# Making Evaluation Reports Easy to Determine “What Works, Why, and for Whom”

*Mary Sutter, Grounded Research and Consulting LLC, Oakland, CA*

*Brandy Brown, PhD., CLEAResult, East Lansing, MI*

# ABSTRACT

As evaluation of energy efficiency programs across the country undergoes higher scrutiny, how the industry reports results needs review and reform. Absent such change, stakeholders will question the credibility of results. Through discussions of the results of a systematic review within the commercial sector and a systematic review of energy efficiency evaluation methods, this paper shows that evaluation reports typically lack useful data in several important areas. Specifically, this review highlights the difficulty in determining what works, why, and for whom – all fundamental questions that evaluation should regularly answer. Using information culled outside of energy efficiency, the authors provide several examples of how evaluation reporting could be improved to make it easier for both readers of individual reports and those undertaking future systematic reviews to understand what works, why, and for whom.

# Introduction

Multiple disciplines outside of energy efficiency render a larger understanding of issues, trends, and patterns within research on a given subject through the use of systematic review, and use clearinghouses to overcome the difficulty of collecting research on specific topic areas. Clearinghouses serve as a repository for high quality research and exist to help industries describe “what works and why”. As examples, consider that members of the Cochrane Collaboration (<http://www.cochrane.org/>) gather and summarize the best evidence from research in the health care industry. Members of the Campbell Collaboration ([www.campbellcollaboration.org](http://www.campbellcollaboration.org)) perform systematic reviews to summarize the research evidence on the effects of interventions in crime and justice, education, international development, and social welfare. The Institute of Education Sciences created the What Works Clearinghouse (<http://ies.ed.gov/ncee/wwc/>) to tease out what works in education so that educators can make evidence-based decisions. The University of Wisconsin maintains “What Works Wisconsin” (<https://fyi.uwex.edu/whatworkswisconsin/>), an effort that summarizes and disseminates information on effective policies, practices, and programs for youth and their families, schools and communities. Furthermore, the American Evaluation Association (AEA) regularly sponsors research on evaluation-themed articles and studies. The AEA interest in research on evaluation is focused on the relationship between evaluation theories and practice in an effort to advance the discipline (Coryn, et al., 2011). As can be seen by these activities, it is not new for researchers to systematically gather information about program interventions, analyze the data, and disseminate knowledge on evidenced-based effects. Many of these efforts have been in place for ten to twenty years. However, while evaluation of energy efficiency programs has been occurring for over thirty years, researchers have made no attempt to create a similar type of evidence-based peer review process.

While there may have been good reasons in the past for not creating this type of process, the authors believe that now is the time to seriously consider it for four reasons:

1. New people are joining the industry and require a useful place to learn about good program designs.
2. Program designers are tasked with creating programs that provide for even more energy efficiency[[1]](#footnote-1).
3. Potential changes in policy may occur that tie energy efficiency to climate change topics.
4. The number of stakeholders have increased over time (Vine 2012) and these people have multiple and varied interests in the outcome of energy efficiency programs.

The authors have been around long enough to see old program viewpoints repackaged and thought of as new, and while uncertain why this is occurring, it could be due to program designers lacking a ‘macro’ viewpoint supported by easily obtained broad knowledge, such as that available with clearinghouses of information. The authors were interested in determining whether systematic review, if taken up by the energy efficiency industry, is feasible and if this type of review could succinctly present information so that stakeholders can clearly see what is working, why, and for whom.

Unfortunately, for reasons offered in the forthcoming pages, two independent reviews of energy efficiency programs highlighted the wide variation in evaluation reporting and underscored that the energy efficiency industry cannot easily implement a systematic review process similar to other evaluation areas. However, through easy changes to reporting, our industry can begin to build the program design knowledge that will enable a future system so that stakeholders can understand what works, why, and for whom.

