

The Importance of Evaluating Youth Interventions

“...the dearth of rigorously evaluated youth-oriented programs and policies can undermine (their) credibility...”

World Development Report 2007

As noted in the World Bank’s 2007 *World Development Report*,¹ very little is known about the quality or impact of youth interventions because very few programs have been carefully evaluated. As a result, program officials, policy makers, and the community in general have only limited information about how best to implement programs for at-risk youth, about which program features are the most effective, and even which programs yield the greatest benefits. Therefore, in practice, most youth programs are selected and designed on the basis of anecdotal evidence, interviews with potential participants, and the program designer’s personal predilection. Yet, for youth programs to succeed and yield lessons for policy makers and other stakeholders, they must be carefully planned, well implemented, and rigorously evaluated, and the lessons learned from those evaluations must be widely disseminated.

Rigorous, well-controlled evaluations can be expensive. Program officials, especially those managing small projects on limited budgets, may be reluctant to spend their scarce resources on evaluating the program rather than on providing more essential services or more training or compensation to their underpaid and dedicated staff. Unfortunately, failing to invest in evaluation can be shortsighted and self-defeating because it means that lessons cannot be learned or applied to ongoing or other programs. Without evaluations, it is impossible to understand the full impact of our interventions on those whom they are designed to serve. We must be able to understand two things: (i) are we doing the right thing, and (ii) are we doing it well?

The terms “monitoring” and “evaluation” are often used jointly and are abbreviated as M&E. However, they are quite different. Monitoring can be thought of as measuring the process of an intervention, while evaluation involves assessing its impact. Both are essential to the success of a program. Ideally, both monitoring and evaluation should be carried out as an integral part of the program from the outset rather than as an afterthought or a way to showcase its success. In fact, it is extremely difficult to evaluate a program if the evaluation begins only as the program is coming to an end. While such ex post evaluations can be used to inform the design of future programs, their conclusions come too late to be used to modify and improve the program that has just ended.

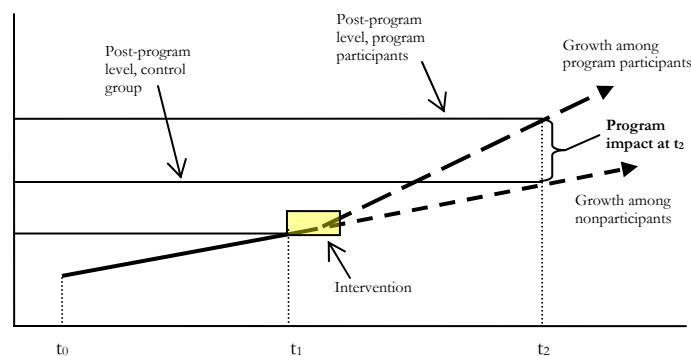
The goal of impact evaluation is to understand what consequences the intervention has had on the program’s participants. It presumes that the program was implemented as intended—for instance, that employees have performed the tasks for which they were hired and the services were delivered as expected. However, this may not always be the case. If there are any problems in the program’s implementation, then this will confound the impact evaluation. If an evaluation finds that a program had no positive results, this may be because the program was ineffective or because it was poorly implemented. This is the purpose of process evaluation, or monitoring.

Process evaluations should be carried out during each phase of the program’s implementation to assess whether the program is actually carrying out its intended activities. Problems inevitably arise and activities are rarely implemented exactly as planned. The job of the evaluators is to understand what has happened in the program in practice, and clearly it is better to obtain this information during the course of implementation rather than relying solely on program documents after the fact.

What Is an Impact Evaluation and What Does It Measure?

Impact evaluations assess any changes in the well-being of individuals, households, communities, or firms that can be attributed to a particular project, program, or policy. In addition to increasing the accountability of policy makers and program managers, impact evaluations yield lessons that policy makers can then use to improve ongoing programs and ultimately to allocate funds more effectively among programs. The key question that impact evaluations seek to answer is what would have happened to those participating in the intervention if they had not participated in the program. Because it is not possible to observe the same group experiencing both the presence and the absence of an intervention, the key challenge is to develop a counterfactual—in other words, a group of people who are as similar as possible in both observable and unobservable dimensions to those who participated in the intervention. Comparing the outcomes experienced by the participating group with those experienced by the control group makes it possible to establish definitive causality (in other words, to attribute any observed changes in welfare to the existence of the program) while removing any potentially confounding factors. As depicted in figure 1, the program's impact equals the difference between the observed outcome and an estimate of the outcome had no program been available.

Figure 1. A Program's Impact—Levels of and Changes in Outcomes



Those in the control (or comparison) group should be as similar as possible to those in the participating group in every respect except that they do not participate in the program. It is important to plan for this in the evaluation, ideally by randomly assigning the eligible population to either the treatment and control groups (for example, by lottery) or by using some rule to select people to be in those two groups as long as this rule does not also affect the goods or services delivered by the program.² With clear and thorough planning, it should be possible to avoid any ethical concerns about denying program participation to any given group of people. In cases where the program will offer its services to different groups of participants sequentially, those who participate later can be used as a control group for the earlier participants, providing that the order in which the services are offered is random.

