Program Evaluation

Authored by

mohammad looti

November 19, 2022

Project evaluation is a systematic method for collecting, analyzing, and using information to answer questions about projects, policies and programs, particularly about their effectiveness and efficiency. In both the public and private sectors, stakeholders will want to know if the programs they are funding, implementing, voting for, receiving or objecting to are actually having the intended effect, and answering this question is the job of an evaluator.

The process of evaluation is considered to be a relatively recent phenomenon. However, planned social evaluation has been documented as dating as far back as 2200BC (Shadish, Cook & Lentish, 1991). Evaluation became particularly relevant in the U.S. in the 1960s during the period of the Great Society social programs associated with the Kennedy and Johnson administrations. Extraordinary sums were invested in social programs, but the impacts of these investments were largely unknown.

Program evaluations can involve both quantitative and qualitative methods of social research. People who do program evaluation come from many different backgrounds, such as sociology, psychology, economics, and social work. Some graduate schools also have specific training programs for program evaluation.

Doing an evaluation

Program evaluation may be conducted at several stages during a program's lifetime. Each of these stages raise different questions to be answered by the evaluator, and correspondingly different evaluation approaches are needed. Rossi, Lipsey and Freeman (2004) suggest the following kinds of assessment, which may be appropriate at different stages:

Assessment of the program's cost and efficiency
Assessment of the program's outcome or impact (i.e., what it has actually achieved)
Assessment of how the program is being implemented (i.e., is it being implemented according to plan?)
Assessment of program design and logic/theory
Assessment of the need for the program

Assessing needs

A needs assessment examines the population that the program intends to target, to see whether the need as conceptualised in the program actually exists in the population; whether it is, in fact, a problem; and if so, how it might best be dealt with. This includes identifying and diagnosing the actual problem the program is trying to address, who or what is affected by the problem, how widespread the problem is, and what are the measurable effects that are caused by the problem. For example, for a housing program aimed at mitigating homelessness, a program evaluator may want to find out how many people are homeless in a given geographic area and what their

demographics are. Rossi, Lipsey and Freeman (2004) caution against doing an intervention without properly assessing the need for one, because this might result in a great deal of wasted funds if the need did not exist or was misconceived.

Assessing program theory

The program theory, also called a logic model or impact pathway, is an assumption, implicit in the way the program is designed, about how the program's actions are supposed to achieve the outcomes it intends. This 'logic model' is often not stated expicitly by people who run programs, it is simply assumed, and so an evaluator will need to draw out from the program staff how exactly the program is supposed to achieve its aims and assess whether this logic is plausible. For example, in an HIV prevention program, it may be assumed that educating people about HIV/AIDS transmission, risk and safe sex practices will result in safer sex being practiced. However, research in South Africa increasingly shows that in spite of increased education and knowledge, people still often do not practice safe sex. Therefore, the logic of a program which relies on education as a means to get people to use condoms may be faulty. This is why it is important to read research that has been done in the area. Explicating this logic can also reveal unintended or unforeseen consequences of a program, both positive and negative. The program theory drives the hypotheses to test for impact evaluation. Developing a logic model can also build common understanding amongst program staff and stakeholders about what the program is actually supposed to do and how it is supposed to do it, which is often lacking(see Participatory Impact Pathways Analysis).

Assessing implementation

Process analysis looks beyond the theory of what the program is supposed to do and instead evaluates how the program is being implemented. This evaluation determines whether the components identified as critical to the success of the program are being implemented. The evaluation determines whether target populations are being reached, people are receiving the intended services, staff are adequately qualified, etc. Process evaluation is an ongoing process in which repeated measures may be used to evaluate whether the program is being implemented effectively.

Assessing the impact (effectiveness)

The impact evaluation determines the causal effects of the program. This involves trying to measure if the program has achieved its intended outcomes. This can involve using sophisticated statistical techniques in order to measure the effect of the program and to find causal relationship between the program and the various outcomes. More information about impact evaluation is

found under the heading 'Determining Causation'.

Assessing efficiency

Finally, cost-benefit or cost-effectiveness analysis assesses the efficiency of a program. Evaluators outline the benefits and cost of the program for comparison. An efficient program has a lower cost-benefit ratio.

Determining causation

Perhaps the most difficult part of evaluation is determining whether the program itself is causing the changes that are observed in the population it was aimed at. Events or processes outside of the program may be the real cause of the observed outcome (or the real prevention of the anticipated outcome).

Causation is difficult to determine. One main reason for this is self selection bias. People select themselves to participate in a program. For example, in a job training program, some people decide to participate and others do not. Those who do participate may differ from those who do not in important ways. They may be more determined to find a job or have better support resources. These characteristics may actually be causing the observed outcome of increased employment, not the job training program.

