

Looking in the Mirror: Applying Metaevaluation to Energy Efficiency Program Evaluations

Brandy Brown, Western Michigan University, Kalamazoo, MI

ABSTRACT

Metaevaluation is used to systematically assess the quality of evaluation products, confirm that evaluations deliver sound findings and conclusions, are useful to the client, are credible, are ethically conducted, and are done as cost-effective as possible. Energy efficiency evaluation practice can augment efforts to improve its methods, policies and processes by integrating metaevaluation into its practice. This paper examines the applicability of metaevaluation to energy efficiency evaluation by presenting lessons learned from a metaevaluation of recent publically available energy efficiency evaluations. It concludes by suggesting next steps in incorporating metaevaluation into energy efficiency to empower programs and portfolio managers with more credible findings and results for incorporation into their programs, thus, making larger savings for the same effort a possibility.

Introduction

Evaluation by nature is multidisciplinary in that it is not restricted to one domain, but applies to many. One overarching theme within evaluation is that evaluators commonly seek go beyond proving how a program is performing; they seek to improve the program. Improving and informing program practices is at the root of the evaluation theory. In fact, evaluators have used this focus on improvement reflectively, in an introspective examination of evaluation practice. Over the last few decades, this reflectiveness has brought attention to the variety of evaluation in distinct disciplines. Evaluation has assorted, context specific applications; however, the overarching fundamentals are the same. Evaluators operate under certain theoretical assumptions that serve as the guide for conducting evaluation. Theories act as the lens through which an evaluator sees when selecting research questions and accompanying research design and methods. Energy Efficiency evaluation, therefore, is conducted with certain assumptions and steps that are unique to our industry. We must proactively examine our work with in this context and in comparison to the larger evaluation community to ensure its progress.

Over the last twenty years, evaluators have emphasized evaluating their work to ensure that they are producing quality, methodologically sound evaluations. Evaluators continue to build a base of empirical evidence on the practice of evaluation thereby elucidating areas for improvement. Indeed, there is increased scrutiny of the value and cost-effectiveness of energy efficiency programs due to disagreements around the quality and validity of the findings of energy efficiency program evaluations. In the last few years the state and federal attention to energy efficiency evaluation has grown. Inconsistent methods open the door to attacks on the validity of evaluation results and confuse sound methods with junk science. Energy efficiency evaluators understand that practice may differ from theory and desire to communicate best practices. However, energy efficiency evaluation currently lacks empirical evidence that depicts energy efficiency evaluation as practiced does not exist.

Current calls for national standards of practice have become a focal point within the field, as energy efficiency evaluators desire to communicate methodological best practices. But prior to that work, there must be a thorough study on the current common practices in energy efficiency evaluation. Because this field of energy efficiency is relatively young in comparison to other fields, no empirical studies have been published that explore what methods are used to evaluate energy efficiency programs. Evaluating the

practice of energy efficiency is one way of establishing an empirical baseline for the common methodologies underlying energy efficiency evaluation. After we identify the common practices empirically, we can explore the quality of our work and revise standards accordingly. We can learn high quality lessons by evaluating our work, searching for patterns across programs, and finding connections across other disciplines. By increasing evidence that supports the utility and value of energy efficiency evaluation, we will increase the assurance in its significance. This need for stronger empirical evidence of energy efficiency evaluation directs attention toward a method for researching evaluations: metaevaluation.

The remainder of this paper describes metaevaluation and sheds light on potential areas for incorporating metaevaluation into energy efficiency evaluation practice. All insights described in this paper resulted from the author's experience conducting a qualitative, descriptive study that systematically reviewed the central methodological characteristics of energy efficiency evaluations by applying metaevaluation to 91 energy efficiency evaluation reports. The data was examined for emerging themes in accordance with program evaluation standards and research design principles.

Overview of Metaevaluation

Metaevaluation provides a means of systematically assessing the quality of one or more evaluation products and to examine common themes. Its purpose is to confirm that evaluations deliver sound findings and conclusions that are useful to the client, credible, ethically conducted, and as cost-effective as possible. It retrospectively examines finished evaluations and addresses overarching questions across projects. Metaevaluation is different from meta-analysis, which synthesizes quantitative data from multiple studies to create a more comprehensive understanding of a particular topic. For example, a meta-analysis of energy efficiency impact evaluations would gather the qualitative data from impact evaluations of the one program type, like a HVAC program, and synthesize the findings to gain insight to the energy savings of HVAC programs evaluations within a defined jurisdiction. Rather metaevaluation judges the merit, worth and significance of an individual evaluation or set of evaluations. A metaevaluation of energy efficiency impact evaluations would examine the quality of one or more evaluations based on specific criteria, such as accuracy. Therefore, though meta-analysis would need to be program specific. Metaevaluation is not.