It is common to compare outcomes for the same individuals both before and after they have participated in the intervention, but this can often produce misleading results. This method produces a “counterfeit counterfactual”—in other words, one that does not accurately reflect what would have happened to the individual during the same time period in the absence of the program. This can yield an incorrect estimate of the program's impact and may cause analysts to come to erroneous conclusions.

How Do We Identify Relevant Outcomes and Indicators?

Ultimately, any program should improve the lives of the young people who participate in it, whether by reducing their risky behavior, improving their academic achievement and progress, or helping them to obtain and retain career jobs in growth industries. The outcomes and indicators to be evaluated should be chosen based on the specific goals and activities of the program in question. Outcomes are observable characteristics of the program's beneficiaries and, whether they materialize in the short or the long term, they should be associated with measurable indicators. In devising an evaluation strategy, program planners should bear in mind the difference between outcomes and outputs. Outputs are the goods and services provided by the program to the beneficiaries, whereas outcomes are the changes observed in those beneficiaries as a result of the program. For instance, one measure of the output of a training program is the number of young people who actually receive the training, while one of the direct outcomes is any changes in the employment status of those young people. Program planners need to keep in mind both direct and indirect anticipated outcomes when deciding which indicators to measure. For example, a program that increases the likelihood of employment and income among youth may also have an effect on the participants' risk behavior. Also, they should be open to the possibility that the program may produce some completely unexpected outcomes. For example, in an evaluation of a program that aimed to reduce dropout rates by providing students with uniforms, Duflo et al. (2006) found that the program also—unexpectedly—reduced pregnancy rates among the program's participants.³

Youth interventions are often difficult to evaluate as they are generally diffuse in nature and scope and have outcomes that cover several different sectors,⁴ making it difficult to isolate the impact of each individual component. Difficult as this is, it is essential to do so because information on each component of a youth intervention helps policy makers to identify the optimal combination of youth investments. Comprehensive interventions may be preferable to narrow sectoral ones, but we can't always be sure which components in what combination are necessary to achieve the desired outcomes.

Similarly, interventions aimed at children and young people can have consequences that only materialize well into the future or are not directly related to the intervention itself. For example, high-quality early childhood development immediately improves children's health, nutrition, and education but can also reduce delinquency, teen pregnancy, and incarceration and increase employment status and earnings among those children in their later youth and adulthood. This one intervention yields benefits in at least four different areas—health, education, crime and justice, and labor—during the participants' lifetimes.

Young people are often more difficult to study than adults. They are an unusually mobile group, moving from one place to another without warning and making use of many different services from various organizations and localities. Therefore, program planners need to make some provision for tracking individuals in the evaluation sample. Also, interviewing minors raises both logistical and ethical concerns. Because minors cannot be expected to give informed consent to participate, they should not be interviewed without the consent of their parents; however, this must be balanced by the need to maintain and protect their privacy. These challenges must be addressed, for example, by sharing the questionnaire (but not the answers) with parents to obtain consent for a private interview with the young person. These steps ensure that the evaluation is carried out in an ethical way and that the program's outcomes are adequately measured.

What Are the Elements of a Good Evaluation?

Good evaluations of youth development programs should at minimum include the following features:

- a description of the intervention, including its planned activities, methods of delivery, duration, and expected outcomes
- clearly defined goals and indicators that directly reflect those goals (outcome indicators)
- clearly defined indicators of how well the program is being implemented (process indicators)

- a baseline survey of the population being studied, both treatment and comparison (control) groups, which ideally should be conducted before individuals are assigned to each group
- the identification of a comparison group that is as similar as possible to the treated group in all key observable characteristics (random assignment is preferable but is not always possible)
- periodic follow-up surveys both to monitor process and to evaluate outcomes
- a careful accounting of all the program's costs (which is often overlooked but is essential to understanding whether the program is successful in economic terms as well as in strictly outcome terms)

In sum, impact evaluations are essential for testing the effectiveness of interventions that are designed to improve the lives of young people by establishing whether or not there is a causal link between the intervention and its anticipated outcomes. By providing critical feedback on which interventions work and which do not, impact evaluations can improve the design and outcomes of all future youth programs.

Endnotes

1. World Bank. *World Development Report 2007: Development and the Next Generation, 2007*. Washington, D.C.: World Bank.
2. Proper evaluation requires that the process of assignment to either treatment or control group does not also affect the intervention that is provided to the treatment group; this is the requirement of *unconfoundedness*.
3. Duflo, Esther, Pascaline Dupas, Michael Kremer, and Samuel Simezi. 2006. "Education and HIV/AIDS Prevention: Evidence from a Randomized Evaluation in Western Kenya." Background paper for the *World Development Report 2007*. World Bank, Washington, D.C.
4. World Bank. 2007. "Evaluating Youth Interventions." *Youth Development Notes* 2(5).