# Description of Systematic Reviews

A systematic review is not the same as a literature review. “A systematic review is a key step towards finding what works. It uses transparent and explicitly defined procedures to find, evaluate and synthesize the results of relevant research. The process is designed to minimize bias and to give as accurate as possible an indication of the effect of an intervention or program.”[[2]](#footnote-2) According to another site that uses systematic reviews:

* A literature or comprehensive review brings together information on a topic in order to provide an overview of the available literature on a certain subject. Research materials are gathered through searching one or more databases and qualitatively brought together in the review. Literature reviews can be the first step in perusing a topic for a further study to get an idea of the current state of the science available but they can also be their own publication.
* Systematic reviews look at a topic more in depth using a scientific method. By looking at not only the available literature, but also theses/dissertations, abstracts/conference proceedings, and other grey literature sources, systematic reviews seek to be all-encompassing in showing results on a topic. To complete a systematic review, a team of researchers select a question to be answered and specify eligibility criteria for their resources before synthesizing the information to answer their question. Multiple databases are searched in order to find every possible article on the topic. Not only are the results of the searches presented, but the search strategy, assessments and interpretations of research are also included in this form of review.[[3]](#footnote-3)

As an example, within work performed for the Campbell Collaborative, systematic reviews have four clear requirements:

* Clear inclusion/exclusion criteria
* An explicit search strategy
* Systematic coding and analysis of included studies
* Meta-analyses (where possible)

A systematic review can go beyond a typical meta-analysis in which results from any study included in the meta-analysis must be of a specific format. Systematic reviews can include any number of studies and any study design (i.e., experimental, quasi-experimental, or qualitative designs are acceptable). Having this ability to bring multiple study designs into a strategic review enables broader inclusion and stronger results.

# Application of Systematic Reviews to Energy Efficiency Reporting

The two reviews described below took place several years apart and highlight the difficulty our industry faces in using past evaluations to support new program designs. Each of these efforts applied an improvement and accountability approach which commonly seeks to promote program improvement through assessing program processes and outcomes, while informing stakeholders of results. This allows all program stakeholders to make informed decisions regarding what works and for whom (Stufflebeam & Coryn, 2014). The first review assessed the methodological patterns found in energy efficiency evaluation, while the second review examined commercial sector contractor training. Both highlight the need for modifying reporting practices to provide our industry with the required program design knowledge to build an understanding of what works, why, and for whom.

## Energy Efficiency Methodology Review

A Metaevaluation of Energy Efficiency Evaluation by Brown, 2014, was designed to describe and catalog the central methodological characteristics of energy efficiency evaluations. The systematic review of 91 energy efficiency evaluations documentedthe commonalities and differences among them and assessed conformance to industry standards of evaluation (Brown, 2014). The remainder of this section describes the methods of this study and findings that shine a light on energy efficiency evaluation reporting practices.

To commence the study, data collection activity involved an extensive search for energy efficiency evaluations within formal databases such as the Consortium for Energy Efficiency (CEE), the American Council for an Energy-Efficient Economy (ACEEE), International Energy Program Evaluation Conference (IEPEC), Educational Resources Information Center (ERIC), the Northwest Energy Efficiency Alliance (NEEA), the California Measurement Advisory Council (CALMAC), and New York State Energy Research and Development Authority (NYSERDA). Search terms included combinations of energy efficiency and evaluation phrases such as: *energy efficiency, commercial, industrial, evaluation, program evaluation, process evaluation, and impact evaluation.*

This search produced 348 evaluations, which were subsequently filtered by program and scrubbed by removing market studies or evaluation reports that lacked program descriptions and/or outcomes. The remaining 334 evaluations were scrutinized for inclusion with selection criteria of program characteristics, evaluation characteristics, temporal relevance, and transparent outcome data (adapted from Scott-Little et. all 2002). In total, 91 evaluation reports (31 impact evaluations and 60 process evaluations) met the inclusion criteria and were utilized for the metaevaluation and systematic review of energy efficiency evaluations.

Evaluating the methodological quality of the selected evaluations included an examination of the extent to which the reported methods in the sampled evaluation reports were congruent with the California Energy Efficiency Evaluation Protocols: Technical, Methodological and Reporting Requirements for Evaluation Professionals (TecMarket Works Team 2006)[[4]](#footnote-4), and whether the energy efficiency evaluations included standard research design elements. (Shadish, Cook, & Campbell, 2002).

Overall, the study findings elucidate the need to reform current reporting practices because the majority of evaluation reports typically lack useful data in several important areas. Taken as a whole, the absence of useful data prevents the assessment and eventual informed use of the research.

The ability to accurately assess evaluation for methodological quality depends on the presence of detailed descriptions of evaluation methods. This information was largely absent in the sample of energy efficiency evaluations reviewed in this study. 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 65 (approximately 70%) did not mention what methods were used for analysis. In general, accurate assessments of quality were impeded by the absence of information on specific evaluation methods.

