



NOTE 8: Increasing the Relevance of the Impact Evaluation

Although a standard quantitative impact evaluation in and of itself can be of great value to our program and organization, there are a variety of options to further enhance the quality of the analysis and increase the relevance of the results. First, impact evaluations can answer a variety of research questions, including, but well beyond, average program effects. Second, when combined with qualitative tools, impact evaluations can be much more informative than when relying purely on quantitative methods. Finally, the results of an impact evaluation can be leveraged for further analysis, for example to weigh total program benefits with total program costs (through cost-benefit analysis). This final note provides a brief overview of these tools so practitioners can make the most of their impact evaluation.

Measuring a Variety of Impacts

The basic impact evaluation answers the question “Did the program work”; that is, did it affect the outcomes of interest as defined in our program and learning objectives? The question of whether the program as a whole had an impact is an important one, but it is by no means the only question we may want ask.

First, it may be useful to have a more nuanced picture of the program’s actual impact. This includes obtaining a better understanding of the following questions:

- Do outcomes vary across different groups of beneficiaries (e.g., boys benefit, but girls do not)?
- What is the short-term versus the long-term impact of the intervention?
- Does the program have positive or negative spillover effects? Are there any intended or unintended outcomes beyond the actual target group?

Second, we may also be interested in testing crosscutting designs (CCDs), testing how the effectiveness of our program changes as we modify the design. CCDs investigate the following questions:

- Is one program design more effective than another? We may want to compare alternative interventions (providing start-up grants versus start-up loans for young entrepreneurs, for example), or test the most effective combination of program components (training alone, training plus internship, and training plus internship and mentoring).
- What is the most effective dosage of program activities? For example, should we provide 20, 50, or 100 hours of training?

If properly designed, impact evaluations can provide answers to these questions, though it will be difficult to answer all questions with a single impact evaluation. Because each intervention will have different priorities and learning objectives, we can design the impact evaluation to answer the questions most relevant to our program. By addressing a broader set of questions, we can improve the relevance of the evaluation findings. Yet, it is also important to understand that additional data are required to evaluate elaborate questions and crosscutting designs (see table 8.1).

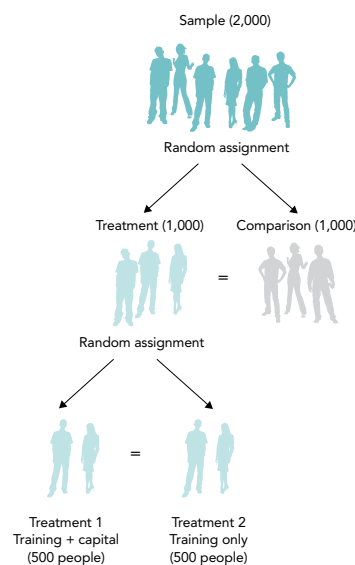
TABLE 8.1 Categories of impact evaluation questions

Question	Description	Additional Data Requirements	Sample Evaluation Result and Interpretation
What is the overall program impact on outcomes A, B, and C in group X? In context Y?	This is the standard impact evaluation question.	n/a (standard data collection based on the method chosen)	The average impact of the training program on the income of youth is +\$20 per month. The program has a positive impact on income.
Do the outcomes vary across population groups?	Interventions often affect groups differently (heterogeneity of impacts). Measuring only average impact may hide these differences, so we need to break down impacts by population group.	- Sociodemographic information of participants and comparison group (age, gender, income level, etc.) - To be able to disaggregate the results, the number of people covered by the evaluation (the sample size) needs to increase with each category of information that is to be analyzed	The average increase in income is \$40 for boys, and \$0 for girls. Older youth benefit more than younger youth (\$30 versus \$10 on average). Therefore, the program is not equally effective for all participants. We need to understand why groups benefit to a different extent and possibly adapt the program's targeting and design to accommodate particular groups.
What is the short-term versus the long-term impact of the program?	The change in outcomes may not be constant over time. Short-term effects may vanish, while long-term effects may not be visible for years after the intervention ended.	Data over an extended period of time (in practice, it often means following treatment and comparison groups for several years)	At the end of the program, we observe an average monthly income for participants of -\$5 (a loss) compared with the controls. Two years after the program, the average increase in monthly income for the treatment group is \$20. Those who participated in the training were not able to work as much as their peers during the training, so they lost income. Over time, however, the training paid off and participants were able to secure incomes higher than those of their counterparts who did not participate. Looking only at short-term outcomes may provide misleading results.
Does the program have spillover effects?	The program may have indirect effects on nonparticipants (positive and negative).	Data beyond the treatment and comparison group, to include family or community members	Not only do participants have a \$20 higher average income, their neighbors also experienced a \$5 increase. Participants apparently passed on the knowledge to others.
Is program design A or program design B more effective?	There is often ambiguity about the best possible program design. Questions can relate to comparing alternative interventions or combinations of program components.	- Several treatment groups (one receives design A, one receives design B, etc.) - The number of people covered by the evaluation needs to be large enough to be able to create more than one treatment group as well as a comparison group.	The average increase in income is \$5 for those who received training and \$30 for those who received training and an internship. Thus, providing practical work experience in addition to training appears to significantly improve impact.
What is the most effective dosage of the intervention?	More is not always better; finding the right balance of how much service to provide is important to maximize impact on the one hand and minimize costs on the other.	- Several treatment groups (one receives design A, one receives design B, etc.) - The number of people covered by the evaluation needs to be large enough to be able to create more than one treatment group as well as a comparison group.	The average increase in income is \$0 for those who received one month of training, \$20 for those who received three months, and \$20 for those who received six months. Although 1 month of training was insufficient, six months of training had no additional benefit compared with three months of training. The optimal length of the training seems to be about three months.

