NOTE 4: Choosing the Right Type of Evaluation

The most serious mistakes are not being made
as a result of wrong answers.
The truly dangerous thing is asking the wrong question.

—Peter Drucker

Although a good monitoring system is critical to knowing whether our intervention is moving in the intended direction, it does not necessarily answer the question how or why changes are coming about, nor does it prove that any observed changes in outcomes are the result of our intervention. To complement the information we obtain from our monitoring system, we need evaluations. Evaluations are periodic assessments of the relevance, efficiency, effectiveness, impact, and sustainability of our intervention. The type of evaluation best suited for our project will depend primarily on our information needs. Therefore, the first step to any evaluation is to define what we want to learn. These learning objectives as well as our operational context, in turn, will determine which type of evaluation is right for our program.

What Is the Purpose of the Evaluation?

As a first step to deciding if an evaluation is necessary and which design should be chosen, it is crucial to clearly define what we want to get out of the evaluation. *What decision will be informed by the evaluation and what kinds of information are needed to make that decision?* Do we want to know more about how well our programs are being implemented, whether our programs are meeting their objectives, or whether our beneficiaries are actually better off as a result of our intervention? As program managers and evaluators, we must first establish our questions and learning objectives and then select the most appropriate evaluation tool to provide the necessary information (Karlan 2009).

Broadly speaking, evaluations address three types of questions (Imas and Rist 2009):

- **Descriptive questions** seek to describe processes, conditions, organizational relationships, and stakeholder views (*What is going on in our project?*).
- **Normative questions** compare what is taking place to what should be taking place. They compare the current situation with the specific objectives and targets that have been defined (*Has our project been implemented as intended? Is it performing as expected?*).
- **Cause-and-effect questions** examine outcomes and try to measure what difference an intervention makes. They ask whether objectives have been achieved as a result of our project (*What is the impact or causal effect of our program on outcomes of interest?*).

Which of the above questions we should ask is ultimately up to us, based on the specific intervention.

Organizing our questions. In practice, we may have many questions across all categories that we would like to answer. An effective way to organize all the possible evaluation questions is through our results chain (see table 4.1). In fact, if a good monitoring system is in place (see note 3), there should be consensus around our project logic in terms of implementation and results, which in turn makes it easier to identify the critical learning objectives along all stages of the intervention. Descriptive and normative questions can relate to all levels of the results chain; however, cause-and-effect questions specifically refer to outcomes and higher-level outcomes.

TABLE 4.1 Examples of evaluation questions

	Inputs	Activities	Outputs	Outcomes	Higher-Level Outcomes
Descriptive	• How does the cost of the program compare to similar interventions? • What are the qualifications of service providers? • What are other ongoing interventions?	• Do youth know about the program and how they qualify to join? • What delivery mechanisms are being used? • To what extent does the program implementation differ by site?	• How many youth participate (by age, sex, etc.)? • Who drops out? • What services are used the most?	• Are participants satisfied with the program? • Are there any observable changes in participant knowledge, attitudes, etc.? • How many program participants find employment within 3 months?	• Is local youth unemployment rising or falling? • Are household incomes evolving?
Normative	• Do we spend as much as we have budgeted? • Are the staff and financial resources adequate? • Is the program duplicating other efforts?	• Is the process for selecting participants fair and equitable? • Is the program implementation delayed? • Are operational manuals being followed?	• Do we achieve the desired gender balance in participants? • Will we reach the goal of training 5,000 youth per year?	• Does participant income increase by 20%, as planned? • Do 80% of beneficiaries find a job within 3 months of graduation, as required? • What, if any, are the unintended positive or negative effects?	• Are more households becoming self-sufficient? • Are more households reaching food security?
Cause-and-Effect	n/a	n/a	n/a	• As a result of the job training, do participants have higher paying jobs than they otherwise would have? • Does including internships increase the effectiveness of technical training offered? • Does the program affect boys and girls differently?	• Does the project contribute to reducing poverty in the area? • What other impacts does this intervention have on the living conditions of the wider community?

The connection between evaluation questions and evaluation criteria. Another way to think about evaluation questions is to think about the common criteria for evaluation as originally defined by the Organization for Economic Cooperation and Development (OECD). As already mentioned, evaluations are periodic assessments of the *relevance*, *efficiency*, *effectiveness*, *impact*, and *sustainability* of our intervention (OECD 1991). Taking a closer look, we realize that *relevance*, *efficiency*, and *effectiveness* primarily relate to normative questions, while *impact* refers to causality. Questions relating to *sustainability* can be either normative (is the intervention likely to be continued after donor funding ends?) or cause-and-effect (are the observed impacts sustainable over time?). None of these is purely descriptive, though normative questions naturally incorporate descriptive ones. Table 4.2 maps each criterion to the corresponding type of evaluation question.

