WFP) initiated its decentralized evaluation practice, moving evaluation

commissioning for country evaluations to country level. In parallel, to support the growing of capacity in countries, WFP established a structure of regional evaluation advisers to support groups of countries in each of its regions and an intensive online evaluation training program for staff members when they were first asked to act as ECMs. The specificity of the guidance for decentralized evaluations was comforting to newly minted ECMs in country offices, while it provided time for WFP to invest and expand its evaluation capacity throughout the organization. We have seen similar evolution paths followed by some of our foundation clients.

Pitfall #6: Making the Proposal Process Too Burdensome for Smaller Evaluation Firms or Those Representing Marginalized Communities

If we make the procurement process complicated and time-consuming to prepare and submit a proposal, we will likely prevent the participation of those evaluators we want proposals from in the first place. Requiring a very detailed design in the RFP means we are asking evaluators to do a lot of the evaluation work for free, since they are only paid if and when they have won the contract. Commissioners need to be aware that preparing proposals costs money, and smaller or less-established evaluation firms will be hard-pressed to prepare long, detailed proposals. As mentioned in Pitfall#5, there are ways to address this by lightening the RFP phase and putting most of the design work into the contract itself in an inception phase, where the evaluators can be paid for their design work.

2.4.3 Good Practices in Writing Evaluation Terms of Reference

To avoid the pitfalls discussed above, here are some practices that we have found useful in our experience with writing and responding to evaluation TOR.

Practice #1: Help Your Organization Clarify Its Needs and Convey Them in the Terms of Reference

Begin by imagining that the readers of your TOR are considering whether to submit a curriculum vitae (CV) or proposal and are trying to figure out what they need to include in their proposals. The most important characteristic of a good TOR is clarity and the most important thing you will do is to engage in the discussions needed within your organization to get clarity. Clarity is not simple because frequently there are many stakeholders with different points of view. Furthermore, in some cases, satisfying one requirement will directly contradict another requirement; for example, on the one hand, you wish to ask for a participatory evaluation with extensive consultations while on the other hand, you may wish to minimize the cost and limit the timeline. These are choices TOR authors need to push their colleagues to address so the TOR is clear on expectations and boundaries and reflects an organizational consensus and not only the author's point of view. Consultation within the organization is worth the investment to ensure your TOR will lead to an evaluation that matters—it is relevant to the organizations and key stakeholder needs so that it has the potential to be used.

The bottom line is, ensure that your TOR

- conveys who the evaluation is for, what is it trying to inform, and who will make those decisions and by when;
- includes your draft questions but invite input;

- shares the values that are important for your organization, which should inform the way the evaluation should be conducted and what it needs to cover (such as equity or sustainability);
- comments on methodology but invite the experts to suggest their ideas; and
- provides the budget level and the assumptions you used as well as timeline parameters.

Practice #2: Know When to Remain Open and Flexible

While it is important to be clear on what is required, it is also important to be clear where the TOR invites input and ideas from evaluators. Those areas might include the following:

- What other questions do evaluators think might be important given the purpose and context of this evaluation?
- What considerations (risks and opportunities) are important for designing the evaluation and selecting methods? What do evaluators think is the best methodology?
- How are evaluators planning to engage stakeholders?
- How do evaluators think about ensuring the evaluation reaches and includes all key stakeholders and underserved minorities?
- The TOR can also explicitly invite responders to push back if they think something is unrealistic or ask for reflections on the TOR itself or on how they would go about constructing a learning agenda.

Giving responders areas where they can showcase their thinking, knowledge, expertise, and creativity will serve you in the selection process and is likely to be helpful no matter who is selected. Therefore, within the procurement rules of the commissioner, consider being explicit about the values that are important for the organization and program and specifying the required approach rather than the methodologies needed.