In terms of the methodological commonalities and differences of publicly available energy efficiency evaluations, evaluations were more alike than different. Typical methodological characteristics among this sample of energy efficiency evaluations largely consisted of non-experimental design, with non-probability convenience sampling, with data collection relying heavily on document review, interviews, and surveys. Additionally, the evaluation reports do not describe methods of analysis. Although an assessment of common potential threats to validity could not be completed due to incomplete information, this research was able to identify clear threats to statistical conclusion validity, internal validity, and construct validity with impact evaluations accounting for the statistical conclusion validity issues, and process evaluations accounting for the internal validity and construct validity threats.

Even though this assessment focused on completed evaluation reports from a fixed time period in one jurisdiction, and were interpreted through the lens of a single reviewer, these results have implications for the way in which energy efficiency evaluation is reported, specifically in relation to applicable standards, approaches, methods used, and the intersection of those issues in practice. This study showed that although energy efficiency evaluation protocols/standards do exist, implementation of these standards in practice varies greatly. This finding speaks to the need for greater professional reporting standards for both impact and process evaluations. Any set of standards should allow for the audiences of energy efficiency evaluation to adequately assess the quality of evaluation results and determine the importance of those results in terms of designing successful energy efficiency programs. Failing to present credible evidence and demonstrate justifiable conclusions calls hard work into question. As such, evaluators need to substantiate evaluation claims by walking their audience through the analysis and enabling the readers to make informed decisions about what works, why, and for whom.

## Commercial Sector Review

The author knows that in many areas of the country, commercial programs provide significant savings, yet are always under pressure to increase those savings. Because many programs include vendors within the program design, the author wanted to understand if training of commercial sector contractors increased program savings. This is a relevant question as it costs money to add a training component to a commercial program, even one that already uses contractors. Good program design may need to include training, but there is no hard evidence that by doing so helps the program obtain more savings. The systematic review included commercial and market effects studies, impact and/or process studies, and excluded residential, industrial, or agricultural sector studies, impact only or market characterization studies. Searches included study reports and conference papers, but no grey literature. After scanning through 331 documents and 55 papers within 19 conference proceeding, the author narrowed the studies for further review to 23 studies and 11 conference papers. After close review of only eight of the 34 documents, it became clear that determining savings from program training efforts simply could not be accomplished given the current reporting. For example, while half of the reports included savings and process information (which had potential to look at both savings and contractor training), only two reports included sufficient information about the program design to begin to understand program implementation and whether contractors were included in the program and trained as part of the process. Within those two, no specifics were provided about training that would enable us to understand if the training was short or long or what type of topics were covered within the training (as some trainings are simply about how to fill in applications while others help contractors learn to ‘sell’ energy efficiency). As such, after this short review (and ~40 hours into the overall effort), the author abandoned hope of performing a useful systematic review and turned, instead, to how reporting could be improved.

# Known Energy Efficiency Reporting Templates

# Templates can help transition to better reporting. However, before delving into how evaluators could improve reporting to enable systematic review, this paper would be remiss if it did not discuss what type of energy efficiency reporting are currently available. Multiple organizations have foreseen the need for warehousing energy efficiency data and created templates that detail how evaluators must provide the data so that others can analyze the information in the databases. Table 1, below, describes the databases (which have specified data templates) about which the authors are aware (and acknowledge that others may also be available).