If programs could use random assignment, then they could find a strong correlation or association. Causation is not something that can be proved through correlation. A program could randomly assign people to participate or to not participate in the program, eliminating self-selection bias. Thus, the group of people who participate would be the same as the group who did not participate.

However, since most programs cannot use random assignment, causation cannot be determined. Impact analysis can still provide useful information. For example, the outcomes of the program can be described. Thus the evaluation can describe that people who participated in the program were more likely to experience a given outcome than people who did not participate.

If the program is fairly large, and there are enough data, statistical analysis can be used to make a reasonable case for the program by showing, for example, that other causes are unlikely.

Reliability, Validity and Sensitivity in Program Evaluation

It is important to ensure that the instruments (for example, tests, questionnaires, etc) used in program evaluation are as reliable, valid and sensitive as possible. According to Rossi et al. (2004, p. 222), 'a measure that is poorly chosen or poorly conceived can completely undermine the worth

of an impact assessment by producing misleading estimates. Only if outcome measures are valid, reliable and appropriately sensitive can impact assessments be regarded as credible'.

Reliability

The reliability of a measurement instrument is the 'extent to which the measure produces the same results when used repeatedly to measure the same thing' (Rossi et al., 2004, p. 218). The more reliable a measure is, the greater its statistical power and the more credible its findings. If a measuring instrument is unreliable, it may dilute and obscure the real effects of a program, and the program will 'appear to be less effective than it actually is' (Rossi et al., 2004, p. 219). Hence, it is important to ensure the evaluation is as reliable as possible.

Validity

The validity of a measurement instrument is 'the extent to which it measures what it is intended to measure' (Rossi et al., 2004, p. 219). This concept can be difficult to accurately measure: in general use in evaluations, an instrument may be deemed valid if accepted as valid by the stakeholders (stakeholders may include, for example, funders, program administrators, et cetera).

Sensitivity

The principal purpose of the evaluation process is to measure whether the program has an effect on the social problem it seeks to redress; hence, the measurement instrument must be sensitive enough to discern these potential changes (Rossi et al., 2004). A measurement instrument may be insensitive if it contains items measuring outcomes which the program couldn't possibly effect, or if the instrument was originally developed for applications to individuals (for example standardised psychological measures) rather than to a group setting (Rossi et al., 2004). These factors may result in 'noise' which may obscure any effect the program may have had.

To conclude, only measures which adequately achieve the benchmarks of reliability, validity and sensitivity can be said to be credible evaluations. It is the duty of evaluators to produce credible evaluations, as their findings may have far reaching effects. A discreditable evaluation which is unable to show that a program is achieving its purpose when it is in fact creating positive change, may cause the program to lose its funding undeservedly.

The Shoestring Approach

The "Shoestring evaluation approach" is designed to assist evaluators operating under limited budget , limited access or availability of data and limited turnaround time, to conduct effective

evaluations that are methodologically rigorous(Bamberger, Rugh, Church & Fort, 2004). This approach has responded to the continued greater need for evaluation processes that are more rapid and economical under difficult circumstances of budget, time constraints and limited availability of data. However, it is not always possible to design an evaluation to achieve the highest standards available. Many programs do not build an evaluation procedure into their design or budget. Hence, many evaluation processes do not begin until the program is already underway, which can result in time, budget or data constraints for the evaluators, which in turn can affect the reliability, validity or sensitivity of the evaluation. > The shoestring approach helps to ensure that the maximum possible methodological rigour is achieved under these constraints.

Budget Constraints

Frequently, programs are faced with budget constraints because most original projects do not include a budget to conduct an evaluation (Bamberger et al., 2004). Therefore, this automatically results in evaluations being allocated smaller budgets that are inadequate for a rigorous evaluation. Due to the budget constraints it might be difficult to effectively apply the most appropriate methodological instruments. These constraints may consequently affect the time available in which to do the evaluation (Bamberger et al., 2004). Budget constraints may be addressed by simplifying the evaluation design, revising the sample size, exploring economical data collection methods (such as using volunteers to collect data, shortening surveys, or using focus groups and key informants) or looking for reliable secondary data (Bamberger et al., 2004).

Time Constraints

The most time constraint that can be faced by an evaluator is when the evaluator is summoned to conduct an evaluation when a project is already underway if they are given limited time to do the evaluation compared to the life of the study, or if they are not given enough time for adequate planning. Time constraints are particularly problematic when the evaluator is not familiar with the area or country in which the program is situated (Bamberger et al., 2004). Time constraints can be addressed by the methods listed under budget constraints as above, and also by careful planning to ensure effective data collection and analysis within the limited time space.