Practice #3: Clarify Your Values About Who Should Be Involved and the Implications

Evaluators need information on who you consider important to include in the evaluation. This will provide a more culturally and contextually responsive proposal and a more realistic budget. While the TOR will not specify all methods, the extent of the stakeholder consultation required should be reflected in the timeline and be adequately budgeted. Here are some questions to help guide our thinking:

- Who needs—and has the right—to be engaged and informed directly, and who can be represented in different ways? For example, if several stakeholder groups have recently produced analysis that represents their point of view on key questions, these do not need to be repeated.
- If stakeholders are dispersed, will you require travel or online consultations?

- How might the evaluation engage representatives of some key groups (to limit the numbers) and then, after data collection, consult with a wider number of stakeholders to engage in evaluative sensemaking, validate findings, and develop recommendations or action plans?
- If the evaluation will require deep engagement across distances or in different languages, the budget must enable it. We have seen examples of RFPs calling for a participatory approach for people in rural communities with an expectation of one-week visits when it requires two days to get to a site. This is simply untenable.
- If you do not have the resources for meaningful participation but still believe in it, invite responders to suggest strategies for increased participation within your means.

Practice #4: Develop Realistic Budgets

When commissioning an evaluation, the evaluation budget needs to be developed by someone who understands evaluation enough to ensure that the provision for resources is adequate and allocated appropriately. Frequently, procurement regulations make it impossible to share the total budget allocation with bidders because it is assumed that cost competition will be most advantageous for the evaluation commissioner. In fact, the result is frequently an inefficient proposal process that focuses on lowballing the competition and under-resourcing the evaluation. A different and more productive way to stimulate competition is to let everyone know how much money is set aside for the evaluation or the cost range expected and focus the competition on the most cost-effective use of the available resources to get the best evaluation possible. Otherwise, the range of assumptions made in different proposals is so wide that they are not comparable. And, if the commissioner selects the lowest-cost proposal, it will either not be done well or the “smart” contractor will later negotiate a cost adjustment to make the budget more realistic. The U.S. Government has a category of proposals where its selection is based on “lowest cost technically acceptable,” and we have seen bidders propose unrealistically low prices, which have been later renegotiated and increased. The most respectful and fiscally responsible practice is to conduct cost analysis and aim for cost realism, not lowest cost.

Practice #5: Include Evaluation Competencies in Criteria for Assessing Evaluators

Most of the attention in a TOR is given to background and methods and less on laying out the requirements for evaluator competencies. This might have been partly because evaluation communities have only recently developed competency frameworks and partly because it is assumed that someone who understands this type of program will be the best person to evaluate it. We have seen a TOR that values content expertise the most (perhaps because TOR writers are most familiar with it) while showing little appreciation of the evaluation expertise needed and a devaluing of context and cultural expertise. An evaluation of a transportation program, for example, may ask for a transportation expert (content expertise) and a survey designer (limited evaluation expertise) without mention of expertise in context and relevant stakeholder communities (e.g.,

remote station travelers, unions, etc.). Overall, we recommend centering evaluation expertise for an evaluation TOR, placing a high priority on contextual and cultural expertise, and placing content expertise in an advisory capacity.

2.5 SELECTING A STRONG EVALUATION TEAM (OR SOLE EVALUATORS)

In developing an evaluation TOR, commissioners of evaluation focus first on the evaluation scope, questions, timeline, and budget, and this sometimes leaves less time to develop the specifications for the type of evaluation team they need. In fact, when the TOR is developed by non-evaluators, it is difficult to provide any kind of specificity about what is needed. And yet, given that evaluation is not generally a standardized product, selecting qualified evaluators who are a good fit for the specific evaluation is critical.

Although the practice of evaluation is millions of years old, our existence as a profession and a discipline is much more recent (Scriven, 1991). One consequence is that clients and people working in other disciplines are often not aware that there is any such thing as specialized evaluation knowledge and know-how.

2.5.1 Avoid Pitfalls in Evaluator Competency Specifications

The following are some frequent pitfalls in developing staffing requirements for an evaluation.

Pitfall #1: Hiring Researchers to Conduct Evaluation

When evaluation emerged as a profession separate from research, the first evaluators were social scientists. Still today, many evaluators get into evaluation through social science research. These researchers need to change their frame of mind to perform well in evaluation.