Table 1. Known Energy Efficiency Databases

| Database Manager | Database URL | Purpose of Information |
| --- | --- | --- |
| NEEP | <http://www.neep.org/initiatives/emv-forum/regional-energy-efficiency-database> | Serves as a regional platform for the consistent reporting of Northeast and Mid-Atlantic electric and gas energy efficiency programs.  Currently, REED contains program level data for the following ten states: Connecticut, Delaware, the District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New York, Rhode Island, and Vermont. |
| CPUC | <http://eestats.cpuc.ca.gov/Views/EEDataPortal.aspx> | The public site for California Program Administrator (PA) energy efficiency program data (PA-reported measure-level program tracking data).  This site currently has data for program cycle 2010-2013 and 2013-2015 |
| LBNL | <https://emp.lbl.gov/publications/flexible-and-consistent-reporting>  Standardized Annual Reporting Workbook (v1.1 November 2015) | Using data that efficiency program administrators report to state regulators, LBNL maintains their Demand-Side Management (DSM) Program Database, now at nearly 6,000 program-years of program spending and impacts data from 36 states. According to the website, LBNL uses the database to characterize and inventory efficiency programs and to calculate and report on the cost of saving energy.  The cost of saved energy can be used for:   * Weighing multiple energy demand and supply resource options * Comparing efficiency program performance * Forecasting loads and the role of efficiency * Assessing Energy Efficiency Resource Standards (EERS) and other efficiency policies * Assessing options for compliance with environmental regulations * Enabling better insight into future efficiency spending and savings |
| LBNL | <https://emp.lbl.gov/what-it-costs-save-energy> | The Excel-based template is designed to produce consistent, useful metrics on program investments and performance for small to medium-sized administrators of public power efficiency programs, including American Public Power Associate members. |
| LBNL | <https://emp.lbl.gov/projects/eproject-builder> | The electronic Project Builder (ePB) standardizes data collection for ESPC projects nationwide—across ESCOs and all market sectors, including: federal, state, local, K-12, universities and colleges, public housing, health, industrial and private commercial. ePB provides authorized users with a streamlined, standardized, and secure online platform for collecting, housing and reporting their ESPC project data.  ePB enables energy service companies (ESCOs) and their customers to:   * Upload, track and access ESPC project-level information for the life of the performance contract * Quickly generate data for project and portfolio reports * Develop project scenarios using standardized amortization calculations * Benchmark new ESPC projects against historical project data |
| The Climate Registry | <https://www.theclimateregistry.org/thoughtleadership/energy-efficiency/> | The National Energy Efficiency Registry will allow states to track initiatives within their own programs as well as demonstrate progress towards energy goals and potential compliance with existing and future state and federal environmental regulations. |
| U.S. Energy Information Agency | <http://www.eia.gov/efficiency/programs/inventory/> | The intent of making this information available is to facilitate further research by enabling the rapid location and comparison of EM&V reports in terms of geography, scope, and other relevant characteristics. This inventory is national in scope with the exception of California, for which the CALMAC system provides a ready source for EM&V information.  EIA did not plan to maintain this spreadsheet over time, so interested parties will need to verify any changes to EM&V documents or records status made after the spring of 2013. |

The above databases are needed for purposes of tracking savings and cost effectiveness across the nation, as described in Billingsley (2015)[[5]](#footnote-5) Their templates are detailed, but with little information on specific program design that would enable a systematic review. Beyond even a systematic review, though, Klass (2015) also indicates that evaluation of existing programs are severely hampered by lack of consumption data, and several states are considering creation of consumption databases. As this possibility becomes a reality, it is important to connect consumption data to energy efficiency programs as this could provide a rich set of information on what is occurring outside of a program and be a piece of understanding what is working for whom.

For now, the efforts to date are targeted towards the end goal for energy efficiency programs – the savings. While these efforts are absolutely required, they fall short of collecting useful design information. The next section presents information for inclusion in reports to enable a systematic review and a template that evaluators could deploy and indicate where a specific data piece is already being collected in one or more of the databases described above.

# Three Areas for a Potential Expanded Energy Efficiency Reporting Template

Many reports provide information that could be used for a systematic review, but the data is difficult to extract and lacks consistency. Because the authors believe that starting small has a higher likelihood of being implemented (as opposed to creating a full schematic for reporting), this paper puts forward only three main areas where evaluators could choose to consistently include in their reports: 1) information about the program, 2) information about the evaluation, 3) savings results (if any) and a short description of findings in ‘plain language’ that anyone can understand.

The following sections expand on each of these areas.

**Information about the Program**

The commercial review brought out several areas that made it difficult to determine if an evaluation would be useful for systematic review. For examples, only two of the five relevant reports included program budgets and none applied a universal program type (although inclusion of a program type was not expected). Many reports did include a relatively good list of activities, but none had information on the expected outcomes of program interventions.

The bullets below indicate the four specific program description areas to include and the reasons why these are important.