Within the field of evaluation, the key purpose of metaevaluation is as a research tool in to assess evaluation practice and compare it to evaluation standards. Evaluation theorists agree that metaevaluation could increase our collective knowledge about evaluation practice and eventually improve evaluation utility through expanding research on evaluation and assessing methodological rigor. There have been several notable efforts to promote the practice of systematically testing evaluation theory and methods of practice. Shadish, Cook, and Leviton look at different evaluation theories and their applications. Coryn et. al (2011) empirically examined the practice of theory-driven evaluation. Christie's work on bridging the gap between evaluation theory and practice notes empirically that evaluation theory is applied in practice is not typically fully implemented. Christie and Miller have each examined how evaluation theory relates to evaluation practice. Miller developed a framework for studying the relationship of theory to practice and showed that adoption of evaluation theory does influence strikingly different methods. Extending this practice of research on evaluation to energy efficiency allows us to demonstrate and validate industry best practices. Moreover, it allows us to advance debates about appropriate and inappropriate uses of evaluation design and methodology within our field. Some within our industry claim that it has adapted a "business as usual" approach to energy efficiency evaluations, which ignores contemporary evaluation methods. Some evaluators prefer experimental designs and related methodology, while others prefer non-experimental or quasi-experimental designs. Furthermore, many evaluators focus on increasing research quality and rigor. There are several approaches and standards to employ in assessing methodological rigor, however one requires metaevaluation to systematically apply standards to finished evaluation.

Several scholars have developed metaevaluation tools to ease the application of evaluation standards to program evaluation. One of the best-known standard tools is the Program Evaluation Metaevaluation

Checklist¹ originally published by the Joint Committee of Standards for Educational Evaluation in 1988. This checklist operationalizes a set of widely accepted program evaluation standards and assigns a metaevaluation score based on an evaluation's utility, feasibility, propriety, accuracy, and evaluation accountability. Within energy efficiency, evaluation standards vary by jurisdiction. This variation makes it difficult to gauge which standards are most applicable, however, is still plausible to develop metaevaluation criteria for energy efficiency evaluations with respect to our unique approaches. For example, in the metaevaluation I conducted on energy efficiency evaluations, I selected commercial and industrial program evaluations located in the same jurisdiction and I differentiated between process and impact evaluations. By utilizing specific steps for case selection increased the likelihood that the cases would employ similar methodologies and be accountable to similar standards. Similarly, we can use metaevaluation in energy efficiency by comparing finished evaluation products with relevant evaluation standards.

Conducting Metaevaluation in Energy Efficiency

Conducting a metaevaluation entails applying evaluation methodology to an existing evaluation. Although the methodology may differ depending on whether the subject evaluations are process evaluations or impact evaluations, the general evaluation methodology is the same. One could examine whether impact evaluations adhere to accepted guidelines or assess them on evaluation standards such as utility, feasibility, propriety, accuracy, and evaluation accountability. Process evaluations metaevaluation would require a different method of assessment from impact evaluations in terms of methods, i.e. you would not hold the methods of a process evaluation to the same standards of rigor as an impact evaluation. However, you could develop a technique to assess rigor of a process evaluation based upon qualitative standards of inquiry. Also, you could choose to focus on a broader context-specific standards or focus on evaluation of a single program type, like evaluations of agricultural energy efficiency programs. In any case, the general framework for evaluation methodology applies and is delineated into five actionable steps:

- (1) Defining the purpose of evaluation;
- (2) Identifying and organize evaluation criteria;
- (3) Collect data and compare to the evaluation criteria;
- (4) Determining merit; and
- (5) Synthesizing results.

The first step in conducting a metaevaluation within energy efficiency is to determine the purpose of the study. Are you seeking to assess the evaluations utility, accuracy, or adherence to standards? Each of these items would require a different metaevaluation approach. For example, to assess an evaluation's utility you would need to contact the evaluation end-users (utility or program staff) and ask them how useful the evaluation was. To assess accuracy or adherence to standards, you would need to review the evaluation methodology through a document review.

The next step would be to identify and organize evaluative criteria. Following the previous example of assessing evaluation utility, you would need to determine what utility means and develop a way to rank it. For example, you could operationalize utility into separate indicators, each designed to measure some subset of evaluation use. You may decide to ask whether the utility administrator or program staff acted on any of the evaluation findings or whether they changed the program design as a result. Each of these indicators would be organized in preparation for the next step. It is important to note that several instruments have been developed in different fields to measure common evaluation performance criteria. Since creating and validating a new instrument can be a lengthy process, we may benefit from adapting already validated instrumentation from other fields.

