MATHEMATICA Policy Research

Is it Baked Yet?

Using Evaluability Assessment to determine readiness for Impact Evaluation

With the rise of systematic reviews designed to inform evidence-based decision-making, policy makers and program managers must identify or design robust interventions that demonstrate effectiveness and promise for scale-up. However, if the intervention is not fully developed or ready to for implementation with fidelity, then impact evaluations have limited usefulness for rigorous evaluation. Conducting an evaluability assessment is one way to determine whether an impact evaluation is warranted. This brief provides an overview of evaluability assessment (EA) and its application as a tool for helping prepare and assessing whether an intervention is ready for rigorous impact evaluation.

EA is not a new approach. In fact, implementers and researchers began conducting evaluability assessments in the 1970s. While evaluability assesessments led to some important accomplishments, they fell out of favor by the late 1980s. EAs were viewed by some as too ambiguous to be functional and by others as a way to indefinitely postpone rigorous evaluations. We explore in this brief EA's uses and limitations for program development and learning. We then present several examples of EA's application in recent evaluations, including our experience with a pre-primary education intervention in Tanzania. Next we provide recommendations for using EA in the pre-evaluation process.

WHAT IS EVALUABLITY ASSESSMENT?

Evaluability assessment is a tool for pre-evaluation learning. Rigorous impact evaluations are costly, not only because of the financial and human resources involved, but also due to the opportunity cost of evaluating one potentially impactful intervention over another. Given the costs imposed by impact evaluation, innovation testing and learning can help to examine the feasibility and advisability of evaluations prior to program roll out or scale-up. As impact evaluations are evolving into standard practice, it is increasingly essential to embed critical learning activities into the implementation and evaluation process.

Evaluability assessment, coined by Joseph Wholey and the Urban Institute in the late 1970s, is one such type of learning. EA "assesses the extent to which programs are ready for future evaluation and helps key stakeholders come to agreement on realistic program goals, evaluation criteria, and intended uses of evaluation information"(Wholey, 2010). (See Box 1 for further details.[1]) Considered a process that engages stakeholders (not something 'done' to an agency or program), EA incorporates

[1] We summarize guidance from Wholey (2010).

elements of program theory, stakeholder evaluation, process evaluation, needs assessment, and other techniques (described in Box 2). Dedicating resources to determine the suitability of a full impact evaluation serves as a way to better develop programs and to avoid scale-up and evaluation when it is unlikely to be successful or meaningful. Yet EA can be implemented in a cost effective way by embedding monitoring and learning activities within program implementation.

EA is best conducted by an external team with access to key documentation and personnel because independent assessors can bring objectivity in a way that is difficult for program implementers to manage internally. In practice, the evaluability assessor would participate in intervention planning and implementation activities, conduct a comprehensive review of program documentation, and review management and monitoring data collected by various implementers and M&E partners.

Box 1: Components of an evaluability assessment

EA components:

- Involve intended users and other key stakeholders
- Clarify the program design
- Explore program reality
- Assess the plausibility of the program
- Reach agreement on any needed changes in program design or in program implementation
- Reach agreement on the focus and intended use of any further evaluation

Though these components might, at first glance, read as steps, evaluators widely recognize that the EA process is one in which several components may be completed simultaneously or repeated multiple times.

EVALUABILITY ASSESSMENT: THEN AND NOW

EA is on the upswing.

Throughout the late 1970s to early 1980s, the United States Department of Health and Human Services and the Department of Education used evaluability assessments fairly consistently to determine program readiness for evaluation. Out of 57 health and education programs from 1972 to 1984, over half were considered ready for a formal evaluation, with the size and seniority of the program influencing whether it was selected for EA (Rog, 1985).

Though its popularity in the federal government later waned after Wholey's departure, an overview of EA from 1986 to 2006 shows that, since 1995, EA's have been conducted more frequently. In addition to direct applications, evaluators have put forward a number of methodologies to outline and encourage the practice of EA. Renewed federal efforts to connect funding to performance have led to a variety of fields

and programs making use of EA, from social services and healthcare to community development, education, energy, and technology (Trevisan, 2007). For example, the National Institute of Justice (NIJ) began to incorporate EA as a pre-evaluation activity. Over the period 2002-2003, the NIJ examined 461 anti-crime programs and commissioned evaluability assessments for 57 of these projects. Of these 57, 16 programs received or were under review for impact evaluations (Chemers and Reed, 2005). Including EA in the pre-evaluation learning process in this way ensures that evaluation resources are directed to programs likely to produce meaningful results.

Davies (2013) further documents that EA is experiencing an uptick in popularity among international development agencies and organizations. In particular, the Inter-American Development Bank (IADB) maintains a strong tradition of applying EA as a criterion in project design and approval. This not only allows the IADB to track the evaluability of lending operations, but provides valuable information on how problems can be corrected and the bank's overall operations improved by, for example, identifying weak oversight or improper risk management (Soares et al., 2010).

