

NOTES

YOUTH DEVELOPMENT



Evaluating Youth Interventions

Youth development projects aim to improve the lives and livelihoods of young people around the world. Interventions for youth are often multi-sectoral in nature, ranging from job- and life-skills development to programs for better health and nutrition. Rigorous impact evaluation is key to producing the knowledge base required by policymakers and practitioners to choose among different options, and implement the most cost effective projects. This note outlines some approaches to producing evidence of what works in the context of youth development projects, and looks at expanding the set of outcome indicators to more fully capture the effects of these projects on the welfare of young people around the world.

The challenge of incorporating impact evaluation into youth projects

... few solid evaluations of youth programs in developing countries unambiguously identify the causality from policy to program to effect ... many (youth) programs fall into the promising but unproven camp ...

—World Development Report 2007 (WDR07),
Development and the Next Generation

The evaluation of youth development projects poses special challenges, both conceptual and the logistical, particularly if they are multi-sectoral. Youth development projects are often diffuse in nature and scope, extend over a long period of time, vary widely across applications, and have outcomes across a range of sectors. These challenges must be addressed in an evaluation to ensure that causality is well established and that outcomes are adequately measured. For example, when looking at the effects of a youth intervention on employment, we know that obtaining a job is also a function of health and schooling. Alternatively, we may want to know if discouraging girls from early marriage is more effective when girls are in school. We can use impact evaluation to isolate the impact of any one component of a youth intervention, test the optimal combination of interventions in different contexts, or look at potential spillover effects across populations.

Although evaluations are usually narrowly defined, the multifaceted nature of youth transitions means that interventions can have unexpected outcomes. Recent evaluations of youth programs use a wider range of outcome indicators to capture these different impacts. For example, interventions focusing on education have also been shown to affect risky behavior: conditional cash transfers may reduce alcohol and smoking, early child development may reduce crime, violence and teen pregnancy, and additional schooling may lower the incidence of teen pregnancy and HIV/AIDS (1).

When considering the evaluation of youth projects there are also a number of logistical considerations to keep in mind. Young people are exceptionally mobile, and it is important to make provisions to track individuals in the evaluation sample over time. Similarly, when interviewing minors, issues of parental consent are important, while at the same time providing the necessary safeguards to protect the young person's privacy. The remainder of this note outlines key aspects of an impact evaluation design, and considers a number of issues that are unique to the evaluation of youth development projects.

Elements of effective impact evaluation design

An impact evaluation design allows us to isolate the effect of a youth development program on a given outcome, or to test the optimal combination of interventions in different contexts. Impact evaluation helps us understand “what is the effect of

X on Y?” For example: what is the effect of a youth training program on employment? Ideally, this would be estimated by comparing the employment status of an individual with and without the training program at the same point in time. Given that we will never observe the same individual in two different states at the same time, impact evaluation must attempt to construct a plausible alternative for comparison, or *counterfactual*: that is “what would have happened to the youth without the training program?”. As depicted in Figure 1, the program impact is the difference between the observed outcome (the continuous line) and an estimate of the outcome had no program been offered (the dashed line—i.e. the counterfactual).

Counterfactuals are estimated using *control* groups, that is, a group of individuals who do not participate in a program. Identifying a valid counterfactual is critical to good impact evaluation. Typically, identifying the control group entails determining why one group of individuals was treated and the other was not. Doing this retrospectively can be challenging, especially if the two groups were not randomly selected. There may be unobserved differences between those in the treatment group and those in the control that affect the outcome, and this will confound measurement of the impact of treatment. When working prospectively in the planning phase of an intervention, one can either explicitly select—or preferably randomly assign—individuals into treatment and control groups.

Identifying a control group

By knowing **Who** is eligible, **Where** the intervention will go and **When** the intervention will be delivered, we can identify a control group that can be used to estimate a valid counterfactual for the estimation of a program's impact. By working within the context of program planning and operations, we can minimize the ethical concerns that may arise by denying treatment to the control group. For example, in the early stages of program implementation, budgetary and logistical constraints usually limit the number of eligible youth or groups that can receive the intervention. Everyone who is eligible will receive the intervention, just not all at the same time. When a project can't go everywhere at the same time, managers must use some rule to determine where the project will begin and how it will scale up. Provided we understand the scaling-up rules, the individuals who do not receive the intervention in the early stages can provide valid controls for those who do.