TABLE 4.2 The connection between evaluation criteria and evaluation questions

Criteria	Description	Details	Type of Evaluation Question
Relevance	Do the objectives match the problems or needs that are being addressed?	• To what extent are the objectives of the program still valid? • Are the activities and outputs of the program consistent with the overall attainment of its objectives?	Normative
Efficiency	Is the project delivered in a timely and cost-effective manner?	• Is the program or project implemented in the most efficient way? • What are the costs per output/beneficiary and how do these compare with similar interventions?	Normative
Effectiveness	To what extent does the intervention achieve its objectives?	• To what extent were the intended results achieved? • What are the major factors influencing the achievement or nonachievement of the objectives?	Normative
Impact	What are the positive and negative changes produced by the intervention?	• What are the higher-level outcomes resulting from the program or project? • What real difference has the activity made to the beneficiaries?	Cause-and-effect
Sustainability	Are there lasting benefits after the intervention is completed?	• To what extent do the benefits of a project continue after donor funding ceases? • What are the major factors that influence the achievement or nonachievement of sustainability?	Normative or cause-and-effect

Source: Based on OECD (n.d.)

[Tip]

Make sure you identify the audience for the evaluation and what that audience wants to know. Some evaluations may be demanded within the organization by program staff or management. Donors or policymakers may require others. Internal and external information needs may be different, leading to different evaluation questions. Involving stakeholders in defining and prioritizing your evaluation questions is therefore crucial.

Source: Adapted from Rubio (2011).

Prioritizing our questions. No type of question is a substitute for the other, though normative questions usually include and build on descriptive ones. All are looking at different aspects of the project and provide a different type of information that can be useful. If we want to focus on results, however, then cause-and-effect questions have a special appeal. In fact, if our goals are to identify promising youth livelihood interventions and to prove what effects our intervention really have, then cause-and-effect questions should be a part—if not a priority—of our program's learning objectives.

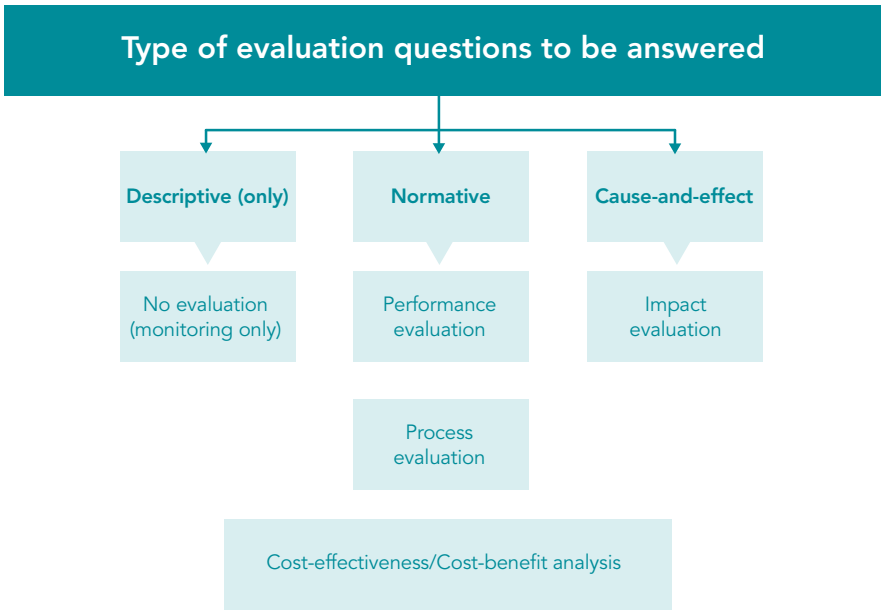
Each of these three kinds of questions—descriptive, normative, and cause-and-effect—leads to different considerations for the type of evaluation to be set up. Program managers and evaluators can allocate a potential question into one of the three types and then consider the implications of each type of question for the development of an evaluation design. Thus, by choosing a set of evaluation questions we define the menu of appropriate monitoring and evaluation tools that will allow answering them (GAO 1991).

Linking Evaluation Questions to Evaluation Design

There is no "one size fits all" evaluation template. Ultimately, the choice of the evaluation should depend on the preceding questions, not our own methodological preferences or those of the internal or external evaluator. This may seem obvious, but it is not always common practice.