CCDs help identify more than just the overall program impact; they also evaluate specific program features and why these do or do not work. For example, a program may provide vocational and entrepreneurial skills training, such as carpentry or tailoring, along with small start-up capital for businesses. The provision of cash grants could be expensive or politically difficult, and so the program director may wonder if the start-up capital is necessary, or if participants are able to implement their training without the capital. A CCD can help determine the best program design in this case.

CCDs require at least two treatment groups that receive different combinations or dosages of the program. These two groups can then be compared at the endline, and the difference between the two groups is the impact of the specific design. Using the example above, the program may conduct an evaluation in which a sample of 2,000 participants is randomly assigned into treatment and comparison groups. The treatment group can then be further randomized into two treatments. In treatment 1, the training and start-up capital are provided. In treatment 2, only the training is given (see figure 8.1).

FIGURE 8.1 Outline of an impact evaluation with a crosscutting design component



The impact of providing start-up capital can then be determined by comparing those in treatment 1 to those in treatment 2 at the endline. The impact of the training is determined by comparing those in treatment 2 to those in the comparison group.

Using Mixed-Methods Approaches

It is important to keep in mind the limitations of quantitative impact evaluation methods. If used in isolation, there is a risk that we will not be able to understand the complexity of program results and adequately interpret the impacts that may be identified. In order to have a solid understanding of the dynamics of an intervention and to be able to explain why things may be working, it is important that impact evaluation techniques are embedded in a framework of strong monitoring and process evaluation. Overall, we believe that using mixed methods—that is, explicitly adopting both quantitative and qualitative methods in the impact evaluation design—can significantly improve the learning in and about our programs.

As [Bamberger, Rao, and Woolcock \(2010, pp. 6–7\)](#) and [Leeuw and Vaessen \(2009\)](#) point out, there are several ways in which mixed methods can strengthen quantitative impact evaluation:

- Quantitative impact evaluations usually do not collect information on the quality of program implementation. Understanding the implementation process is crucial to understanding how program implementation affected program results and to correctly interpreting findings to differentiate whether disappointing results are due to weaknesses in program design or in implementation. Solid monitoring is therefore a prerequisite for effective evaluation and can be complemented with additional process analysis tools such as key informant interviews, direct participant observation, and focus groups.
- Incorporating qualitative methods can aid understanding of how and why the effect of the intervention may have varied across the target populations. Even though quantitative techniques can be designed to capture impact heterogeneity across groups, they cannot provide a clear understanding why these heterogeneities may have occurred.
- Although quantitative designs alone may be unable to capture the range of local circumstances in which each program is implemented, mixed methods can help provide detailed contextual analysis and document differences in the quality or speed of implementation across program sites. This qualitative information, in turn, can explain the potential differences in the outcomes of programs in different geographic areas.
- Many outcomes of youth livelihood interventions (such as mental health, empowerment, or household relations) are complex and multidimensional and may not be captured with quantitative methods. Mixed methods allow for tracking qualitative indicators and provide selected case-study analysis to help better understand the dynamics and results of the intervention. For example, small structured and semi-structured qualitative interviews in which participants are free to express real-life stories that fall outside categories of quantifiable information can help round out an understanding of a program's impact. Qualitative methods may also be better suited for collecting information on sensitive topics, such as reproductive health or violence.
- Qualitative methods may help identify appropriate indicators in the first place. For example, a focus group may yield important information about beneficiary concerns and how they expect the intervention to affect their lives.