Suppose that 100 localities are identified as the areas of highest youth unemployment, but budgetary and logistical constraints only permit coverage of a training program in 50 localities during the first year. One fair way to assign the benefit is to give each locality an equal chance of receiving the benefit; for example, by using a lottery to select the localities that will receive the intervention this year. In that case, the localities that will receive the program in the future serve as a counterfactual control group to the localities that receive the program in the first year. On average, there will be no differences (observed or unobserved) between the two groups before the program is rolled out, and assignment to treatment and control groups is by design unrelated to any characteristics of the localities.



Therefore, differences in outcomes between the two groups following program implementation can be attributed to the causal effect of the program, since the only difference between the groups is that one received training and the other did not.

When randomization is not possible, other good options for identifying valid control groups can be found using program eligibility rules. For example, interventions are often targeted to groups or individuals that meet certain criteria, such as poverty: those with incomes just below the threshold are eligible, while those just above are ineligible. Arguably, pre-intervention differences between two individuals with incomes on either side of the threshold are very small, and differences in outcomes after the intervention can be largely attributed to the intervention itself.

Identifying relevant outcomes and indicators

Program activities produce *outputs*, and the resulting changes observed in the beneficiaries is the *outcome*. For example, in the case of vocational training, an outcome is employable skills,

Box 1. Outcome indicators

Let's consider a job training program in a post conflict setting characterized by inter-ethnic conflicts, designed to increase employment skills, as well as reduce inter-ethnic conflict, and risky behaviors, while promoting tolerance, and civic participation. Direct and indirect expected outcomes to consider are as follows²:

Direct Outcomes

1. Competencies in training program skills (e.g., basic business, finance, and accounting knowledge)
2. Business activities (e.g., size, profitability, employment, youth employed, revenues and revenue growth, return on investment, etc.)
3. Credit and capital activities (sources of credit and equity raised)
4. Economic status (e.g., employment, wages, days employed, average earnings, asset ownership)
5. General skill competency (e.g., literacy test score, numeracy test score, English language skills test score, computer skills test score)

Indirect outcomes

6. Risky behaviors (e.g., school absenteeism or dropout, inactivity - neither in school nor in work, substance use, early sexual initiation, unsafe sexual practices, criminal behavior)
7. Violent behavior (e.g., hostility, participation in fights, carrying a weapon, participation in riots or violent protest, attitudes towards the use of violence)
8. Ethnic and religious attitudes (e.g., ethnic and religious tolerance, ability to articulate another ethnic group's point of view)
9. Political and community participation (e.g., membership in community groups, civic participation, participation in peaceful protests, political extremism)

2. The outlined outcome indicators have been defined for the impact evaluation of the Post Conflict Fund for Kosovo Youth Development (in particular for the *Business Development for Young Entrepreneurs* component). For further information, please contact Silvia Paruzzolo (sparuzzolo@worldbank.org)

Box 2. A Counterfeit Counterfactual

A commonly used counterfactual that may produce misleading results is the comparison of the *same individual before and after the intervention*. For example, say that the youth employment rate in a given community is 30%. A training program enters the community and trains young people in employable skills. After the training, it is observed that youth employment has increased to 60%. Was the training program a success? If, for example, positive economic growth also affected youth employment over the same period, the contribution of the program to the increase in employment may be only minor or zero. On the other hand, in the case of an economic recession over the same period, youth employment would have been much lower, and so the simple change in employment rates will underestimate the true impact of the program. Thus, just comparing an outcome for the same individual before and after the introduction of a program may lead us to erroneous conclusions about a program's success.

while an output is receiving the training. Outcomes are observed characteristics of the beneficiary, and not of the program; and whether short or long term, should have measurable proxies or indicators.

Any evaluation should have some idea how and why the intervention leads to the expected outcomes. Impact evaluation should include a review of program implementation, or a "process evaluation," to understand this chain of events. Some programs do not work because planned activities are not carried out as planned. When a program is poorly implemented, there may not be a great need to delve deeply into all the hypothesized causal links in the chain.

The selection of relevant outcome indicators is a critical step in the design of an impact evaluation, and should be guided by the logical framework that connects program activities to direct outcomes. These direct outcomes may in turn lead to other, more indirect outcomes. Examples of measurable outcomes are described in Box 1.

Frequently, the diversity of objectives of an intervention makes selecting valid indicators difficult. For example, projects that aim at providing skills may have a direct impact on competencies and employment, but may also have equally important indirect impacts on reducing risk behaviors. It is necessary to anticipate both direct and indirect outcomes, keeping in mind that direct outcomes may not always be the most relevant from a social and policy perspective.