* *Program budget range* *for evaluation period*– This information lets a reviewer understand if this is a large or small program in terms of the absolute value. At times, an evaluation looks across more than a single year. When this occurs, it helps the reader to have the full program budget, not just an annual value. Additionally, including the statistic of the program budget as a percentage of the whole portfolio budget (when known) lets a reader know quickly if this specific program was a large ‘player’ within the overall energy efficiency portfolio. Often, larger programs receive more rigorous evaluations (although this is not a hard and fast rule.). When a study is not program specific (such as a baseline study or a market study), providing the total energy efficiency budget provides the reader with a good idea of the level of interventions occurring within the market. (This data point is included in the LBNL energy reporting tool.)
* *Program type –* Many different names could be ascribed to a program. It is useful if evaluators adopt a single typology so that all reports use the same labeling. LBNL has worked to create and vet a typology and is currently applying it to track impacts from several thousand reports (LBNL 2016). As such, the authors recommend adopting this typology. (This data point is included in the LBNL energy reporting tool and the NEEP form, although the points represent different typologies.)
* *Program Activities –* All programs have many different activities undertaken to effect change. Inclusion of a bulleted list clearly lays out what occurred within the program intervention.
* *Program Theory of Change –* This does not need to be a long description, or include a logic model. However, inclusion of this information provides a reader with a quick understanding of how the program operates and expected outcomes. To work best, this should include discussion of more than the final outcome of saving energy, as it is in many of the immediate and intermediate outcomes where findings can support what is working and why.

**Information about the Evaluation**

As previously mentioned, there are several evaluation specific areas where additional information could improve the larger body of knowledge for the energy efficiency industry. The bullets below indicate the four evaluation areas to include, the reasons these areas are important, and provide some specific suggestions when reporting.

* *Evaluation Budget Range –* For purposes of a systematic review, the evaluation budget is a proxy for the rigor associated with the effort (although not the only piece of information that helps to understand the evaluation rigor). Therefore, it is important for a reader to understand the evaluation budget. The majority of evaluations cost between 3% and 4% of a program’s budget, although the evaluation budget may be higher for new or small programs because of a desire to delve deeply into a new program or simply because 4% of a very small program is insufficient to capture any primary data. However, market evaluations often do not have associated programs and therefore, the actual evaluation budget provides the best information. Regardless of whether this piece of information is provided as an actual dollar amount or a percent of the program budget, it is not vital to know the exact budget, so the following set of five ranges is one possible solution for providing a budget[[6]](#footnote-6):
  + Under $25,000
  + $25,001 - $75,000
  + $75,001 - $150,000
  + $150,001 - $350,000
  + Over $350,000
* *Evaluation Research Questions –* A reader should be able to quickly find and understand what questions the evaluation plans to answer as this is key to know whether the study should be included in any systematic review. This information can help the reader understand if the data collection and analysis align well with the study questions. For example, if a study question was “How has the program intervention affected the market?” and there is no nonparticipant information collected, then the reader knows that the study most likely will provide partial answers to the question.
* *Evaluation Research Design –* This information is a second proxy for evaluation rigor as the accuracy and validity of the results depend on the research design (which includes elements of data collection, the sample design, and sample sizes). These elements provide an indication of the robustness of the evaluation design. Secondary data collection is just as important and should be included. Typically, the available budget dictates the level of data collection, which is why having both pieces of information (i.e., the budget and data collection) is useful. (NEEP 2014 collects confidence and precision values.)
* *Evaluation Analytical Methods –* This component is important for understanding the analytical factors that frame evaluation results. For example, if two impact evaluations of smart thermostats used identical approaches with the same sample population, the results could be different if they applied different data treatment techniques. Depending on the analytical methods applied in a study, the results can produce different program decisions and conclusions. Therefore, it is imperative to disclose how the inputs from the evaluation design were manipulated to inform a conclusion. (NEEP 2014 collects EM&V methods.)

**Results and Plain Language Summary**

Perhaps the area the authors found lacking in most reports was a summary of easily understood findings. Several executive summaries were 8 to 10 pages long, while others provided the findings with little to no context, causing the meaning of the results to lack clarity. The idea of a “plain language summary” comes directly from systematic review templates, as those who perform these reviews must provide the results within 500 words and without jargon. A communication is in plain language if its wording, structure, and design are so clear that the intended readers can readily find what they need, understand it, and use it.[[7]](#footnote-7)

A plain language summary is not required to enable a systematic review. However, the energy efficiency industry has more stakeholders becoming involved who are unfamiliar with terms created over decades of evaluation. Creating an evaluation report summary that policy makers and other members of the public can understand will expand the use of the findings.