More typically, research is funded by grants, has longer timelines, and does not have a visible client. We have seen many nonevaluative research studies with the word *evaluation* on the cover but that clearly are not actually evaluations. What do we mean by that? They tend to describe what happened (quantitatively and/or qualitatively) but they are not at all clear about whether whatever happened was significant and/or to what extent. In other words, there are no explicit conclusions about merit, worth, or significance. The reader is supposed to figure that out.

An evaluation typically has a shorter timeline, the people commissioning it need information and lessons from the evaluation to inform upcoming decisions, and stakeholders tend to be more involved in decisions that influence the evaluation. This is why evaluator competency models include management and interpersonal domains alongside methodology and professional behavior and why context has emerged as its own domain rather than receiving an honorable mention under methodology.

Pitfall #2: Valuing Subject Matter Expertise Over Evaluation Expertise

A common related problem is assuming that all one needs to evaluate *X* is a deep knowledge of subject matter *X*. While subject matter expertise certainly needs to be used in evaluation, evaluation expertise is critically important for figuring out how best to

weave the subject matter expertise into the evaluative conclusions in a coherent and valid way. In fact, subject matter experts frequently disagree on the standards of performance for a given topic; selecting one expert over another may mean very different evaluation results. Good evaluators are skilled at identifying differing beliefs, assumptions, and values and bringing them to the surface for discussion and consideration. Good evaluation makes such difficult and necessary conversations possible.

Our observation has been that studies performed by researchers with deep subject matter expertise (Pitfalls #1 and #2 together) but little evaluation expertise tend to get lost in the details and sidetracked by questions that are of academic interest to the subject matter expert but of little relevance to decision makers. This is why it is important to commission a team that will have an evaluation specialist as the lead rather than a content expert.

Pitfall #3: Devaluing Context Expertise

TORs will frequently identify content and evaluation expertise but are silent related to context expertise. What do we mean by *context*? If, for example, the evaluation concerns a program that serves incarcerated populations, what should the evaluators know about the context of prison life and the needs of incarcerated people? If the program serves only women, would an evaluation team of all men be adequate? What if the program relates to refugee camps where gender-based violence is a very possible occurrence—do we expect that women program participants will speak openly to men evaluators? When programs serve people of color, including Indigenous groups, how will this be factored into the evaluation team competency requirements? There are other kinds of context knowledge that are important—knowledge about the local systems, the political and economic environment, and cultural and social practices.

In fact, the makeup of the evaluation team both in terms of competencies and in the identities of the evaluators themselves will greatly influence the effectiveness and credibility of the evaluation. And let us be clear, assuming that an evaluation team has sufficient cultural expertise when the only Black or Brown faces on the evaluation team are a couple of research assistants is not OK.

For those evaluators who believe that “value-free” evaluation is desirable or even possible, we assert that there can be no “value-free evaluators.” The challenge is often implicitly characterized as a language/translation issue, where it is assumed that the needed outcomes are very straightforward and culture/value-free. Instead of asking for value-free evaluation, it is best to specify the values implied in that request; in our experience, *value-free* might mean prioritizing efficiency, not taking sides/considering all sides, using the same standards across the board regardless of location or stakeholder group, and so on. Unpacking the concept of a value-free or objective evaluation will help clarify the values and standards that are being prioritized by this evaluation.

2.5.2 Strategies for Selecting an Evaluation Team with the Right Competencies

Screening people on paper is difficult, but here are some practices that can be helpful.

Practice #1: Emphasize the Relevant Evaluation Competency Domains

One of the lessons from the pitfalls mentioned above is that evaluation expertise should be prioritized. How do we achieve that? The emergent global consensus on evaluation

competencies (see Figure 1.5 in Chapter 1) is making a significant contribution to developing great clarity in how to seek the right evaluators for each evaluation. So, decide on the evaluation competency model that best fits your organizational context and include a reference in the TOR asking respondents to address how their team meets these qualifications. Also, use the evaluation competencies as a checklist to see what you need to emphasize for your particular evaluation. For example, a large and complex evaluation should include some language under the evaluation management competency domain; an evaluation that emphasizes high engagement of stakeholders might include language under the methodology/qualitative and the interpersonal domains—and so forth.