Data Constraints

If the evaluation is initiated late in the program, there may be no baseline data on the conditions of the target group before the intervention began (Bamberger et al., 2004). Another possible cause of data constraints is if the data have been collected by program staff and contain systematic reporting biases or poor record keeping standards and is subsequently of little use (Bamberger et al., 2004). Another source of data constraints may result if the target group are difficult to reach to

collect data from - for example homeless people, drug addicts, migrant workers, et cetera (Bamberger et al., 2004). Data constraints can be addressed by reconstructing baseline data from secondary data or through the use of multiple methods. Multiple methods, such as the combination of qualitative and quantitative data can increase validity through triangulation and save time and money. Additionally, these constraints may be dealt with through careful planning and consultation with program stakeholders. By clearly identifying and understanding client needs ahead of the evaluation, costs and time of the evaluative process can be streamlined and reduced, while still maintaining credibility.

All in all, time, monetary and data constraints can have negative implications on the validity, reliability and transferability of the evaluation. The shoestring approach has been created to assist evaluators to correct the limitations identified above by identifying ways to reduce costs and time, reconstruct baseline data and to ensure maximum quality under existing constraints (Bamberger et al., 2004).

Methodological challenges presented by language and culture

The purpose of this section is to draw attention to some of the methodological challenges and dilemmas evaluators are potentially faced with when conducting a program evaluation in a developing country. In many developing countries the major sponsors of evaluation are donor agencies from the developed world, and these agencies require regular evaluation reports in order to maintain accountability and control of resources, as well as generate evidence for the program's success or failure (Bamberger, 2000). However, there are many hurdles and challenges which evaluators face when attempting to implement an evaluation program which attempts to make use of techniques and systems which are not developed within the context to which they are applied (Smith, 1990). Some of the issues include differences in culture, attitudes, language and political process (Ebbutt, 1998, Smith, 1990).

Culture is defined by Ebbutt (1998, p. 416) as a "constellation of both written and unwritten expectations, values, norms, rules, laws, artifacts, rituals and behaviours that permeate a society and influence how people behave socially". Culture can influence many facets of the evaluation process, including data collection, evaluation program implementation and the analysis and understanding of the results of the evaluation (Ebbutt, 1998). In particular, instruments which are traditionally used to collect data such as questionnaires and semi-structured interviews need to be sensitive to differences in culture, if they were originally developed in a different cultural context (Bulmer & Warwick, 1993). The understanding and meaning of constructs which the evaluator is attempting to measure may not be shared between the evaluator and the sample population and thus the transference of concepts is an important notion, as this will influence the quality of the data collection carried out by evaluators as well as the analysis and results generated by the data (ibid).

Language also plays an important part in the evaluation process, as language is tied closely to culture (ibid). Language can be a major barrier to communicating concepts which the evaluator is trying to access, and translation is often required (Ebbutt, 1998). There are a multitude of problems with translation, including the loss of meaning as well as the exaggeration or enhancement of meaning by translators (ibid). For example, terms which are contextually specific may not translate into another language with the same weight or meaning. In particular, data collection instruments need to take meaning into account as the subject matter may not be considered sensitive in a particular context might prove to be sensitive in the context in which the evaluation is taking place (Bulmer & Warwick, 1993). Thus, evaluators need to take into account two important concepts when administering data collection tools: lexical equivalence and conceptual equivalence (ibid). Lexical equivalence asks the question: how does one phrase a question in two languages using the same words? This is a difficult task to accomplish, and uses of techniques such as back-translation may aid the evaluator but may not result in perfect transference of meaning (ibid). This leads to the next point, conceptual equivalence. It is not a common occurrence for concepts to transfer unambiguously from one culture to another (ibid). Data collection instruments which have not undergone adequate testing and piloting may therefore render results which are not useful as the concepts which are measured by the instrument may have taken on a different meaning and thus rendered the instrument unreliable and invalid (ibid).

Thus, it can be seen that evaluators need to take into account the methodological challenges created by differences in culture and language when attempting to conduct a program evaluation in a developing country.

Utilization of Evaluation Results

There are three conventional uses of evaluation results: persuasive utilization, direct (instrumental) utilization, and conceptual utilization. Persuasive utilization is the enlistment of evaluation results in an effort to persuade an audience to either support an agenda or to oppose it. Unless the 'persuader' is the same person that ran the evaluation, this form of utilization is not of much interest to evaluators as they often cannot foresee possible future efforts of persuasion.