# Template

# The style of energy efficiency evaluation reports varies. The authors intent is not to change these styles, but to include the information presented above as a short appendix with a known format so that all readers can quickly and easily find relevant information. Chosen premises were that this template must:

* Be easy to complete
* Be relatively short
* Include main pieces of information on methods used, data collected, and results found
* Be easy to present electric and/or gas results if there are savings
* Work for process or impact evaluation
* Include other protocols as useful

# Table 2, below, provides the template with sample size examples.

Table 2 Evaluation Study Template Useful to Enable Systematic Reviews

| Data Description | Data | | | | |
| --- | --- | --- | --- | --- | --- |
| Program Information | | | | | | |
| Program Name |  | | | | |
| Program Budget Range |  | | | | |
| Sector | Use LBNL's program typology | | | | |
| Simplified Category | Use LBNL's program typology | | | | |
| Detailed Category | Use LBNL's program typology | | | | |
| Activities | List activities | | | | |
| Theory of Change | Short description | | | | |
| Evaluation Information - | | | | | | |
| Evaluation Name |  | | | | |
| Evaluation Budget Range |  | | | | |
| Type | Impact; process | | | | |
| Research Questions | List the research questions by impact and process; one line for each and link to data collection shown below. | | | | |
| Evaluation Information -Data Collection | | | | | | |
| Primary | List each primary data collection type and number (new line for each) | *n* | *RP a* | *Sample Design* |
| Primary |  |  |  |  |
| Secondary | provide grouped type of secondary information | | | | |
| Evaluation Information – Analytical Methods | | | | | | |
| Quantitative | Engineering model; statistical model and any data treatment; descriptive statistics; IPMVP option; inferential statistics not in a model | | | | |
| Qualitative | Contact summary sheet; coding; pattern coding; memoing; case analysis meeting; interim case summary; vignettes; pre-structured case; sequential analyses; within-case displays: exploring and describing; within-case displays: explaining and predicting; cross-case displays: exploring and describing; cross-case displays: ordering and explaining | | | | |
| Other | TRM algorithms; TRM deemed values | | | | |
| Results | | | | | | |
|  |  | *Electric* | | *Gas* |
| Gross Savings | Provide the table and page numbers in the report that shows the savings | Table X, page Y | | Table W, page Z |
| Net Savings | Provide the table and page numbers in the report that shows the savings | Table X, page Y | | Table W, page Z |
| Findings | Evaluation findings in a short form. Can use technical terms. | | | | |
| Plain Language Summary | | | | | | |
| Plain language text of no more than 500 words that describes evaluation conclusions so that a person not involved with EE understands what the 'take-away' is for the report. | | | | | | |

# a Relative Precision

# Next Steps

# This paper describes a possible template that evaluators could use. The author’s proposals begin small, with a template that fits a short appendix and suggested categories for budgets, program typology, and analytical methods. Each of these components would support the future growth of a robust body of knowledge. The next step in this process is for evaluators to begin to use this or a summary that may not have the exact outline here, but meets the spirit of this paper’s template. This could begin by choice or through the request of evaluation funders. The authors also hope that organizations that are maintaining the energy databases described above (or those about to begin collecting report data) would consider adding in those few variables in this template they do yet plan to include. This type of piggybacking would most likely be of high benefit.

# Using a template such as suggested here is a positive step, especially if an organization could take on the job of reviewing evaluations. However, it will only get our industry part of the way to being able to understand what works, why, and for whom. We need to think critically about how energy efficiency evaluation is currently being practiced and contemplate how to advance its practice for the benefit of the industry (Brown, 2014). As described in Granger (2015), evaluations must answer questions relevant to policymakers and practitioners – questions around which of two similar programs to fund or if adding a new intervention within a program will bring about more savings. Some methods, such as Random Control Trials (RCTs) are “black box” in terms of what interventions are working and whether those interventions are working differently for different type of participants. An RCT study may do an excellent job describing whether a programs is ‘winning’, but often provides only the average savings with results that cannot always be assured with a different population. Through reporting beyond energy savings, and including a theory of change and findings for immediate and intermediate outcomes associated with that theory, a future set of researchers could use current evaluations to undertake systematic reviews and begin to determine what works, why, and for whom.