Practice #2: Ask the Evaluators to Provide an Example from Their Experience That Demonstrates the Top Competency Requirement for Your Evaluation

Let us say you need to keep track of and organize findings in an easy-to-access database of evaluation sources and methods because you will repurpose the data in the future. Ask candidates to give you an example of an evaluation where they had to work with big data sets and where indexing and retrieval of data were well done. If you need to have evaluators who have managed a politically charged issue, ask for an example of that. The personality of the respondent and their experience will come through more clearly than reading their CV.

Practice #3: Hold Oral Interviews as Part of Your Competitive Process

You will be working closely with your selected team and you will want to trust them with sensitive information and access to your people and data as well as your thoughts for the future. When you think of the potential benefit of a strong evaluation team and the risks of hiring the wrong one, you will see that including an oral part to the competition will give you valuable information and allow you to consider how well you would work with this group of people.

Practice #4: Check References

Never skip this step. Check references for team members and for the organization. We have found ourselves acting as members of evaluation commissioning groups, and we are grateful for this practice every time. In one case, a reference check of the work of a well-known evaluation organization (and one that the whole commissioning team thought was a shoo-in for the work) revealed that the organization was overstretched, changed its personnel frequently, and was sloppy in keeping track of the data and demonstrating the inference path of the findings, conclusions, and recommendations. In contrast, the reference checks of the team selected (that was less well known) revealed that a weakness of the team was that they needed to be encouraged to speak up because they were very insightful and methodical and had great instincts. We were very glad to select the second team, and they did a fabulous job in the evaluation.

Ultimately, you want an evaluation team that is flexible, attentive, and respectful. You want to trust the team to handle your sensitive issues and information responsibly. You want them to act with integrity and to hold to the path of independence and impartiality with empathy and understanding, and with a commitment to leaving your organization and programs better off for having intervened with high-quality evaluation. Picking a good team is worth your time.

2.6 SELECTING THE BEST EVALUATION IMPLEMENTOR

If you have done a good job with the TOR, this step will be much easier because you will have all the information you need. Selecting a person or team to conduct the evaluation may take place by competing the evaluation widely, to a preselected group of individuals or organizations, or through a sole-source process. We discuss these processes below.

2.6.1 Competitive Selection

Evaluations (especially large ones) are competitive, with some exceptions. Managing the selection process well is critical. There is a final section in your TOR that is essential for the selection, and that is the evaluation criteria for your evaluation and a scoring sheet that includes the maximum points that can be awarded for good performance in each criterion and the weight of each criterion (see the following, for example):

- Technical approach: 25 points
- Organizational track record: 25 points
- Evaluation team expertise: 50 points
- Cost ceiling (published in the proposal): *X* amount (if exceeded, the organization is disqualified)
- Total points: 100

Under each criterion, you need to specify what elements are needed for a high score, and this should be aligned with the TOR and be quite detailed. Also, each criterion should have a “bar”—a minimum level of acceptable performance. This is critically important for ensuring that exceptionally high scores in one area cannot make up for inadequate performance in another.

In some cases, the ECM might select the evaluation team without consulting others, but for most evaluations, some consultation is preferred or required. In public and international organizations, the selection process is usually more formal. It involves appointing an evaluation selection committee (about five people including the ECM who developed the TOR) and having the members of the selection committee review and score all or subsets of the proposals independently. A good practice mentioned above is the ability to ask questions of clarification to the proposing teams to make a better-informed decision. After the scoring is compiled and shared in the group, the committee members meet together to discuss the results and make a recommendation. In some cases, the recommendation is informed by a reference check conducted during the review; sometimes, the reference check is conducted only for the two or three finalists.

In theory, if you have done a good job with your TOR, this decision should be easy—and sometimes, it is. Next are some pitfalls we have seen in the selection process and how to avoid them.