Figure 4.1 provides an overview of available evaluation options depending on the type of questions we want to prioritize.[2]

FIGURE 4.1 From evaluation questions to evaluation design



No Evaluation

If a program manager requires only descriptive information about the intervention, for example, because the project is in a very early stage and the objective is to obtain some general information about how the program is being implemented, then a full-fledged evaluation may not be necessary. In that case, the knowledge obtained from monitoring may well be sufficient. Obviously, this requires the existence of a well-functioning monitoring system, with a clearly defined results chain, indicators, data collection tools, and the like (see note 3). If such a system is in place, descriptive information about the program should be available relatively easily.

Performance Evaluation

Performance evaluations assess how well program objectives have been formulated (see criteria in note 2) as well as the program's progress in achieving these objectives (Rubio 2011). They also ask whether the established results framework is appropriate; that is, whether there are inconsistencies among resources, activities, and objectives, and whether priorities or timelines should be adapted to better achieve the agreed objectives. Such evaluations can be carried out across all stages of implementation, but they are particularly common for mid-term reviews (when their focus is on learning for program management) or at program completion (when their focus is on accountability and lessons learned for future interventions). Typically carried out by an independent evaluator, performance evaluations can be implemented relatively quickly and at moderate cost because they rely heavily on desk research and selected interviews.

2 There are other types of evaluations focused on other levels of aid delivery (including sectors, themes, and aid effectiveness) that are not considered in this note. This note is limited to the evaluation of projects and programs.

Sometimes, however, performance evaluations may incorporate more extensive data collection, such as a before-and-after comparison of participant outcomes or additional qualitative tools. While useful for general quality assessment purposes, performance evaluations do not provide absolute certainty about whether the changes observed occurred because of the particular intervention.

Process Evaluation

Unlike performance evaluations, which focus primarily on the achievement of objectives, process evaluations are geared to fully understanding how a program works and seek to assess how well a program is being implemented. They determine whether there are gaps between planned and realized activities and outputs and try to understand the reasons for gaps. Building on descriptive information such as what activities are being offered and who is participating in the program (or who is not), they identify ways to improve the quality of the services offered. A process evaluation may be carried out at specific milestones as an early-warning system or may be conducted when problems such as delays in implementation or beneficiary dissatisfaction have already been detected through standard monitoring procedures (World Bank 2002). Process evaluations tend to rely on a mix of quantitative and qualitative tools, including key informant interviews, user satisfaction surveys, direct observation, and focus groups.

Impact Evaluation

Impact evaluations are the only type of evaluation to specifically answer cause-and-effect questions in a quantifiable manner. Such questions require us to determine not only whether the desired outcomes occurred but also whether those outcomes occurred *because the program was implemented.* As Gertler and colleagues (2011, p. 4) note, this focus on causality and attribution "is the hallmark of impact evaluations" and determines the set of methodologies that can be used. (Note 6 provides an overview of appropriate tools.) To estimate the causal effect of a program on outcomes of interest, any method chosen must estimate the so-called *counterfactual,* that is, what would have happened to program participants in the absence of the program. To do this, impact evaluations require finding a comparison group; that is, a group of people who, in the absence of the intervention, would have had similar outcomes to those of program recipients (Duflo, Glennerster, and Kremer 2006). This is what makes impact evaluations different from all other evaluations. As a result, they tend to require more time and quantitative skills, and they typically cost more than other evaluation types. Based on the information they provide, impact evaluations are particularly useful to inform strategic questions, from scaling up effective interventions to curtailing unpromising programs (Rubio 2011). Moreover, they increase the global knowledge base about the relative effectiveness of different types of livelihood interventions in reaching certain outcomes and help us understand which program design options (dosage, delivery channel, etc.) are most important within a specific program category.

[Definition]

A **counterfactual** refers to the estimated outcomes for program participants in the absence of the program. The counterfactual answers *what would have happened to the beneficiary had the program not taken place.*

Cost-Effectiveness and Cost-Benefit Analyses

Cost-effectiveness and cost-benefit evaluations assess monetary and nonmonetary program costs and compare them with alternative uses of the same resources and the benefits produced by the intervention (Baker 2000). *Cost-effectiveness analysis* (CEA) measures the cost per output or outcome (e.g., $300 per youth trained, $500 per job created) and compares this cost to similar interventions of our own and other

organizations. It thus answers the question about how much output or outcome we get per dollar spent (descriptive) and whether there is a gap with our expectations (normative). *Cost-benefit analysis* (CBA), in turn, weighs the total expected costs against the total expected benefits (outcomes) of an intervention, where both costs and benefits are typically expressed in monetary terms. For instance, if our program were to help 500 youth find and keep jobs or set up sustainable small businesses, we would (1) estimate the aggregate benefits in terms of higher incomes, better health, lower crime, etc., and (2) compare these benefits to the overall costs of the intervention. Since cost-benefit analysis looks at the value of the benefits achieved, it requires a credible estimate of the degree to which the program influenced the outcomes of interest, thereby making it very useful in combination with impact evaluations (for a more detailed description, see note 8). Box 4.1 provides links to examples of the evaluation types discussed above.