# References

Billingsley, Megan, Hoffman, Ian M., Stuart, Elizabeth, Schiller, Steven R., Goldman, Charles A., LaCommare, Kristina. 2015. *The Program Administrator Cost of Saved Energy for Utility Customer-Funded Energy Efficiency Programs.* [*https://emp.lbl.gov/sites/all/files/lbnl-6595e\_0.pdf*](https://emp.lbl.gov/sites/all/files/lbnl-6595e_0.pdf)

Brown, Brandy. 2014. *A Metaevaluation of Energy Efficiency Evaluation.* Kalamazoo, MI. Western Michigan University.2013 International Energy Program Evaluation Conference, Chicago

Coryn, CLS, Noakes, L., Westine, C., Shröter, D.C. 2011. *A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009*. American Journal of Evaluation, 32, 199-226.

Granger, Robert C. and Maynard, Rebecca.2015. *Unlocking the Potential of the “What Works” Approach to Policymaking and Practice: Improving Impact Evaluations.* American Journal of Evaluation 2015. Vol. 36(4) 558-569.

Klass, Alexandra B. and Wilson, Elizabeth J. 2015. *Energy Consumption Data: The Key to Improved Energy Efficiency.* Social Sciences Research Network Electronic Paper Collection.

LBNL. *What it Costs to Save Energy.* Accessed in August 2106 at <https://emp.lbl.gov/what-it-costs-save-energy>

Northeast Energy Efficiency Partnerships.2014. *Model EM&V Methods Standardized Reporting Forms for Energy Efficiency. Version 1.0.*

Schiller, Steven R. and Schwartz, Lisa C. 2016. *Coordinating Demand-Side Efficiency Evaluation Measurement and Verification Among Western States: Options for Documenting Energy and Non-energy Impacts for the Power Sector.* [*https://emp.lbl.gov/sites/all/files/lbnl-1005776\_0.pdf*](https://emp.lbl.gov/sites/all/files/lbnl-1005776_0.pdf)

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.

Shadish, W. R., Cook, T. D., & Campbell, D. T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton Mifflin.

Stufflebeam, Daniel L., Coryn, Chris L.S., (2014) *Evaluation Theory, Models, and Applications.* San Francisco, CA. Jossey-Bass

TecMarket Works Team. 2006. *California Energy Efficiency Evaluation Protocols: Technical Methodological, and Reporting Requirements for Evaluation Professionals.* [*http://www.calmac.org/publications/EvaluatorsProtocols\_Final\_AdoptedviaRuling\_06-19-2006.pdf*](http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf)

The Cadmus Group Inc. 2011. *Scoping Study to Evaluate Feasibility of National Databases for EM&V Documents and Measure Savings.* [*https://www4.eere.energy.gov/seeaction/system/files/documents/emvscoping\_databasefeasibility.pdf*](https://www4.eere.energy.gov/seeaction/system/files/documents/emvscoping_databasefeasibility.pdf)

Vine, Edward, Hall, Nick, Keating, Kenneth, M., Kushler, Martin, and Prahl, Ralph. 2012. *Emerging issues in the evaluation of energy-efficiency programs: the US experience.* Energy Efficiency (2012) 5:5-17.

1. For example, in October 2015, California legislators passed a law (SB350,<https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160SB350> ) requiring a doubling of energy efficiency by 2030. Meeting this legislative mandate requires knowing what works and why for the design of future cost effective programs. [↑](#footnote-ref-1)
2. Quote from Campbell Collaboration website. <http://www.campbellcollaboration.org/artman2/uploads/1/Campbell_mal_Campbell_Collaboration_Improving_the_evidence_base_for_social_policy_and_practice.pdf> [↑](#footnote-ref-2)
3. <https://library.mskcc.org/blog/2014/08/systematic-review-vs-literature-review-whats-best-for-your-needs/>) [↑](#footnote-ref-3)
4. This study was completed prior to the debut of the Uniform Methods Project. [↑](#footnote-ref-4)
5. At the time this paper was written, implementing the Clean Power Plan remained in limbo. If this plan begins to be put in place, understanding savings across the nation is very important. [↑](#footnote-ref-5)
6. Ranges based on the authors personal experiences working across the nation and over many evaluations. Each subsequent level often allowed addition of more primary data collection and more analysis, thus increasing the evaluation rigor the assessment was able to undertake. These ranges are, by no means, indicative of actual evaluation rigor. [↑](#footnote-ref-6)
7. Copied from centerforplainlanguage.org. This site has a beneficial checklist for plain language writing. [↑](#footnote-ref-7)