Practically speaking, incorporating qualitative elements into our impact evaluation can take many forms, including open-ended survey questions, selected in-depth interviews and case studies, focus group discussions, participatory tools like the “Most Significant Change” technique (see [Davies and Dart 2005](#)), participant observation, and the like. To learn more about participatory monitoring and evaluation, consult [Catley and colleagues \(2010\)](#), [Sabo Flores \(2008\)](#), [Powers and Tiffany \(2006\)](#), and [Gawler \(2005\)](#).

At the same time, qualitative data alone are not well suited to identify program impacts. Using mixed methods, therefore, allows us to combine the strengths and offset the weaknesses of both qualitative and quantitative evaluation tools, allowing for an

overall stronger evaluation design (box 8.1 provides an example). In fact, the combined use of several research methods increases the credibility and validity of our results.

It is important to note, however, that using mixed-method designs can involve additional costs, time, and logistical challenges. In addition, it is often the case that the professional divisions among disciplines and researchers can make “building a multi-disciplinary team time consuming and challenging” (Bamberger, Rao, Woolcock 2010, p. 17).

BOX 8.1 Example of mixed method evaluation

In an evaluation of Junior Achievement's (JA) *Our Nation* curriculum, evaluators combined a range of quantitative and qualitative research methods. *Our Nation* is one of several JA Worldwide globally distributed programs for elementary schools and consists of a series of lessons for students aged 9–11 that examine issues related to entrepreneurship, resources needed for business, and globalization. On the one hand, the evaluation relied on an experimental design with random assignment of students to treatment and comparison groups. Comparison students were from the same states and regions as the treatment students but their classes were randomly assigned to receive the program after the evaluation was completed. Moreover, the evaluators conducted several case studies, using teacher, volunteer, and JA staff interviews, student focus groups, and classroom observation.

The quantitative evaluation results demonstrated some positive impacts on students' content knowledge related to entrepreneurship and globalization but no effects on levels of school engagement and the acquisition of 21st century skills. In addition, the qualitative tools allowed for an in-depth understanding of the mechanisms at work. On the one hand, qualitative tools confirmed a good quality of implementation, with the majority of sessions being implemented according to JA guidance and high levels of satisfaction reported by students, teachers, and volunteers. However, they also indicated challenges to good program delivery, including, for example, insufficient time for volunteers to cover all the contents. Finally, the qualitative evaluation results suggested ways to improve the program, including extending time for sessions, reducing the difficulty of the vocabulary, and providing more teaching guidance for volunteers.

Source: RMC Research (2009).

Cost-Benefit and Cost-Effectiveness Analyses

In many cases, organizations may use different strategies to tackle the same problem. For example, to increase employability, we may want to improve career counseling or improve training. Even when a single strategy is pursued, we may take different approaches to implementation, such as using either public or private training providers. If these various programmatic or implementation strategies were shown to have the same impact, for example, if they each were shown to improve the probability of employment three months after the intervention by 50 percent, would we be equally happy implementing one approach over the other? Probably not. It is not enough to know that an intervention works, for whom, and in what context; we also need to know at what cost.

Having a realistic estimate of the costs, in turn, allows us to answer the following questions:

- How can we choose among alternatives? Which program is the most cost-efficient given a certain level of impact?

- Would we be able to scale up? If costs are high, it is unlikely that we will be able to reach a large number of beneficiaries.
- Is any intervention always better than none? If the total costs outweigh the total benefits of the program, maybe the resources are better spent somewhere else.

Analytical Tools

The two tools commonly used to answer the above questions are cost-effectiveness analysis (CEA) and cost-benefit analysis (CBA):

Cost-effectiveness analysis identifies the full cost of a program and relates these costs to specific measures of outputs or outcomes (\$500 per person trained, per job created, per HIV/AIDS infection prevented, and the like). CEA thus tells us how much output or outcome we get per dollar spent, thereby identifying the most efficient allocation of resources when we compare alternative programs against the same criterion (see table 8.2).

TABLE 8.2 Cost-effectiveness estimates for Jóvenes programs

Country	Program	Cost Per Participant (in 2005 US\$)
Argentina	Proyecto Joven	\$1159
Chile	Chile Joven	\$825–\$1051
Peru	PROJoven	\$697

Source: Betcherman et al. (2007).