Pitfall #1. A Selection Committee That Does Not Represent Key Stakeholders

If the composition of the selection committee seems biased, the risk of opposition and sabotage of the evaluation is high, even if the selection committee does a good job at applying the criteria fairly. Our suggestion is to lean into the discomfort of potential conflict and include those who may have a different view of what is important in selection. In

one of Tessie’s experiences, an international organization experienced conflict between the perspectives of its headquarters and its regional office for a contract that would take place in the region. The headquarters manager, adept in this dangerous political nuance, created a more elaborate selection process than what he had originally intended: He included regional representatives in the selection committee and assigned Tessie’s company (as a neutral partner) to conduct site visits and lead the scoring process, thus protecting the integrity of the selection process through transparency and negotiation.

Pitfall #2. Favoring People and Organizations You Like

If members of your selection committee express predictions about which they expect to be the best proposal, admonish them about the cost of this bias. In one such case, Tessie (who was leading an evaluation steering committee) saw the selection committee evaluate a proposal (that was mostly boilerplate) higher because the members knew the team leader personally. These committee members were so sure in their choice that they refused to conduct reference checks, so Tessie (a firm believer in reference checks) undertook them herself. To Tessie’s shock, the committee’s higher assessment of this weak proposal persisted even after the reference checks came out quite negative for this favorite team leader. The problem here was twofold: (1) Favoritism did not result in the selection of the best evaluation implementor and (2) Prejudging based on prior knowledge was unfair to those who submitted proposals and unethical on the part of the steering committee. Now, Tessie uses this story in preparing selection committees.

Pitfall #3. Manipulating the Questions to Bidders

This is another manifestation of Pitfall #2. In this case, selection committee members pose questions to their favorite bidder to give them a chance to strengthen their bid. In cases where you ask for an elaboration, a fairer strategy is to ask all bidders to elaborate on that same aspect of the proposal or to provide all bidders with the top weakness in their proposal and give them a chance to elaborate. U.S. federal government proposal processes require transparency in questions and answers prior to proposal submission and require any elaboration requests by the government to be shared with all bidders until they enter negotiations with a selected organization.

Pitfall #4. The Points Contradicting Your Intuition

When points and intuition are not aligned, more discussion is warranted. Ask yourself and the selection committee, *What would happen if I went against the point selection? What am I concerned about?* In responding to these questions, you may either come up with an insight that influences the points awarded, which would update and improve the scores, or you might identify a personal positive bias that you have based on friendship or a negative bias that you have based on past experience that you may have to let go. You might also articulate a criterion that should have been identified but was not or an elaboration to an existing criterion that was unclear. The answer is not straightforward and, in the end, you will have to make an ethical decision about fair decision making and your responsibility to the stakeholders.

2.6.2 Competing Among a Preselected Group

In the interest of efficiency, commissioners might share an RFP with a smaller group of potential evaluation implementers who have been prequalified for different types

of evaluation. The prequalification process may be informal or formal. Some foundations or nonprofits, for example, create a roster of consultants by word of mouth or simply reach out to consultants or organizations that have worked with the commissioner before. In the U.S. federal government, the prequalification process is a competition that results in a prequalification contract, usually called an indefinite quantity contract (IQC) or an indefinite delivery indefinite quantity (IDIQ) contract.

The ultimate type of preselection is awarding a sole-source contract to an evaluation implementor without competition, which is typically done with a justification that it is more efficient than conducting a wider competition.

We empathize with commissioners' efforts to use their resources more efficiently and shorten the process through preselection or prequalification. However, there are two important pitfalls to manage in a preselection process: favoritism and barriers to entry for smaller or less-resourced candidates.

Favoritism falls under Pitfall #2 (Favoring People and Organizations You Like). Over time, the favored evaluation team might stop being independent or impartial and become an extension of the hiring unit of the evaluation.