Direct (instrumental) Utilization

Evaluators often tailor their evaluations to produce results that can have a direct influence in the improvement of the structure, or on the process, of a program. For example, the evaluation of a novel educational intervention may produce results that indicate no improvement in students' marks. This may be due to the intervention not having a sound theoretical background, or it may be that the intervention is not run according to the way it was created to run. The results of the evaluation would hopefully lead to the creators of the intervention going back to the drawing board

and re-creating the core structure of the intervention, or even changing the implementation processes.

Conceptual Utilization

But even if evaluation results do not have a direct influence in the re-shaping of a program, they may still be used to conscientize people with regards to the issues that form part of the concerns of the program. Going back to the example of an evaluation of a novel educational intervention, the results can also be used to inform educators and students about the different barriers that may influence students' learning difficulties. A number of studies on these barriers may then be initiated by this new information.

Variables Affecting Utilization

There are five conditions that seem to affect the utility of evaluation results, namely relevance, communication between the evaluators and the users of the results, information processing by the users, the plausibility of the results, as well as the level of involvement or advocacy of the users.

Guidelines for Maximizing Utilization

Quoted directly from Rossi et al. (2004, p. 416).:

Evaluators must understand the cognitive styles of decisionmakers
Evaluation results must be timely and available when needed
Evaluations must respect stakeholders' program commitments
Utilization and dissemination plans should be part of the evaluation design
Evaluations should include an assessment of utilization

Internal Versus External program evaluators

The choice of the evaluator chosen to evaluate the program may be regarded as equally important as the process of the evaluation. Evaluators may be internal (persons associated with the program to be executed) or external (Persons not associated with any part of the execution/implementation of the program). (Division for oversight services,2004). The following provides a brief summary of the advantages and disadvantages of internal and external evaluators adapted from the Division of oversight services (2004), for a more comprehensive list of advantages and disadvantages of internal and external evaluators, see (Division of oversight services, 2004).

Internal evaluators

Advantages

May have better overall knowledge of the program and possess informal knowledge of the program
Less threatening as already familiar with staff
Less costly

Disadvantages

May be less objective
May be more preocuppied with other activities of the program and not give the evaluation complete attention
May not be adequately trained as an evaluator.

External evaluators

Advantages

More objective of the process, offers new perspectives, different angles to observe and critique the process
May be able to dedicate greater amount of time and attention to the evaluation
May have greater expertise and evaluation brain

Disadvantages

May be more costly and require more time for the contract, monitoring, negotiations etc.
May be unfamiliar with program staff and create anxiety about being evaluated
May be unfamiliar with organization policies, certain constraints affecting the program.

Paradigms in program evaluation

Potter (2006) identifies and describes three broad paradigms within program evaluation . The first, and probably most common, is the positivist approach, in which evaluation can only occur where there are "objective", observable and measurable aspects of a program, requiring predominantly quantitative evidence. The positivist approach includes evaluation dimensions such as needs assessment, assessment of program theory, assessment of program process, impact assessment and efficiency assessment (Rossi, Lipsey and Freeman, 2004).

The second paradigm identified by Potter (2006) is that of interpretive approaches, where it is argued that it is essential that the evaluator develops an understanding of the perspective, experiences and expectations of all stakeholders. This would lead to a better understanding of the various meanings and needs held by stakeholders, which is crucial before one is able to make judgments about the merit or value of a program. The evaluator's contact with the program is often over an extended period of time and, although there is no standardized method, observation,

interviews and focus groups are commonly used.

Potter (2006) also identifies critical-emancipatory approaches to program evaluation, which are largely based on action research for the purposes of social transformation. This type of approach is much more ideological and often includes a greater degree of social activism on the part of the evaluator. Because of its critical focus on societal power structures and its emphasis on participation and empowerment, Potter argues this type of evaluation can be particularly useful in developing countries.

Despite the paradigm which is used in any program evaluation, whether it be positivist, interpretive or critical-emancipatory, it is essential to acknowledge that evaluation takes place in specific socio-political contexts. Evaluation does not exist in a vacuum and all evaluations, whether they are aware of it or not, are influenced by socio-political factors. It is important to recognize the evaluations and the findings which result from this kind of evaluation process can be used in favour or against particular ideological, social and political agendas (Weiss, 1999). This is especially true in an age when resources are limited and there is competition between organizations for certain projects to be prioritised over others (Louw, 1999).

CDC framework

In 1999, the Centers for Disease Control and Prevention (CDC) published a six-step framework for conducting evaluation of public health programs. The publication of the framework is a result of the increased emphasis on program evaluation of government programs in the US. The six steps are:

Engage stakeholders
Describe the program.
Focus the evaluation.
Gather credible evidence.
Justify conclusions.
Ensure use and share lessons learned.