Program Evaluation Methodology

The Degrees of Evidence Framework

What is the PSD-IAI?

The PSD Impact Assessment Initiative is funded by USAID under the Accelerated Microenterprise Advancement Project. PSD-IAI conducts impact assessments to create learning about and improve the effectiveness of new generation economic growth programs, identifying specific impacts that PSD interventions have on pro-poor growth.

The PSD-IAI is managed by DAI and led by a team of internationally recognized evaluation experts. Among other activities, PSD-IAI has designed and implemented longitudinal impact assessments of USAID PSD programs in Kenya, Brazil, Zambia, and India. For information on work being done by PSD-IAI and access to all PSD-IAI publications, go to www.microlinks.org/psdimpact or contact Jeanne Downing at jdowning@usaid.gov.

The lack of credible information on program effectiveness severely limits the ability of USAID and other development organizations to practice evidence-based development programming. Owing to a variety of factors, USAID finds it difficult to craft an evaluation strategy that meets its diverse information needs with appropriate levels of transparency and credibility. One important factor contributing to this outcome is the confusion and controversy surrounding evaluation methodologies, their purposes, relative credibility, and appropriateness to answer various questions.

The Degrees of Evidence framework for monitoring and evaluation developed by the PSD-IAI offers a comprehensive and practical framework that clears up the confusion surrounding evaluation methodologies and that allows program planners to assess their methodological options according to their purpose, rigor, and cost and thereby match M&E methodologies to the questions asked, the level of credibility sought, and the level of resources available.

DEGREES OF EVIDENCE FRAMEWORK

The Degrees of Evidence framework provides an approach for understanding and gauging methodological rigor, which is in turn helpful both for understanding the differences among different evaluation methodologies and for matching evaluation methodologies to the questions asked, the level of credibility required, and the available budget. The Degrees of Evidence framework defines rigor using multiple criteria and in the process demonstrates the uses and value of alternative evaluation methodologies.

The basis for all systematic M&E is a causal model that lists program activities and shows the specific outputs, outcomes, and impacts to which these activities are expected to lead.

Within the Degrees of Evidence framework, methodological rigor is determined by the extent to which the research design adheres to the following methodological principles.

1. Methodological validity, which is in turn measured along four dimensions

    a. *Internal Validity*: The extent to which the findings of an evaluation accurately represent the causal relationship between an intervention and an outcome or impact in the particular circumstances of that evaluation. Internal validity requires that observed changes can be attributed to the program and not to other possible causes, which must be ruled out. In other words, the evaluation must establish a valid counterfactual – a picture of what would have happened in the absence of the intervention. Internal validity is a primary criterion for evaluating development programs, but other criteria are also important.

b. *External Validity*: The extent to which the findings obtained from an investigation conducted under particular circumstances can be generalized to other persons, places, and times. If different outcomes or impacts occur in circumstances that differ from those of the particular investigation, the findings lack external validity.

c. *Construct Validity*: The degree to which legitimate inferences can be made from the evaluation study to the underlying theoretical concepts (variables) included in the causal model. Were the treatment, outcome, and other (i.e., mediating) variables appropriately defined and measured in the empirical work?

d. *Statistical Conclusion Validity*: Whether researchers have correctly applied statistical methods and identified the statistical strength/certainty of their results.

2. Triangulation: The evidence of program effectiveness is stronger to the extent that it is supported by multiple sources of evidence. In particular, mixed method evaluation designs using different combinations of quantitative and qualitative research methodologies allow researchers to triangulate toward more credible evaluation findings.

3. Methodological Transparency: The research methodologies are well documented and their weaknesses and related implications are identified.

4. Sound Data Collection Methods: Data collection methods follow accepted good practice, including the use of competent researchers and the implementation of sound quality control measures.

5. Methodological Appropriateness: The research methodology is appropriate to answer the research question(s). This principle incorporates the fundamental concept that the selection of the research methodology is driven by the research question. Program evaluation is not a pre-determined research methodology in search of applications but the matching of research methodologies to the questions asked, as well as to the political, resource, and field constraints faced by researchers. Starting with the question rather than the methodology and taking into account relevant constraints will often point researchers toward methodologies outside their typical realm of preference or experience.

On each of the criteria for rigor, there exists a continuum from 'sound' to 'unsound.' Overall rigor is determined by how well the evaluation scores across all of the criteria. It is quite possible for an evaluation to score high on certain criteria but low on others. For example, a research design may score high on internal validity but low on construct or external validity. Such a study may be technically impressive but offer little useful guidance to development programmers and policy makers.

IMPLICATIONS OF THE DEGREES OF EVIDENCE FRAMEWORK

To capture the lessons to be learned from hundreds of diverse programs, USAID will need an evaluation strategy that will necessarily utilize a diverse set of evaluation methodologies to answer a diverse set of questions related to the effectiveness of its economic and social programs. Creating such a strategy, however, is not easy. It requires both knowledge of the various methodological options and an understanding of their rigor, use, and cost. The Degrees of Evidence framework does this. It provides a practical framework that allows planners to match evaluation methodologies to the questions asked and its broader learning agenda within established budget constraints. It allows USAID, moreover, to assess the quality of the evidence produced as a result in a more accurate and meaningful way, which in turn allows it to gauge more effectively the relevance of the evidence for program planning and design. In this manner, the Degrees of Evidence framework is a valuable resource in any institutionalized effort to create a unified and systematic evaluation strategy and feedback loop.