BOX 4.1 Examples of evaluation by type

Performance evaluations

- Human Sciences Research Council. 2007. *Mid-term Review of the Expanded Public Works Programme: Synthesis Report*. Pretoria: Southern Africa Labour and Development Research Unit, University of Cape Town; Rutgers School of Law; and ITT (UK). http://www.hsrc.ac.za/research/output/outputDocuments/5465_Hemson_MidtermreviewofEPWPsynthesisreport.pdf
- Education and Employment Alliance. 2010. *An Evaluation of Partnerships in Support of Youth Employability: Global Report*. http://www.iyfnet.org/document/1436

Process evaluations

- Miller, E., and MacGillivray, L. 2002. *Youth Offender Demonstration Project Process Evaluation*. Chapel Hill: Research and Evaluation Associates Inc. http://wdr.doleta.gov/opr/fulltext/YODP_final.pdf
- The Lewin Group, Inc. 2003. *Evaluation Design for the Ticket to Work Program–Preliminary Process Evaluation*. http://www.lewin.com/content/publications/2526.pdf

Impact evaluations

- Attanasio, O., Kugler, A. and Meghir, C. 2009. "Subsidizing Vocational Training for Disadvantaged Youth in Developing Countries: Evidence from a Randomized Trial." IZA Discussion Paper No. 4251. Bonn: IZA. http://ssrn.com/abstract=1426738
- Mensch, B., Grant, M., Sebastian, M., Hewett, P., and Huntington, D. 2004. "The Effect of a Livelihoods Intervention in an Urban Slum in India: Do Vocational Counseling and Training Alter the Attitudes and Behavior of Adolescent Girls?" Working Paper No. 124, New York: The Population Council. http://www.popcouncil.org/pdfs/wp/194.pdf

Cost-effectiveness and cost-benefit analyses

- Elias, V., Nunez, F., Cossa, R., and Bravo, D. 2004. *An Econometric Cost-Benefit Analysis of Argentina's Youth Training Program*. Washington, DC: IDB. http://www.iadb.org/res/publications/pubfiles/pubR-482.pdf
- Jastrzab, J., Masker, J., Blomquist, J., and Orr, L. 1996. *Evaluation of National and Community Service Programs—Impacts of Service: Final Report on the Evaluation of American Conservation and Youth Service Corps*. Bethesda, MD: Abt Associates Inc. http://www.abtassociates.com/reports/ccc_youth_0596.pdf
  (*Note*: This is an impact evaluation and a cost-benefit analysis combined.)

Does Our Operational Context Fit the Desired Type of Evaluation?

As noted by the GAO (1991, p. 15), "It is one thing to agree on which questions have highest priority and to choose an evaluation design. It is quite another to know whether the questions are answerable and, if so, at what costs in terms of money, staff, and time."

After formulating the right questions and identifying a potential type of evaluation, we need to assess the operational context of the intervention to understand what evaluation can be implemented within the given constraints.

Timing

Questions about *what kind* of information is needed are closely related to the question of *when* the results of the evaluation need to be available. Knowing when they need to be available determines when the information needs to be collected.

When is the Demand for Evaluation Identified?

Planning well in advance gives more flexibility in choosing an appropriate evaluation tool. For example, many impact evaluation methods need to be planned even before implementation starts. Planning an evaluation should ideally be part of the program planning (a "prospective evaluation"). In many cases, however, information needs may arise suddenly, for example as a result of sudden problems on the ground, or a request from a donor. Similarly, operational constraints, such as implementing quickly to disburse funds, may dictate the timetable for evaluation. Although these constraints are unavoidable in real life, they reduce the options for evaluation that may be available under such circumstances.

At What Stage of the Program Is the Information Needed?

