

Magic Ingredients in Evaluability Assessments

Anne West, The Cadmus Group, Inc., Portland, Oregon
Ben Bronfman, The Cadmus Group, Inc., Portland, Oregon

ABSTRACT

What difference does an evaluability assessment (EA) make? It depends.

Conducting early EAs for utility energy-efficiency programs focusing on emerging and innovative technologies has identified a number of critical data elements and processes requiring attention. Revisiting these programs to conduct a full impact and/or process evaluation six to 12 months later revealed mixed results.

For some programs, the EAs were highly effective, leading to programs where necessary data were collected for evaluations, and early course corrections improved program delivery. For others, data were neither collected nor issues not dealt with.

Conducting evaluations months after the early evaluability offers several insights into the process that will improve the functionality and usefulness of assessments. We find that two ingredients must be present for early assessments to achieve their intended outcome and full potential:

1) Understanding: program managers and third-party implementers need to understand the issues, why data are needed, and how it all relates to successful program evaluations and verification of savings.

2) Follow-through: program managers and third-party implementers must follow through to collect the identified data or direct attention to potential issues. Evaluators must employ successive iterations of the assessment to ensure data are collected.

Introduction

Program evaluations in the energy industry are becoming increasingly more complicated, integrating process and impact evaluations and market assessments. We conduct evaluations across energy-efficiency and demand response programs, renewable energy technologies, and, more recently, climate change initiatives. Utilities that have never offered demand-side management (DSM) programs are now planning and offering them. Where we once evaluated programs one or two at a time for a given utility, we now evaluate suites of 10 or more offerings simultaneously. New third-party implementers are entering the business, along with new technologies or new applications of more mature technologies.

Evaluators continue to face challenges and problems in conducting evaluations within complex policy and management environments, which can obscure program goals and objectives as well as the types of information needed. Stakeholders and market actors may not understand the relevance of data collection or how it ties to evaluating programs later, or how data are used to verify energy savings. Complete and accurate data collection is critical as more sophisticated analyses are conducted. However, the evaluation community continues to see some of the same problems year to year and evaluation to evaluation. For example, data are incomplete, contact information is missing, and proposers do not fully research market assumptions.

At the same time, it is encouraging that evaluators are more commonly becoming involved early in the evaluation process. Further, there are additional tools to ensure data are being collected and programs can be evaluated. This paper discusses refining and further operationalizing one such tool: the Evaluability Assessment.

Evaluability Assessments

Evaluation literature often cites Joseph Wholey as the developer of EAs (Wholey 1979). Described by Wholey and others, the purpose of an EA is to determine whether decision makers and stakeholders have expressed measurable objectives and expectations, and to determine whether data exist, or will exist, which are necessary to fully evaluate a program. In some fields, the EA is used to determine whether the program is ready to evaluate, and the EA is conducted only when there are measurable outcomes. Others use the EA for formative evaluations to develop or modify programs (Trevisan 2007). Wholey further suggested and discussed using the EA to develop a “menu of monitoring and evaluation possibilities” to improve management and governance of public and nongovernmental organizations (Wholey 2005).

Trevisan (2007) in *Evaluability Assessment from 1986 to 2006* discussed the development and use of EAs, finding 900 sources referencing EAs. Of these, 23 were EA studies published across several disciplines. Fields included physical and mental health sciences, education, criminal justice, welfare and social justice, nonprofit performance measurement, and public agency and management performance. Some published papers discussed practical applications of EAs to improve program planning and the usefulness of evaluations. None of the published documents Trevisan reviewed included papers within the arena of energy-related evaluation. Literature searches by these authors and Trevisan found commonalities in approach and purpose, but little discussion or display of actual methods or tools developed for EA.