Collecting data for the evaluation of youth development programs

The success and reliability of an evaluation rests heavily on the quality of the data used. Since primary data collection can represent the lion's share of an evaluation budget (2), data collection strategies need to be carefully considered. Samples should be representative of the target population and include sufficient sample sizes to detect the desired effect size (power calculations can help determine required sample sizes). Survey methods should also be carefully considered,



especially when collecting sensitive data. For example, in the case of risk behavior outcomes, experience suggests that audio or computer assisted self-interviewing (CASI) can help young people to discuss candidly a range of sensitive and potentially embarrassing subjects.¹ Compared to face-to-face interviews, both Audio-CASI and self-administered questionnaires have been shown to provide better prevalence estimates of youth risk behavior in culturally conservative societies, especially for particularly stigmatized or legally sanctioned behaviors (4).

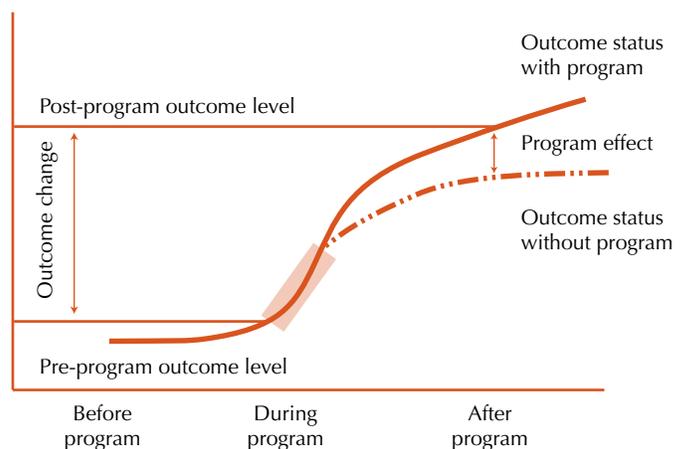
Measuring all the program's costs and benefits

Finally, for the purpose of informing policy decisions, an evaluation is not complete until one considers the costs of the program. Impact is only one criterion for program selection. The program must be effective in both a statistical or clinical sense and an economic sense; the most effective program may not be the most cost-effective one. An intervention may have a profound impact on participants, but if it is extremely expensive, it may not make sense to implement or continue it. It may be preferable to select a program that has a smaller impact, but is much less costly.

This highlights the importance of measuring all of the costs and all of the benefits of a given program. Some benefits, and some costs, may not become apparent until some time after the intervention. And as noted above, a program's benefits may be unrelated to its original goals. Similarly, the program may have social costs as well as financial costs, and all of the resources used will have *shadow* costs—that is, even volunteers to a program have potential alternative uses, and it is the job of the evaluation to determine whether the intervention presents the best feasible use of these scarce resources. Finally, the program's *average* costs may not be a good indicator of *marginal* costs—that is, what it will cost to scale up the program.

1. With AUDIO-CASI, prerecorded questions are presented through headphones and on a computer screen. Answers are given using numbered keys on a computer keyboard. This obviates the need for interviewers, but given the costs of the technology may not reduce overall survey costs.

Figure 1. Outcome level, Outcome change and program effect (impact)



Source: Adapted from Rossi et al. (2004)

References and Recommended Reading

- (1) World Bank. 2006. *World Development Report 2007: Development and the Next Generation*. New York: Oxford University Press. Also downloadable at: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,menuPK:384336~pagePK:149018~piPK:149093~theSitePK:384329,00.html>.
- (2) Rossi P.H., Lipsey M.W., Freeman H.E. 2004. "Evaluation – A systematic approach" (7th ed.), Thousand Oaks, Calif.: SAGE Publications.
- (3) Baker J. 2000. "Evaluating the Impacts of Development Projects on Poverty: A Handbook for Practitioners." Washington, D.C.: The World Bank.
- (4) MacMillan H.L. 1999, "Computer survey technology: a window on sensitive issues", *CMAJ* Specialty Spotlight.
 - Duflo E., Glennersterzand R., Kremer M. 2006. "Using Randomization in Development Economics Research: A Toolkit". Also downloadable at <http://www.povertyactionlab.com/papers/Using%20Randomization%20in%20Development%20Economics.pdf>.
 - The Poverty Lab webpage (<http://www.povertyactionlab.com/>).
 - The World Bank impact evaluation webpage (<http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,menuPK:384336~pagePK:149018~piPK:149093~theSitePK:384329,00.html>).