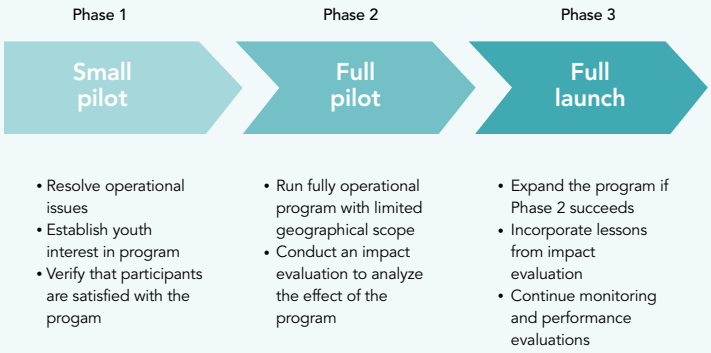
Information needs vary depending on the program lifecycle. For example, a program that has just been planned may require a cost-effectiveness analysis to help determine whether or not to implement the program. Alternatively, for a recently launched intervention, we may need to know how well program procedures are followed and whether any adjustments are necessary to guarantee successful program operation in the future (Rubio 2011). Many times, these information needs can be estimated even before the program begins, and so can the approximate timing of the evaluation.

How Long Does the Evaluation Take?

How long an evaluation takes partly depends on the methods used for collecting and analyzing data, which differ according to the type of evaluation, and on the breadth and depth desired for the particular study, which differ within each type of evaluation. In general, it is fair to assume that pure performance evaluations can be done in one to three months, since they rely heavily on desk research and a limited number of interviews. Process evaluations, in turn, can vary significantly in scope. They may be as fast as performance evaluations, but may take up to six months or longer when complex processes are being analyzed. Impact evaluations tend to be the most time consuming of all (six months to two or more years), since their methodology needs to be well planned and new data collection may be required. Cost-benefit analysis itself can take less than a month if all the necessary data are available. If information first needs to be collected, it can take much longer.

Box 4.2 illustrates at what point in a program's lifecycle different evaluation strategies are best conducted.

BOX 4.2 Lifecycle of a program and suitable evaluation strategies



Source: Adapted from World Bank (2007b, p. 18).

Phase 1. *The first pilot of an innovative and relatively untested youth livelihood intervention is about to start. What evaluation should be used?*

At the earliest stages of a program, we usually need to make sure that everything is being done as planned. Conducting an impact evaluation at this time is not recommended because the results would not reflect the true quality of the program. It is more appropriate to focus on monitoring and process evaluation until the program is fully operational and implementation issues common in setting up new programs are resolved. Qualitative data collection methods (e.g., key informant interviews, focus groups) can be particularly useful in these early stages as they may answer *why* certain elements are or are not working as intended. This initial pilot phase of the program is often referred to as a "feasibility study" to obtain "proof of concept"; that is, to see whether the program can actually be implemented as planned.

Phase 2. *The intervention has been running for one year, and early operational issues have been resolved. Monitoring shows that beneficiaries are satisfied with the program. Should we expand the program or replicate it elsewhere?*

Now may be the time for an impact evaluation. The program is up and running, and we are confident about the quality of implementation. An impact evaluation will allow us to confirm that the program is having an effect on the outcomes of interest. We can also use the impact evaluation to compare the effectiveness of program design alternatives (e.g., different combinations of activities, different intensities of activities) if we are still uncertain about specific design elements. The evaluation will also help us understand some potential unintended effects (positive or negative). As a result of the information obtained through an impact evaluation, we can make the decision on whether substantial funds should be invested in the program or not.

Phase 3. *The impact evaluation yielded very positive results overall. Do we still need to evaluate?*

Although positive results do not imply that the program would work similarly well in different contexts, we can now be fairly confident about the accuracy of our theory of change and the combination of activities. This is a good basis for expanding the program to more participants or replicating it in similar sites. Unless we want to significantly modify our intervention, another impact evaluation will probably not be necessary. However, we need to be certain that the quality of implementation remains high and that we achieve our objectives. Monitoring on all levels, including outcomes, must remain a fundamental component of our program. Moreover, independent performance evaluations in regular intervals can help verify the continued relevance and quality of the program.

Resources

Some otherwise desirable evaluation methods may not be feasible if we don't have the human and financial resources to carry them out. It is important to assess the skills and funding available in our program or organization to ensure they are in line with the needs for the evaluation we envision.

Skills

Conducting quality evaluations requires special skills that may not always exist in a program or organization. In that case, and to ensure neutrality, it is often useful to hire external evaluators. Table 4.3 summarizes some of the major skills required to conduct the various types of evaluations.