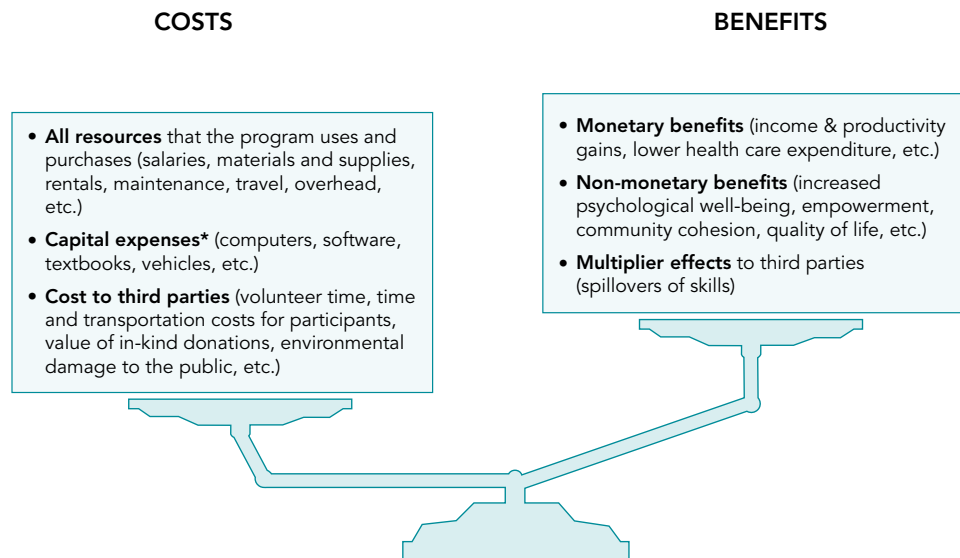
Cost-benefit analysis also identifies and quantifies the full cost of a program and further weighs those costs against the dollar value of all program benefits. Knowing the net benefits and net costs of the intervention, it is then possible to calculate the ratio of benefits to costs and to determine the return to society on the organization's investment. For example, the benefits:cost ratio is 2:1 if net benefits are \$1,000 per person and net costs are \$500. Overall, CBA seeks to determine whether benefits outweigh costs; that is, whether society is richer or poorer after making that investment.

Both CEA and CBA can be used before the intervention or during or after the program. However, only retrospective analysis will provide practitioners with the full information of actual costs and benefits to determine the overall success of the intervention. In fact, an impact evaluation is a necessary condition for having a reliable estimate of the program's direct and indirect benefits.

Capturing All Benefits and Costs

Cost-effectiveness and cost-benefit analyses require capturing, quantifying, and comparing all known costs (and, for CBA, known benefits) of the program to everyone directly or indirectly affected by the intervention: the implementing organization, the program beneficiaries, the government, and others (see figure 8.2).

FIGURE 8.2 Weighing costs and benefits



*Purchase of materials whose use exceeds one year.

.....
Social return on investment (SROI) is variation of CBA that compares extra-financial benefits relative to the resources invested. It assigns financial proxy values to all those outcomes identified by stakeholders that do not typically have market values. To learn more, please consult the SROI Network's Web site: <http://www.thesroinetwork.org/>

Trying to put a dollar value on many intangible benefits may be difficult and subjective, and it can represent a big challenge, especially for CBA. Hence, CBA is usually considered most useful when there are multiple types of benefits and consensus about how to quantify them in monetary terms (J-PAL 2011).

Calculating Net Benefits for Participants

The benefits of a program can be measured by its impact on individual participants. For example, say a skills training program is found to increase income by \$100 per person, per year, on average. In many areas in sub-Saharan Africa, this is a significant amount of money, and may represent great success for the program.

Assume the program costs \$1 million to implement, with \$400,000 to conduct the training and \$600,000 in overhead, including all staff salaries. If it reaches 1,000 people, the program thus costs \$1,000 per person to implement, with \$400 going toward training and \$600 going toward overhead. Is it worth running?

The answer to this question is based on three criteria.

First, **the program's impact must equal or exceed the impact of giving individuals cash equal to the cost of running the program.** In the example above, the impact must be compared with the effects of giving each person \$1,000 cash. There are two possible scenarios. First, a person actually uses the \$1,000 and purchases training with that money. The cost of the training is still only \$400 per person, which yields the same \$100 per person per year return. The individual then has an extra \$600 to use how they please, and is thus better off than with the program. In a second scenario, a person may use the money for something other than training that is less useful for her over the long term, such as cigarettes or other nonessential consumer goods. In the latter case, the program is worth running.