The Millennium Challenge Corporation also uses the EA prior to launching evaluations of their country investments across sectors such as education, energy, and water and sanitation. In these EAs, prior to launching into study design activities, the evaluator assesses whether the implementers have clearly defined the problem underlying the intervention, the theory of change, the risks and assumptions, risk mitigation strategies, project participants, geographic scope, and metrics for measuring results and accountability. Evaluators share results of this assessment with all stakeholders to gather feedback and consensus around the program theory and implementation. Only then can the evaluator move to the study design and implementation phase.

The above are examples of how EA, when applied correctly, is an extremely useful tool to filter the programs most appropriate for further evaluation as well as to identify problem areas and potential corrections.

Why is EA not commonplace?

Despite the widespread use of evaluability assessments, there is little operational consensus on how to implement an EA. The lack of precise criteria impedes a consistent application of EA between and within fields. For instance, Davies (2013) describes how eight international agencies apply EA: In principle, all of these organizations use EA to articulate the program design, availability of information, and institutional context. However, in practice, these same agencies differ substantially in how they implement EAs.

EAs typically encompass one or several evaluation activities, such as process evaluation, needs assessments, and stakeholder evaluation. As a result, most EAs do not conform precisely to any single methodology. Moreover, an EA may be conducted in principle, but not explicitly called an EA, while various sub-elements of EA may be used separately (Trevisan, 2007). Finally, there is no guideline to compare the usefulness of various EA models as well as limited knowledge on how to integrate EAs into broader evaluation plans.

Thus, despite decades of application and discussion, there remains no clear,

standardized guidance on how to conduct an EA and different approaches are used by different groups and vary considerably by discipline. While this ambiguity has certainly not impeded its application, it is may limit the appeal or reduce trust in the approach among funders, policy makers, and evaluators.

TOWARD A MODERN CONCEPTION OF EVALUABILITY

A typology of EA is emerging.

EA is largely about clarification of a program's goals, processes and intended outcomes among all parties rather than as a means to an end (impact evaluation). That is, EA "gives evaluators a credible niche for doing program and organizational development" (Patton, 2008). This iterative process, typically accompanied by external technical assistance, clarifies program design and implementation, may improve communication and motivation among stakeholders, and serves to expand internal evaluative capacity. Regardless of its precise components, EA serves as a conversation between stakeholders who can gain clarity, deeper insights and expertise from this back-and-forth. While a handful of practitioners have highlighted these significant benefits of EA, they have not held a central place in the discussion. Nevertheless, EA has proven increasingly useful in a variety of scenarios. We highlight several below.

Experts use EA to select the most promising programs from a portfolio of many.

A special issue of *New Directions for Evaluation* edited by Laura Leviton (2010) describes the use of EA in the Early Assessment of Programs and Policies to Prevent Childhood Obesity (the Early Assessment initiative) funded by the Robert Wood Johnson Foundation and directed by the Centers for Disease Control and Prevention. The issue makes a case for EA as the first evaluation method in a two-step process to identify obesity interventions with the most potential from a large portfolio of projects. With an outside team determining the degree of implementation, plausibility, and feasibility, programs receive the opportunity to participate in a structured learning phase.

Starting from a portfolio of 458 policy innovations, 128 met the initiative's inclusion criteria and 53 were selected for EA. A final 20 were judged to be both evaluable and highly promising. Ultimately, a low-cost method of learning prior to initiating plans for scale-up ensured a more judicious use of limited resources, directing them to the programs most likely to produce an impact. And, as expected, the programs were effective. Moreover, the overwhelmingly positive reaction from those involved in program development, specifically with respect to the learning that came out of the EA process, suggests that EA can play an important role not only in the evaluation decision but in learning.

Experts apply EA as a framework for planning and implementation in a specific intervention.

In 2013, the Hewlett Foundation contracted Mathematica Policy Research to conduct an impact evaluation of a new, highly collaborative effort to improve learning outcomes for children in Tanzania called "Fursa kwa Watoto" (Opportunities for Children). The goal of FkW was to enhance the quality of existing pre-primary education (PPE) and test innovative approaches to expanding quality PPE in line with current policy initiatives (see Box 3).

Box 3: "Fursa kwa Watoto" (Opportunities for Children)

Dubai Cares, the William and Flora Hewlett Foundation, Children in Crossfire, Center for Social Responsibility Group Africa, Aga Khan University (Dar es Salaam), UNICEF, and Mathematica Policy Research worked with partners in the Tanzania Ministry of Education and Vocational Training (MoEVT), Ministry of Community Development, Gender and Children (MCDGC), and the Prime Minister's Office for Regional and Local Government (PMO-RALG). Fursa kwa Watoto's primary objective is to improve the impact of PPE, as defined by the MoEVT, on children's learning and readiness for school with a quality, cost effective enhancement.