TABLE 4.3 Skills required according to type of evaluation

Skill	Description	Performance	Process	Impact	CBA
Program Design and Monitoring	• Familiarity with youth livelihood programming • Experience in program design • General knowledge of quantitative and qualitative data collection techniques • Country knowledge • A university degree in social sciences	!	!	✓	n/a
Quantitative Data Collection	• Specialized training in the design and fielding of surveys • Some knowledge of quantitative data analysis • Program management skills to build and lead a team of enumerators • A university degree in social sciences	✓	✓	!	n/a
Quantitative Data Analysis	• Specialized training in statistics or econometrics • A master's or doctorate degree in economics, public health, or related field	✓	✓	!	✓
Qualitative Data Collection	• Specialized training in implementation of qualitative techniques • A master's or doctorate degree in sociology, anthropology, or psychology	✓	!	✓	n/a
Qualitative Data Analysis	• Specialized training in coding and analyzing qualitative data • A master's or doctorate degree in sociology, anthropology, or psychology	✓	!	✓	n/a
Valuation	• Specialized training in estimating the costs and benefits of human service programs • A master's or doctorate degree in economics, public health, or related field	✓	n/a	✓	!

! *Required;* ✓ *Desirable*

Funding

The differences in scope and varying forms of data collection and analysis create a wide range of evaluation costs. Relying on desk research and key informant interviews is naturally much cheaper than designing and running new surveys with a large number of people. Performance evaluations are therefore usually the cheapest type of evaluation, while impact evaluations tend to be the most expensive (see table 4.4).

TABLE 4.4 Cost estimates for different types of evaluation

Type of Evaluation	Cost	Factors Influencing Cost
Performance Evaluation	$10,000–$30,000	Scope of the evaluation and salary of the evaluator
Process Evaluation	$10,000–$60,000	Same as performance evaluation, but often uses more data collection tools so evaluation can take longer
Impact Evaluation	$15,000–$1 million+	Cost varies widely depending on methodology used: the more data collected, the more expensive the evaluation becomes (see notes 6 and 7 for more details)
Cost-Effectiveness and Cost-Benefit Analyses	$10,000–$30,000	Depends on whether benefits have previously been measured and whether data are readily available

When all data are readily available, impact evaluations can cost as little as $15,000, though in most cases the cost will be above $100,000. Impact evaluations may seem unrealistic for programs with modest budgets. Yet, their cost may be justified if the intervention is—or will be—running for a long time or at large scale. Moreover, the implementing organization does not always have to bear the full cost of an impact evaluation, but can apply for financial assistance to carry out evaluations (see note 7 for more details on budgeting an impact evaluation).

The Political Context

Different stakeholders within and outside our organization may have potentially competing interests in terms of whether or not an evaluation should take place, the issues to be studied, the type of evaluation and its methodology, the data collection strategy, and who, if anyone, should be hired for the evaluation. All of these factors may result in pressures on the choice of an evaluation and influence the relevance and quality of the planned research. Such pressures may range from hints that certain issues should not be studied to an official disapproval from public authorities to interview certain groups of youth, families, or communities.

It is therefore important to try to understand the various interests and the political environment that exists in the specific context. The following questions will help us begin our analysis:

- What are the local political context and the distribution of power?
- What are the relationships among beneficiaries, program managers, policymakers, donors, and other stakeholders?
- What are the interests of and incentives facing each group of stakeholders to influence the conduct of the evaluation and the design of program? For example, if the program is narrowly targeted to one particular group of youth, those not included will have an incentive to influence the program and evaluation in a way that they, too, can receive benefits.
- If the evaluation shows impact, who are the potential winners and losers from any programmatic or policy reform that could derive from the evaluation?
- Will the local environment allow a rigorous and independent evaluation, and will it support the evaluators to publish their evidence-based findings regardless of political consequences?

An international NGO and its local partner in Brazil decided to conduct an impact evaluation on a youth employability-training program they were implementing jointly. After some push and pull, the eligibility requirements were agreed upon, including that the participant selection would be randomized. However, the local partner had a previous agreement with a private corporation that wanted to influence decisions about which youth would be involved in the program, which would bias any evaluation. This conflict made it unfeasible to effectively conduct the study.

Working to understand stakeholder concerns through continuous and open interaction may help us identify ways to address the pressures and competing interests and to build support for the evaluation. Moreover, it is usually helpful to bring in external evaluators who, in addition to contributing a specific skill set, may have an easier time maintaining their independence.

Types of Programs That Usually Justify an Impact Evaluation

Although performance and process evaluations and cost-effectiveness analyses can be part of every program, impact evaluations and cost-benefit analyses should be applied more selectively. According to Gertler and colleagues (2011), the additional effort and resources required for conducting impact evaluations are best mobilized when the program is (1) strategically relevant and influential, (2) innovative or untested, and (3) replicable.