Bronfman (2008) discussed developing the EA methodology for the energy industry and incorporating it into the program planning, implementation, and evaluation processes for early assessments of 25 utility-sponsored programs. These programs focused on new and emerging technologies or innovative applications of more established technologies offered through the IDEEA/InDEE portfolio¹. Methods discussed in that paper were implemented in early EAs of 28 additional programs, focusing on new and emerging technologies in the subsequent program funding cycle for the same utility. Of late, the EA has been applied in utility program evaluations, most recently at utilities initiating DSM for the first time.

We use EA within the energy industry to inform the evaluation process, to clarify program theory and logic models, modify existing program data collection plans so programs can produce measurable results and accommodate evaluations, modify evaluation plans if warranted, and generally improve both program evaluability and the usefulness of evaluation plans. Opportune times to conduct early EA are when programs have been developed but not yet fully operational. This paper reviews some observations drawn from employing the EA within DSM and energy-efficiency evaluations. We further refine the model and methods operationalized for the 2005 evaluations specific to our discipline of energy-efficiency evaluation, as described by Bronfman.

What’s the Evaluation Plan Got to Do with Evaluability Assessment?

Evaluation allows us to look at program goals and objectives, fundamental program assumptions, and the intervention it offers, be it regarding energy-efficiency equipment or energy education. Evaluation provides an understanding about how an implemented program works as opposed to how planners designed the program. It offers an opportunity to improve the program’s usefulness and efficacy.

The EA offers a systematic framework for reviewing program backgrounds and logic. It assesses whether stakeholders are in common agreement on objectives and goals, and whether assumptions used

¹ Innovative Design for Energy Efficient Applications (IDEEA) and Innovative Designs for Energy Efficiency (InDEE).

accurately reflect the market. The EA offers the opportunity to assess and bring stakeholders into agreement on goals and objectives, and to discuss the intended outcome or result of the program with stakeholders, and gauge their perceptions of an unexpected or surprising result. These discussions can inform and further frame the uncertainty and untested assumptions to determine their appropriate place within the larger evaluation.

If the evaluation includes an assessment of program impacts, or we design the evaluation to reduce uncertainty around a specific technology or performance parameter, program implementers must record the specific data elements. The EA will determine whether data collection tools are adequate and will record data elements for the intended evaluation approach. If, for example, the evaluation intends to estimate free riders, free drivers, or spillover, specific data are required. While we collect data via highly targeted primary research, we cannot conduct primary research through participant surveys and site visits (for example) without participant and nonparticipant contact information or measure installation data.

Keeping the planned (preliminary) process and impact evaluation approaches in mind, we have intentionally conducted EA as exploratory exercises, early enough for program modifications that would facilitate full program evaluations.

We agree with others who state the strength of an early EA is it can potentially improve programs and their performance and ensure intended evaluations can actually take place. (Socio Economic 2003). In conducting an EA, the following issues should be noted:

- Apply the EA with a “light touch.” The approach is a quick, low-cost, and time-limited assessment. It can be built into the early management functions. These are designed to provide rapid feedback.
- Expectations should be realistic. EAs can help establish a framework to move forward. However, it is not possible to remove all the uncertainties, unpredicted outcomes, and new possibilities that typically appear when actually fielding programs.

Evaluability Assessment Approach

The approach and method used by Bronfman (2008) in the 2005 and subsequent EAs included:

- **Interviews** with utility measurement and evaluation (M&E) managers, utility program managers, third-party implementers who developed the program proposal, utility and implementation staff who managed the day-to-day operations, and staff responsible for data collection, storage, and retrieval.
- **Document review and analysis** of any available program documents, including the program description, implementation plan, statement of work, purchase orders, third-party contracts, and program theory *and* logic models. Reviews included technical proposals that described the methodology and assumptions used to estimate ex ante energy savings.
- **Articulating program theory and developing logic models** where none existed. Often, the theory is implicit where it should be stated explicitly. Likewise, intervention logic, intended activities, and outcomes should be diagrammed.
- **Document review and analysis** of the marketing approach, plans, and materials.
- **Examination of program application forms** to assess data elements intended for collection and to compare that to the list of data elements required for later analysis.
- **Examination of program tracking databases** where developed or sample outputs were made available to review, to assess the data collection plan, to identify data elements recorded in the database, and to determine which elements would be stored only in hard or soft copy outside the program tracking database.