The learning collaborative is creating a multi-pronged, evidence-informed intervention designed to impact instructional and classroom environment quality and show a measureable improvement in student learning outcomes by the end of Standard II (grade 2 when students are 8 to 9 years old). A key guiding principle is that the intervention must fit within the government's existing structure, thereby facilitating scale up and sustainability in the future. The primary focus of the model is improved teaching practices through teacher training and feedback on classroom practices (through observation, modeling and coaching), ongoing support for teacher behavior change through mentoring mechanisms, improvements to the classroom environment, the use of locally made learning materials, and training of head teachers, local education officers, and parents in the importance of PPE.

During the early inception and pilot phase of the program, implementing partners and the program funder (Dubai Cares) realized it would take some time for the intervention to be implemented and ready for an impact evaluation. Mathematica's role evolved over time to include providing technical guidance and supporting partners to collectively assess the extent to which FkW is ready to be tested with an impact evaluation. Through a comprehensive evaluability assessment, Mathematica assessed key aspects of the program (such as program fidelity, scalability and sustainability) vis-à-vis evaluation criteria and made specific evaluation recommendations to partners. In doing so, we adapt the general EA framework to our context by drawing on guidance notes from international organizations, critiques of EA, and discussions with professional evaluators and EA experts.

We outline the questions underpinning our EA in Box 4. Prior to conducting the EA, we defined sub-questions relevant to FkW that should be answered in order to fulfill the broad guidance of Wholey (2010). We then identified the documentation necessary to answer these questions exhaustively. This mapping process identifies information gaps clearly for the steering committee. Once

Box 4: Key questions underpinning our assessment

- Do key stakeholders agree on the underlying theory of change and believe that partners can implement the program with the fidelity needed to improve pre-primary instruction and child readiness?
- Are information needs well-defined to monitor the implementation and assess the outcomes of the program?
- Is it feasible and realistic for partners to meet these information needs throughout pilots and potential scale-up?
- Do partners believe that the program is sustainable and scalable?

completed, Mathematica's internal project team as well as external quality assurance advisors used this information to arrive at independent conclusions on FkW's evaluability and provide guidance on how to move forward.

The EA process garnered positive reactions from members of the FkW steering committee. For example, the evaluability assessment has led to a learning and measurement culture among FkW implementing partners so that all partners collect monitoring, learning, and outcome data. As a collaborative, FkW partners reviewed and discussed findings, and this early data informed real-time program improvements. By developing and revising learning tools for each aspect of the intervention, FkW was able to distinguish which program components are most important from those that are not essential to the intervention's success. FkW partners were conditioned to consider each component of the ultimate pre-primary package in order to design a robust intervention that can be implemented with fidelity across Tanzania.

The EA process is valuable.

Our recent experiences and those of other practitioners suggest that EAs can bring about substantial benefits in the evaluation process. We highlight several instances of EA's application, from using it to determine the evaluability of a single, already-commissioned program to applying it as a selection tool to identify the most promising interventions from a set of many. By clarifying a program's design, studying its implementation, and identifying obstacles that impede its feasibility and success, an EA can present evaluators with a clear picture of program readiness.

As most learning typically occurs prior to an evaluation, using EA as a framework to provide structure to this developmental stage can help to maximize the learning incurred. Viewing EA more as a conversation between stakeholders and less as a set of rigorous criteria allows space to consider EA's importance for program learning and development, rather than for the evaluation decision only. Given its functionality and relatively low financial and time costs, EA is well-placed to become a more widely recognized and applied tool in pre-evaluation, particularly for organizations with large program portfolios.

References

Davies, R. *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations*. Working Paper 40 Vol. London: UK Department for International Development, 2013. Web.

Leviton, Laura C., and Marjorie A. Gutman. "Overview and Rationale for the Systematic Screening and Assessment Method." *New Directions for Evaluation* 2010.125 (2010): 7-31. Print.

Miller, Candace, Kate Place, John Tyler, George Kafkas, and Kimberly Boller. "Evaluability Assessment: Fursa kwa Watoto." Mathematica Policy Research, Cambridge. December 2015.

Patton, Michael Quinn. *Utilization-Focused Evaluation*. Sage, 2008. Print.

Soares, Yuri, Alejandro Pardo, Veronica Gonzalez, and Sixto Aquino. 2010. "Ten Years of Evaluability at the IDB". IADB. http://www.oecd.org/development/evaluationofdevelopmentprogrammes/dcdndep/46436272.ppt.

Trevisan, Michael S. *Evaluability Assessment from 1986 to 2006*. 28 Vol. , 2007. Print.

Wholey, Joseph S. "Exploratory Evaluation." *Handbook of Practical Program Evaluation.* Eds. Joseph S. Wholey ed., Harry P. Hatry ed., and Kathryn E. Newcomer ed. 3rd ed.Jossey-Bass, 2010. 81-99. Print.

Fursa kwa Watoto Steering Committee (2015). FkW Programme Document. Harare.