¹ http://www.wmich.edu/evalctr/archive_checklists/program_metaeval.pdf

The third step would be to collect data and compare it to the evaluation criteria. For example, you would interview program administrators and record their responses to different aspects of evaluation utility. It is plausible to assign a rank or score for each aspect of utility at this time. Next, the fourth step involves inferring the merit and importance of the score relative to the ideal case for the criteria in question. For example, you would need to compare the utility rankings of the evaluations to accepted standards and describe the result. The final step would be to synthesize the information and determine the importance of the results across a single criteria. For instance, how useful was the evaluation to the program administrators overall? How useful was the evaluation on the various aspects of utility? Using these actionable steps, it would be possible for evaluators to apply metaevaluation to evaluations in our industry. Metaevaluation can be used on process and impact evaluation

Limitations of Metaevaluation in Energy Efficiency

A thorough metaevaluation approach would likely include document review and interviews to gather information and perspectives from key stakeholders. However, this is not always possible as evaluation reports are sometimes the only tangible byproduct of completed evaluations. Therefore, evaluation reports often become the primary data source for metaevaluation. As evaluation reports are not always a comprehensive report of evaluation activities, assessing quality and compliance to standards is a limitation of metaevaluation. In my review of energy efficiency evaluation reports evaluation methodology was largely absent in a sample of 91 energy efficiency evaluations. For instance, nearly one third of the evaluation reports in this sample did not include a methodology section, over a third did not specify sampling methods, and a majority of energy efficiency evaluation reports did not mention what methods were used for analysis. In general, accurate metaevaluation assessments of quality are impeded by the absence of information on specific evaluation methods. However, evaluation reports within our industry are written for our intended audience, who may be more interested in evaluation results than evaluation methods. We would need to address the absence or methodological descriptions in our reports if we are to embark on comprehensive energy efficiency metaevaluations in the future.

Another limitation is the applicability of general evaluation standards may not be applicable in our context. Therefore, tools developed on these standards may not be useful for application in energy efficiency. For example, I applied the Metaevaluation Checklist to energy efficiency evaluation but found that only one standard, accuracy, was relevant to energy efficiency evaluation. Furthermore, we need to be cognizant of our intra-disciplinary distinctions as well. For example, standards used to assess an impact evaluation do not necessarily apply to the standards that would be used to assess process evaluation. However, depending on a researcher's methodological tilt, one could argue that principles of credible evidence and justifiable conclusions apply to each. It is important that metaevaluation is tailored to be context specific so that it does not render invalid conclusions.

Conclusion

As we work to establish methodological best practices, we need a continuous examination of our methods and practice. Metaevaluation is a useful tool that would help us engage in this continuous assessment of the energy efficiency evaluation in practice and provide stronger supporting evidence for energy efficiency programs in the future. The improved perspective on energy efficiency evaluation that will result from metaevaluation will allow us to evolve our evaluation practices and thus, provide more actionable data to improve energy efficiency programs.

My experience conducting a metaevaluation of energy evaluation shows that conducting metaevaluation is not a stretch from our current evaluation practice. However, it does require a certain level of methodological expertise, time, and resources to conduct. For example, if a metaevaluation focuses on methods, then the evaluator should be knowledgeable in methodology. Likewise, if the metaevaluation

focuses on Commercial New Construction energy efficiency programs, the evaluator should be experienced in that area. Given our often resource constrained programs and the boundaries within which we conduct evaluation, we may need to begin preparing our field for future efforts rather than embarking on full metaevaluation studies in the near future. Specifically, we need to examine whether we can include full, in-depth evaluation reports for fellow evaluators with information that may not be interesting to our clients.

Metaevaluation is one tool at our disposal in examining energy efficiency evaluation to ensure that we continue to deliver sound, useful to our clients. We can use it to critically examine the current practice of energy efficiency evaluation and contemplate how to advance its practice. Metaevaluation, and our increased understanding of our work, will help us to empower programs and portfolio managers with more credible findings and results for incorporation into their programs, thus, making larger savings for the same effort a possibility.

Acknowledgements

Thank you Dr. Chris Coryn, Rob Bordner, and Dr. Rafael Friedmann for your continued support and willingness to challenge my thinking on evaluation.

References

- Alkin, M. C. (2004). *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed.). Thousand Oaks, CA: Sage.
- Chelimsky, E. (1998). The Role of Experience in Formulating Theories of Evaluation Practice. *American Journal of Evaluation* , 19 (1), 35-55.
- Christie, C. A. (2003). What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory. *New Directions for Evaluation* , 97, 7-35.
- Cooksy, L. J., & Caracelli, V. J. (2005). Quality, Context, and Use: Issues in Achieving the Goals of Metaevaluation. *American Journal of Evaluation* , 26 (1), 31-42.
- Coryn, C. L. (2009, September). Contemporary trends & Movements in Evaluation: Evidence-based, participatory & empowerment, & theory-driven evaluation. Kalamazoo, MI, USA.
- Coryn, C. L., Noakes, L. A., Westine, C. D., & Schroter, D. C. (2011). A Systematic Review of Theory-Driven Evaluation Practice from 1990 to 2009. *American Journal of Evaluation* , 32 (2), 199-226.
- Datta, L. (2007). Looking at the Evidence: What Variations in Practice Might Indicate. *New Directions for Evaluation* (133), 35-54.
- Davidson, J. (2004). *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: Sage.
- Donaldson, S. I., & Christie, C. A. (2005). The 2004 Claremont Debate: Lipsey vs Scriven - Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard? *Journal of Multidisciplinary Evaluation* , 2 (3), 60-77.