Strategically Relevant and Influential

How important would the results be for informing future programs, policies, or policy dialogue? If the stakes of an intervention are high—for example because a program requires substantial resources and covers, or could be expanded to cover, a large number of people—then an impact evaluation should be considered. This may apply to new initiatives as well as to existing programs when we need to make decisions about their continuation, expansion, or termination. In fact, even an expensive impact evaluation can be highly cost-effective since its findings may help to produce important improvements in program performance. In fact, in the case of large initiatives, even minor improvements may result in considerable savings to the implementing organization (World Bank 2009).

Innovative or Untested

What is the current state of evidence or knowledge on the proposed program's impacts? If little is known about the effectiveness of the type of intervention, globally or in a particular context, an impact evaluation can add powerful knowledge to our organization and the entire field. This is the case for most youth livelihood programs for which the evidence base is still slim (see box 4.3). In the case where no or only little evidence is available, it is usually recommendable to start out with a pilot program that incorporates an impact evaluation. Even if there is existing evidence about a particular type of intervention, an impact evaluation may be still be warranted if the program is implemented in a different context or if it includes innovative aspects that have not been previously tested.

[Online Resource]

Knowledge gaps and potential research questions for impact evaluation

http://www.iyfnet.org/gpye-m&e-resource3

BOX 4.3 Knowledge gaps in youth livelihood programming

Although the following generalizations must be interpreted with caution, we believe existing evidence on youth livelihoods programs appears to be particularly weak in these areas:

Types of programs: Most evaluations exist in the area of training and skills development, while evidence on all other types of interventions such as subsidized employment for youth, employment services, youth entrepreneurship, youth-inclusive financial services, and targeted programs for excluded groups is relatively scarce.

Design Features: Little is known about the relative effectiveness of program alternatives. Within each type of program, what is the effect of adopting different program components, different pedagogies, dosage, and delivery channels?

Context: Evidence of youth livelihood programs is particularly scarce in the Middle East and North Africa, Asia, and sub-Saharan Africa. Moreover, more evidence is needed regarding what interventions and design features are better suited for rural versus urban contexts, informal versus formal settings, or in postconflict and fragile-states environments.

Beneficiaries: How do different types of programs affect young people differently by age group, gender, level of education and socioeconomic background? What works best for disadvantaged groups? And what are the positive or negative spillover effects of livelihood interventions on peers, families, and communities?

Outcomes: What are the effects of livelihood programs not only on employment and labor market outcomes, but also on risky behaviors, civic engagement, family formation, mental health, and the like? Furthermore, evidence on long-term effects of most interventions is virtually inexistent.

For a review of the existing evidence, see the Youth Employment Inventory (www.youth-employment-inventory.org) and Cunningham, Sanchez-Puerta, and Wuermli (2010).

Replicable

To what extent and under what circumstances could a successful pilot or small-scale program be scaled up or replicated with different population groups? If an intervention design is extremely specific and targets a narrow and particular context, then a process evaluation that would contribute to a smooth implementation would probably be sufficient. If, however, the program can be scaled up or can be applied in different settings, then an impact evaluation is an important step in providing the justification for a program to be replicated.

Table 4.5 presents a table summarizing the evaluation types.

TABLE 4.5 Overview of main evaluation types

	Performance Evaluation	Process Evaluation	Impact Evaluation	Cost-Effectiveness and Cost-Benefit Analyses
What are the main questions answered by this type of evaluation?	• Do programs have clear objectives? • Is the program design appropriate to achieve the objectives? • To what extent have program objectives been achieved? • Do priorities need to be changed?	• Are adequate resources and systems (management, information, etc.) in place? • Is the program being implemented according to design? • What are the actual steps and activities involved in delivering a product or service? • What do beneficiaries or other stakeholders know or think about the program?	• How have participants' well-being changed as a result of the intervention? • Are there any unintended consequences, positive or negative, on program participants?	• Are program costs justified compared with similar interventions? • Are aggregate program costs justified in terms benefits achieved?
When can this evaluation be conducted?	It may be conducted at early stages of implementation, for mid-term review, or at program completion	It may be conducted at any time, once or regularly, to confirm that implementation is on the right track or to understand specific operational concerns	It should be designed during the planning of a program, but the final results will typically not be available till after the program (phase) has been completed	It is commonly conducted during an ex ante analysis to determine whether the program is worth implementing or continuing, or after the program is completed to determine the final costs
How long does it take?	1–3 months (more if before/after analysis is included)	1–6 months	• At least 6 months (retrospective evaluation) • 12–24 months (prospective evaluation)	1–3 months
What data collection and analyses are required?	Desk review of existing documents and selected field visits, possibly complemented by monitoring data analysis, beneficiary and stakeholder interviews, mini-surveys, focus groups, etc.	A mix of interviews with program staff and clients, user satisfaction surveys, record review, direct observation, focus groups, and analysis of monitoring data	Statistical and econometric analysis of survey and administrative data, ideally combined with qualitative data analysis	Desk review of existing program documents and relevant literature as well as key informant interviews
Who carries out the evaluation?	Usually independent evaluator (but can also be internal)	Internal or independent evaluator	Independent evaluation team, including lead evaluator, field coordinator, survey firm	Independent evaluator (can be the same as for performance or impact evaluation)
What skills are needed?	Program analysis, possibly simple quantitative methods	Process analysis, quantitative and qualitative methods	Statistical and econometric analysis, possibly qualitative methods	Valuation and economic analysis of program costs and benefits
What are the costs?	$10,000–$30,000	$10,000–$60,000	Cost can range from $15,000 to $1 million or more, depending on the size and complexity of the program	$10,000–$30,000
What programs are best suited for this evaluation?	Every program	Every program	Programs that are: • Innovative and untested • Strategically relevant and influential • Replicable	• Cost effectiveness: Every program • Cost-benefit: Same as impact evaluation