- **Literature reviews** particularly focused on ex ante energy savings and whether DEER or other white papers supported assumptions. In addition, the California Evaluation Protocols were consulted to determine whether one of the impact evaluation methods could be used, given data collected or collectable.

These methods are consistent with a subset of those named in Trevisan’s research (2007). The research listed additional methods that may be applicable to other energy-efficiency focused evaluations. Additional methods named in other published studies included: site visits, focus groups, questionnaires and surveys, and an evaluability working group consisting of the evaluation team and program stakeholders. EA conducted after programs have matured sufficiently to collect data may include examinations of program statistics, examination of case files, and statistical analysis.

Following the EA approach discussed above, the evaluation team’s steps to finalize the EA process included:

- **Making evaluation decisions** where definitive information or data were desirable but not available. That is, determining whether the evaluation plans needed to change, whether evaluation timelines could be met, or whether stakeholders should and could address shortcomings.
- **Discussing deficiencies** with program managers and implementers who were key decision makers and instrumental in day-to-day operations; for example, identifying additional data capture needed on application forms to facilitate evaluation, and requiring changes to printed forms.
- **Explaining the purpose** for specific data needs. For example, implementers must collect contact information for program decision makers, whether or not they participated. Without contact information, savings cannot be verified, free ridership cannot be assessed, and market barriers and reasons for nonparticipation cannot be explored.
- **Strategic planning** in cases where we identified significant issues or required modifications.
- **Determining whether baseline M&V** was needed, and developing measurement and evaluation options, including metering and cost options.

These steps were consistent with comments in Wholey’s 2005 presentation to the World Bank’s International Finance Corporation. We explained the “reality of the situation” as well as the purpose of the program inputs (data) and use within the evaluation. We explored measurement and evaluation options, and used the EA to inform both evaluation and program decisions. (Wholey 2005).

Evaluability Assessment Results

Since many DSM and energy-efficiency programs require six months to a year or more of post-intervention data before impact evaluations are fully engaged, the approach for using early EA, particularly in these programs, is to determine whether we can evaluate the program later. That is, can we evaluate given data discussing the goals, objectives, assumptions, and pre- and post-intervention data collection plans. Carefully detailing what we do not know is as instructive as detailing what we do know. After completing the EA interviews and document reviews, we found a number of shortcomings and unknowns:

- Few programs included a program theory in the program proposal or subsequent document.
- Fewer still had logic models.
- A common understanding of program goals and objectives did not exist.
- Roles and responsibilities of various departments touching the program were not well defined or commonly understood.

- In some cases, no formal program description existed, particularly where no third-party implementer was involved.
- Some programs had not planned to collect participant and/or nonparticipant contact data, follow-up contact was critical to the evaluation.
- Some programs did not plan to keep records in their tracking database of the specific homes to which services were delivered, or, in a new construction program, to record the specific building plan along with the address where the home was built.
- In one case, documents did not clearly define and implementers did not appear to research, the baseline specific to the targeted geographic area and intended market.

These EAs produced data required to modify program elements, increase stakeholder awareness and understanding, and develop subsequent program theories and logic models; thus, the intent of the EA was satisfied: we determined some programs would not be evaluable unless implementers collected additional data, defined the baseline, or took some other action.

Returning to conduct the evaluations six months or more after the original EA and discussions with key stakeholders, we found some third-party implementers and program managers did not make required changes.

Since the intent of the EA was satisfied and initial follow up completed, why are there cases where the EA does not produce the expected outcome? Staff changes, business changes, institutional (even short-term) memory loss, time's passage, too much to do, lack of a business priority, a ball dropped, pressing financial or marketing issues, and a million other reasons come between the early EA and the full program evaluation, and the ability of evaluators to assess savings and processes.