Barriers to entry are erected because small evaluation firms or some evaluators may not have the networks or resources to become known to decision makers or to pursue opportunities that require investment. If being considered requires being already in the network, how do you enter? Or, if you need to win a prequalification contract to be considered, that may represent a significant endeavor that a small evaluation firm cannot afford.

So, what should a commissioner do to make a process fair, inclusive, and efficient? Let us start by recognizing that commissioners must make realistic decisions that fit their own values and goals and do their best to establish fair and inclusive processes within their own resource constraints. Good evaluation management means value-for-money evaluation decisions, so commissioners must do their best within their means. Having said that, a lot can be done, and we have seen many creative solutions on the part of ECMs:

- Commissioners might include requirements to show preference to small firms, minority evaluators, or minority-run evaluation organizations. In the U.S. federal government, this can be seen in evaluation contracts reserved for competition only among small businesses in qualified minority-run businesses through the Small Business Administration's 8A program or other programs. In foundations and nonprofits, this could be an explicit criterion for selection.
- Commissioners might create a less intensive prequalification process through a short expression of interest and then reimburse a portion of the proposal development process for two or three finalists.
- Commissioners might develop intentional plans to encourage entry and placement of minority evaluators; for example, the American Evaluation Association's Graduate Evaluation Diversity Initiative (GEDI) hosted by different universities and initially supported by The Kellogg Foundation or internship programs being developed by United Nations agencies in different countries to support candidates through EvalYouth, the global network of young and emerging evaluators.

So here is what we are saying: We know our standards of independence, impartiality, fairness, and inclusion are high, and we work in the real world of limited budgets and time. We accept that we all carry our own biases and blind spots. Within this reality, we commit to being thoughtful, not always taking the most expedient path but intentionally including ethical considerations and decisions in our selection process.

2.6.3 Internal Evaluation Teams

In internal evaluation, there is more limited choice. Internal evaluation teams may be small and constant or only one person; in a larger internal evaluation office, there is a selection process usually made by the evaluation office supervisor. Internal teams might be supplemented with outside consultants depending on the evaluation budget, and their selection would fall under one of the categories discussed above. The way to enact impartiality, fairness, and inclusion for internal teams is through their staffing decisions and their evaluation process.

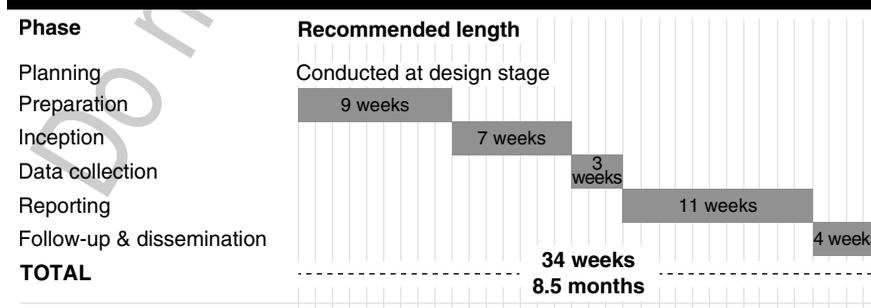
2.7 BUILDING CAPACITY IN COMMISSIONING EVALUATION

Building capacity in commissioning evaluations has a good return on investment by ensuring that more useful and reliable evaluations are commissioned that help the organization learn and grow its programming in more productive directions that serve the public good. Simultaneously, competencies in commissioning better evaluations bring about more strategic thinking and coherent programming because of the questions that ECMs ask of their leaders and colleagues. There is more discussion on this in Chapter 10.

CASE EXAMPLE

Review the distribution of time for a decentralized evaluation by the WFP DEQAS. List possible reasons why it makes sense and the challenges it might create. Based on your analysis, what would be your recommended time distribution allocation if the evaluation were still to take 34 weeks (see Figure 2.2)?

FIGURE 2.2 ■ Recommended Duration of a Decentralized Evaluation



Source: World Food Programme. (2021, April). *Decentralized evaluation—guidance for process and content*. <https://docs.wfp.org/api/documents/9f13fcec2d6f45f6915beade8e542024/download/>