Second, **the program must have equal or greater return than running other programs.** Is it possible that another program could have realized the same or greater

impact per person for less money? This question requires a comparison of results across different program options. The program that has the greatest impact but costs the least is then the best program to continue running.

Finally, **the net present value of the return should be more than the cost of the program**. Present value is a way of thinking about the value of money today compared with its value in the future. Using the current example, we take the value of the money obtained yearly (the \$100 per person return on training) and adjust its value over a period of time according to the *discount rate*, which in most cases equals, the local interest rate. The net present value is simply the sum of the present value adjusted over a period of time. This is represented in table 8.3, using an interest rate of 20 percent, which is a common rate in many developing countries.

TABLE 8.3 Present value and net present value for a yearly return of \$100

Year	Present Value
0	\$100
1	\$83
2	\$69
3	\$58
4	\$48
5	\$40
6	\$33
7	\$28
8	\$23
9	\$19
10	\$16
Net present value	\$517

In this example, today \$100 is valued at \$100. However, at our current interest rate, the \$100 of income today will be worth only \$16 in ten years. Over a 10-year period, the net present value of our \$100 is \$517. Over the entire lifespan of a participant, the net present value will be at most \$600. Thus, a return of \$100 per person per year works out to a maximum return of \$600 per person over their lifetime.

According to these criteria, in today's dollars, our outlay for the program (\$1,000) is greater than the benefit to individuals, even though the cost for training (\$400) is less than the net present value of the training (\$517). Thus, unless people can be induced to take up the training on their own without the need for the overhead budget, or unless the overhead budget can be greatly reduced, the value of the training is not enough to justify the program.

Calculating Net Benefits for Society

In order to assess the net benefits to society we need to consider spillover effects. Spillovers refer to the positive or negative impacts the program has on those who are not directly involved with the program. There are two types of spillover effects that concern us here: multiplier effects, and prize and quantity effects.

Multiplier Effects

Multiplier effects occur when participants in a program impart their skills to others who were not formally associated with the program. For instance, a man trained in carpentry may train his son-in-law. The impact evaluation may measure only the impact the program had on the participant; it may miss the effect the program had on the son-in-law.

Using the example above, the impact on the carpenter is \$100 per year. The impact on the son-in-law may be smaller due to lower quality training, but clearly the training of one person has improved the livelihood prospects for two people. The cost per person is thus lower than the \$1,000 originally calculated.

Indirect benefits may be significant and could justify the costs of a program in some cases. In order to capture these spillover effects, plans should be made during the endline data collection to ask about others who have received training or otherwise benefited from the program.

Prize and Quantity Effects

Even though interventions may target only certain aspects of the population and local market, they can have effects on the larger economy, often referred to as “general equilibrium” effects. For example, if, as a result of our carpentry training, there are additional skilled carpenters in the local economy, competition among them may decrease prices for consumers. It is also possible that a program has negative spillovers on the economy. For instance, introducing extra tailors in an area where there are already a lot of tailors may drive prices so low that some of the tailors go out of business. This effect could significantly dilute the impact of a program.

Another undesired effect may result from negative consequences for nonparticipants. If participants of a particular intervention obtain a competitive edge in the labor market, for example, this may result in other youth not finding a job even though they would have in the absence of the program. Such effects are commonly referred to as displacement effects.

For most programs, prize and quantity effects are likely to be very small and not worth collecting data on. Large programs may wish to explore ways to capture this with an evaluation expert. One possibility may be to randomize the intervention at the community or district level.

Key Points

1. Depending on the learning objectives of a program and organization, it is worth exploring whether an impact evaluation can be designed to measure more than just the average impact of the program. Such additional impact questions can relate to heterogeneity, time-horizon, spillover effects, or the relative effectiveness of different program design options.
2. It is highly advisable to incorporate qualitative research elements into an impact evaluation. Using mixed-methods gives us a more comprehensive and nuanced

understanding of a program's impact, or lack thereof.