In other cases, we found the EA process worked as far as we went, but it was not far enough. In some cases, the sample databases and program applications included necessary data elements, and there were no apparent deficiencies to correct. In practice, the database fields were not populated nor were baseline data collected on the application forms. In one case, because the program manager did not understand the need for specific data elements, the manager acquiesced to the implementer, who did not want to collect data specified on the form (and needed for evaluation).

Sometimes, programs targeting new technologies or new markets are proposed by industry experts, but the experts' assessments of their market are not always right. Thoroughly researched applications for new markets or new technologies are not always possible. The product and market fit may seem obvious to the technology experts; however, the science of market intelligence includes a high percentage of subjective content. What seemed like a plausible market and product during the EA was not, in fact, a good match. What sounds good on paper may not actually exist.

We offer four examples of programs in the 2006–2008 California funding cycle where market intelligence was wrong:

- Transforming the manufactured housing market to meet ENERGY STAR standards without understanding the underpinnings of the roles of retail salesperson.
- Introducing a technically superior lighting product to a market that could not discern it.
- Introducing low-pressure membrane filters for industrial process water filtration to customers who did not own the equipment.
- Targeting the last chance to influence building energy efficiency, but offering incentives after a plan is submitted for permit review and chances are very slim designers and building owners will revise plans, timelines, and budgets.

Faulty market intelligence is hard to detect during an early EA, as were these explorations. However, we propose follow-through, or successive iterations of the EA, can identify some flaws earlier in a program's evolution.

Evaluability Assessment Approach Revisited

While the EA discovered and documented deficiencies, and actively followed up with stakeholders, it was not enough to ensure program managers or implementers made the necessary modifications. We propose two ingredients must be present for EAs to achieve their intended outcome and full potential:

1) *Understanding.* Program managers and third-party implementers need to understand the “reality of the situation” (i.e., the issues, why data are needed, and how it all relates to successful program evaluations and verification of savings). These players may not be active stakeholders directly involved with evaluation planning or interested in evaluation outcomes. Yet, evaluation outcomes impact these people.

2) *Follow-through.* Program managers and third-party implementers must follow through to collect the identified data or direct attention to potential issues. Evaluators must employ successive iterations of the assessment to ensure data are collected or other identified deficiencies are addressed.

Checking back with stakeholders to assess their understanding is critical. Many stakeholders have not been previously involved with evaluation and will not understand the nature of the “reality” or the consequence of not following through on action items. The EA follow-up plans should include interviews with all key stakeholders and market actors involved with the action items. Typically, this includes the implementers responsible for program development and the day-to-day operations.

Follow-through occurs at various stages. Once we identify deficiencies through the EA and the actions needed, the EA plan must also include periodic assessments. Follow-up assessments or early check-in with the program managers and third-party implementers should determine whether data described are being collected, and whether all fields in the database are being populated.

Bronfman (2008) presented an Evaluability Assessment Project Flow Diagram. The Flow Diagram remains essentially unchanged after subsequent evaluations. We recommend using the elements of that flow diagram to envision the early EA process used in DSM and energy-efficiency EAs.

We propose minor additions to the Evaluability Assessment Template presented in the 2008 Bronfman paper. The template was originally developed as the systematic guide for the EA of the 2004–2005 and 2006–2008 program funding cycles for a Southern California utility. We have used the template in EAs for other energy-efficiency program evaluations. Table 1 shows the additions made to the original template, based on additional experience using the template with additional EAs.