- Donaldson, S. I., Christie, C. A., & Mark, M. M. (2009). *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA: Sage.
- Dougherty, A., Sutter, M., & Randazzo, K. (2012). *Supporting and Facilitating Experimentation in Energy Efficiency: Exploring New Validity Constructs to Support Inter- and Intra-Organizational Paths to Success*. Washington, D.C.: American Council for an Energy-Efficient Economy.
- Friedman, R., (2011) A Fresh Look at Evaluation to Support Energy Efficiency in the 21st Century. Bostom, MA: IEPEC
- Gullickson, A. R., & Hanssen, C. E. (2006). Local Evaluation in Multisite STEM Programs: Relating Evaluation Use and Program Results. *New Directions for Evaluation* (109), 87-103.
- Hanssen, C. E., Lawrenz, F., & Dunet, D. O. (2008). Concurrent Meta-Evaluation: A Critique. *American Journal of Evaluation* , 29 (4), 572-582.
- Henry, G. T., & Mark, M. M. (2003). Toward an Agenda for Research on Evaluation. *New Directions for Evaluation* (97), 69-80.
- International Energy Program Evaluation Conference. (2012, August 31). *National Energy Efficiency EM&V Standard*. Retrieved from Iepec.org
- Kushler, M., Nowak, S., & Witte, P. (2011). A National Survey of State Policies and Practices for the Evaluation of Ratepayer-Funded Energy Efficiency Programs. Washington, DC: ACEEE.
- Mark, M. M. (2007). Building a better evidence base for evaluation theory: Beyond general calls to a framework for types of research on evaluation. In N. L. Smith, & P. R. Brandon, *Fundamental Issues in Evaluation* (pp. 111-134). New York: Guilford.
- Messenger, M., Goldman, C., & Schiller, S. R. (2010). *Evaluation Planning, Methods, and Practices Across the United States: Emerging Issues and Opportunities for Cooperation*. Asilomar: ACEEE.
- Miller, R. L. (2010). Developing Standards for Empirical Examinations of Evaluation Theory. *American Journal of Evaluation* , 31 (3), 390-399.
- NAPEE. (2006). *National Action Plan for Energy Efficiency*. Washington, D.C.: NAPEE.
- Patton, M. Q. (2001). Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned. *American Journal of Evaluation* , 22 (3), 329-336.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage.
- Scott-Little, C., Hamann, S., & Jurs, S. G. (2002, December). Evaluations of After-School Programs: A Meta-Evaluation of Methodologies and Narrative Synthesis of Findings. 387-419.

- Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Newbury Park: Sage.
- Scriven, M. (2007, February). *Key Evaluation Checklist*. Retrieved August 31, 2012, from The Evalaution Center: http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf
- Scriven, M. (1980). *The Logic of Evaluation*. Inverness, CA: Edgepress.
- Segerholm, C. (2003). Researching Evaluation in National (State) Politics and Administration: A Critical Approach. *American Journal of Evaluation* , 24 (3), 353-372.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Shadish, W. R., Cook, T. D., & Leviton, L. (1991). *Foundations of Program Evaluation*. Thousand Oaks, CA: Sage.
- Stufflebeam, D. L. (1999). *Program Evaluations Metaevalaution Checklist*. Kalamazo, MI: Western Michigan University.
- Stufflebeam, D. L. (2011). The Metaevaluation Imperative. *American Journal of Evalaution* , 22 (2), 183-209.
- Stufflebeam, D. L., & Coryn, C. L. (2012). *Evaluation, Theory, Models, & Applications* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. (2001). The Meta-Evalaution Imperitive. *American Journal of Evaluation* , 183-210.
- Stufflebean, D. (1980). An Interview with Daniel L. Stufflebeam. 2-4.
- TecMarket Works. (2006). *The California Evaluation Framework*. Oregon, WI: California Public Utilities Commission.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evalaution Users* (3rd ed.). Sage.
- York, D., Witte, P., Nowak, S., & Kushler, M. (2012). *Three Decade and Counting: A Historical Review and Current Assessment of Electric Utility Energy Efficiency Activity in the States*. Washington, D.C.: American Council for an Energy-Efficient Economy.