Source: Adapted from Rubio (2011).

Key Points

1. Our learning objectives are the point of departure for any evaluation. This requires formulating evaluation questions across all levels of the results chain and prioritizing the most relevant ones. In general, evaluation questions can be *descriptive*, *normative*, or *cause-and-effect*.
2. The choice of the evaluation strategy depends on the evaluation questions. Purely descriptive information needs may not require an evaluation, and monitoring may suffice. Normative questions are most commonly answered through *process* or *performance evaluations*. If cause-and-effect questions are the priority, *impact evaluations* are needed. *Cost-effectiveness* and *cost-benefit analyses* answer whether the costs involved in an intervention are justifiable.
3. Only impact evaluations—those that can construct a valid counterfactual—allow us to *prove* whether a program has been successful and to generate knowledge that can potentially be generalized beyond the intervention itself. This differentiates them from all other evaluations types and makes them a key instrument for evaluating youth livelihood interventions.
4. Choosing an appropriate type of evaluation depends on the operational context. It is therefore crucial to understand whether the costs in terms of money, staff, and time for each evaluation are appropriate for a given intervention.
5. Since impact evaluations tend to be the most resource intensive type of evaluation, they should be applied selectively to answer strategic questions or to assess innovative pilot interventions testing an unproven, but promising, approach.

NUSAF Case Study: Deciding Whether to Do an IE

Evaluation Questions

The primary learning objective for NUSAF was to estimate the causal impact of participation in vocational training programs on economic livelihoods and social integration. The questions of interest for NUSAF were whether the Youth Opportunities Program helped to:

- increase the number of businesses started
- lower the levels of unemployment
- increase the number of hours working for pay
- improve community integration and decrease conflict
- reduce poverty
- increase psychosocial well-being

Given the cause-and-effect nature of these questions, an impact evaluation was the evaluation method of choice.

NUSAF was also interested in the effects of the program on local training organizations. Since this cannot be easily identified through an impact evaluation, it was decided that this would be part of the monitoring of the Youth Opportunities Program.

Operational Context of NUSAF

Given that NUSAF was a World Bank–funded program with strong support for the impact evaluation from the Government of Uganda, the operational context for an impact evaluation was favorable.

- Timing: The evaluation strategy was planned from the outset of the program. This allowed for the necessary flexibility to plan a rigorous impact evaluation.
- Resources: The necessary resources could be earmarked and a qualified external team hired to conduct the evaluation.
- Political context: Making the evaluation a priority from the beginning fostered stakeholder dialogue and support.

Features of NUSAF that would Justify an IE

The Youth Opportunities Program was a large cash grant program designed and implemented by the government of Uganda. The size and influence of the program, combined with the expectation of rerunning the program in the future, suggested that evaluating the program was an excellent way to increase local and worldwide knowledge of cash grant training programs. Although these types of programs are increasingly implemented, they are generally untested. In addition, the fact that the program was implemented by the government suggested that such a program is scalable and could be replicated in other countries.

Source: Based on Blattman, Fiala, and Martinez (2011).

Key Reading

Imas, L., and Rist, R. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations,* Washington, DC: The World Bank. (Chapters 6 and 7 are relevant to this note.)
http://books.google.com/books?id=NEsg-BtinIsC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Rubio, G. 2011. "The Design and Implementation of a Menu of Evaluations." PREM Notes, The Nuts and Bolts of M&E Systems, No. 6. Washington, DC: The World Bank.
http://siteresources.worldbank.org/INTPOVERTY/Resources/335642-1276521901256/premnoteME6.pdf

Notes