Table 1. Evaluability Assessment Template Additions

Question
1. Who is the intended audience for the evaluation?
What are their information needs and how will they use the information?
What are the evaluation timelines? When are the evaluation results required?
Does the evaluation audience include the same people impacted by or expected to be involved with evaluation activities?
2. What type of evaluation is required or planned?
What data are needed to conduct this type of evaluation?
3. Is there a description of the staff that will operate the program?
Are staff responsibilities defined and understood by all who touch the program?
Are operational staff fully informed of the evaluation and purpose and activities?
4. Is there a description of the program?
5. Is there an explicit program theory or logic model
Is there a clear definition of a program participant and nonparticipant? What are the definitions?
6. Is there a marketing plan?
Does the marketing plan target the intended market with the appropriate message that could elicit program participation?

Conclusions and Recommendations

Early EAs seek to improve usefulness and timeliness of evaluations, maximizing their utility. After conducting the evaluations of the 2004–2005 IDEEA/InDEE portfolio and early EAs of the 2006–2008 IDEEA/InDEE portfolio, both focusing on new and emerging technologies or innovative applications of more established technologies, we find the conclusions and recommendations offered by Bronfman (2008) are still valid. The EA conducted for the 2006–2008 portfolio improved the overall evaluability of the programs, though some still fell short. Evaluations for other utility-sponsored energy-efficiency programs also point to the value of conducting an early EA.

Conclusion: Without some systematic EA and program manager training, proper evaluation, monitoring, and program performance data are lost, program performance suffers, and, ultimately, utility incentives and programs are at risk.

Recommendation: Program managers should be trained on what constitutes an EA, and how EA benefits program implementation. Program managers should also be trained to recognize the needs program evaluators raised in the EA to ensure appropriate information is available when evaluations commence.

Recommendation: The EA template should be included as part of future RFPs or within early program development to alert proposers and planners to their expanded responsibilities when projects move forward. This starts the process when the program has been or is being designed but is not yet operational. It is particularly worthwhile to introduce the EA at this time to assist in articulating the program theory, developing the intervention logic models, and make changes to facets of program offerings, as needed.

Recommendation: Evaluation professionals should be involved early in the program cycle to assist program managers in conducting an EA, design appropriate early M&V studies, recommend QC procedures, and conduct early program startup process evaluations.

Conclusion: Once we identify deficiencies through the EA and discuss the actions needed with stakeholders and others who will implement the actions, the EA plan must also include periodic follow-up assessments. Follow-up assessments or early check-in with the program managers and third-party

implementers should determine whether they took action (e.g., they are collecting and recording data described), and whether all the fields in the database are populated.

Recommendation: Plan and conduct follow-up assessments or early check-in with the program managers, and third-party implementers, and other actors.

Recommendation: Check back early with stakeholders to assess their understanding of the issues and actions needed. Ensure they understand the “reality of the situation” as well as the purpose of the program inputs (data) and their use within the evaluation.

References

Bronfman, Ben, Samiullah, Shahana, et al. 2008. Integrating Evaluability Assessment into the Program Planning, Implementation and Evaluation Process: Case studies from Southern California Edison’s IDEEA program portfolio. Association for Energy Service Professionals. 18th National Energy Services Conference Proceedings, January.

Evaluating Socio Economic Development. 2003 Sourcebook 2: Methods & Techniques. Evaluability Assessment.

Trevisan, Michael S. & Yi Min Huang. 2003. Evaluability Assessment: a Primer. Practical Assessment, Research & Evaluation, 8(20). <<http://PAREonline.net/getvn.asp?v=8&n=20>>.

Trevisan, Michael. 2007. Evaluability Assessment from 1986 to 2006. American Journal of Evaluation. September 2007 28: 290-303.

Wholey, J. S. 1979. Evaluation: Promise and Performance. Washington DC: the Urban Institute.

Wholey, J. S. 2005. Evaluability Assessment and Its Applications. Monitoring and Evaluation (M&E) Meeting, Presentation to International Finance Corporation, World Bank, May 9, 2005. <<http://www.ifc.org/ifcext/sme.nsf/Content/M&EMeeting>>