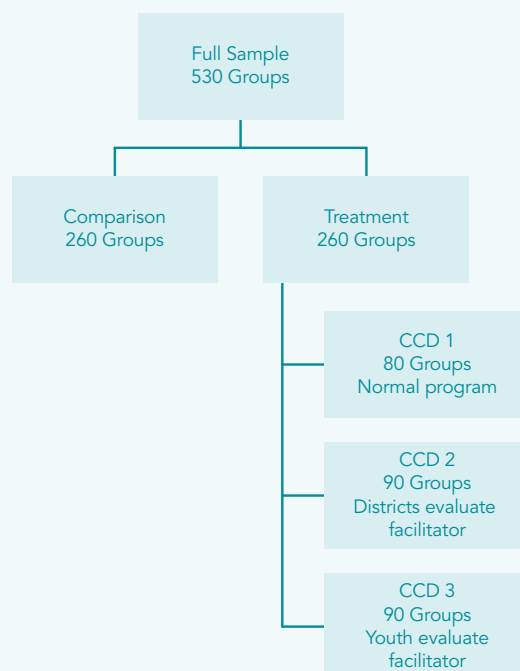
3. Information about the impact of a program may be of limited usefulness unless we also know the costs of designing and implementing the intervention. Any scale up will depend on this piece of information. It is therefore desirable to complement an impact evaluation with a cost-effectiveness or cost-benefit analysis.

NUSAF Case Study: Increasing the Relevance of the IE

Crosscutting Design

In addition to evaluating the overall effects of the Youth Opportunities Program, the impact evaluation was leveraged to test a complementary pilot intervention on an innovative program design variant. Anecdotal evidence from previous rounds of program funding suggested that the quality management, planning, and extension services provided by the district and the community facilitators are key determinants of individual youth group success. The impact evaluation therefore wanted to assess the effectiveness of giving an additional payment to hire a monitoring and extension advisor (MEA) that would be chosen by the group of youths themselves.

The treatment groups were randomly assigned to participate in the crosscutting design. Funded projects were randomly assigned to one of three groups. Treatment group 1 is treated as normal, with no additional intervention. Treatment group 2 has the district officers evaluate the MEAs, and treatment group 3 was given additional resources and asked to evaluate the MEAs themselves.



(continued)

NUSAF Case Study: Increasing the Relevance of the IE (cont'd)

Mixed Methods

The Youth Opportunities Program evaluation took advantage of quantitative and qualitative questions. The quantitative questionnaire was administered to approximately 2,600 youth, while the qualitative questionnaire was administered to about 100 youth. The qualitative questions included the following categories of interest:

1. Quality of group dynamics and cooperation, including process of group formation; group leadership and structures; group decision-making processes; past, present, and future of group activities; benefit and challenges of working in groups; and individual reasons for choosing to work in groups despite challenges.
2. NUSAF funds allocation, including group processes of fund allocation, group funding priorities versus project original plans, and deviation and other unofficial uses of fund.
3. Training experience, including process of choosing group skills training, confidence to apply skills learned, and benefits and challenges of applying skills as a livelihood strategy.
4. Livelihood strategies, including building livelihood after vocational training; risks, success, and failure associated with new livelihood strategies; reasons for success or failure; alternative livelihood strategies, and other strategies to deal with risks and shocks.
5. Empowerment and community participation, including sense of belonging in the communities, civic participation, gender relations, social support, social barriers, and relations with neighbors.

Quantitative data cannot alone bring out the full richness of a program. The responses to these questions will be used to better understand how NUSAF changed the lives of participants, as well as to provide some stories to help understand how the program impacted lives.

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

Key Reading

Bamberger, M., Rao, V., and Woolcock, M. 2010. *Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development*. Policy Research Working Paper 5245, Washington, DC: The World Bank.

http://www-wds.worldbank.org/servlet/WDSCContentServer/WDSP/IB/2010/03/23/000158349_20100323100628/Rendered/PDF/WPS5245.pdf

Bamberger, M., Rugh, J., and Mabry, L. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks: Sage Publications. (See chapter 13.) <http://realworldevaluation.org/>

Cellini, S. R., and Kee, J. E. 2010. "Cost-Effectiveness and Cost-Benefit Analysis." In: Wholey, J., Hatry, H. P., and Newcomer, K. E., eds. *Handbook of Practical Program Evaluation*, 3rd ed. San Francisco: Jossey-Bass.
<http://home.gwu.edu/~scellini/CelliniKee21.pdf>

Knowles, J., and Behrman, J. 2005. *A Practical Guide to Economic Analysis of Youth Projects*. HNP Discussion Paper. Washington, DC: The World Bank.
<http://siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/Resources/281627-1095698140167/KnowlesPracticalGuideFinal.pdf>

Notes
