# Apples to Audits: Challenges Affecting the Reliability of Performance Metrics for ARRA-Funded Energy Programs

Jaime Rossman, Alicia LeDuc, Graham Parrington, Research Services, Washington State Department of Commerce, Olympia, WA<sup>1</sup>

## ABSTRACT

The American Recovery and Reinvestment Act of 2009 (ARRA) substantially expanded federal support for state-level energy efficiency, renewable energy, and energy policy development programs. Program implementation strained administrative capacity at both the national and state level, with funding awarded and many projects initiated before all reporting requirements had been finalized. This resulted in a fluid performance measurement process, which struggled to capture data to properly evaluate the programs' energy and economic impacts.

This paper examines the implementation of ARRA-funded energy programs in Washington State. Drawing on the in-house evaluation team's experience with federal and state reporting processes, the authors analyze the resulting performance metrics, and identify challenges affecting the reliability of the metrics and their utility in calculating the impact and cost effectiveness of the state's portfolio of ARRA-funded programs.

Noting the increased demand for a diverse range of evaluation approaches able to meet the information needs of an expanding field of energy program stakeholders, the authors argue that despite the observed challenges, a portfolio-wide, output-enumeration-based reporting structure like that employed during the ARRA period offers unique advantages. Such a structure promises to accommodate a wide range of distinct program designs within a broadly comparable framework, and well-crafted "desk" evaluations of this type have been able to provide a reasonable level of accuracy in estimating impacts relative to the cost of implementation. National impact evaluations of ARRA-funded energy programs that are currently underway can be used to further improve the accuracy and reliability of output-enumeration-based performance measurement.

## Introduction

The American Recovery and Reinvestment Act of 2009 (ARRA) was intended to stabilize an economy in free-fall and, in doing so, to strategically invest in the nation's infrastructure; develop and retain state and local staff capacity; and stimulate progress on some of the new administration's central policy initiatives. By increasing aggregate demand, ARRA funding sought to slow and eventually reverse economic contraction – to prime the pump and trigger economic growth. Once primed, the flow of funds would be more likely to continue, as businesses and individuals that sold products for ARRA-funded projects in turn spent their earnings on other products and services.

<sup>&</sup>lt;sup>1</sup> The views expressed in this paper are those of the authors, members of an in-house evaluation project team responsible for evaluating ARRA-funded energy programs, and do not necessarily reflect those of their employer, the Washington State Department of Commerce, or the Washington State Energy Office. The project was funded in part by funds made available through the American Recovery and Reinvestment Act (ARRA).

Among the act's investments, ARRA substantially increased allocations for federal programs administered by the Office of Weatherization and Intergovernmental Program (OWIP) within the United States Department of Energy (USDOE), targeted at supporting state-level energy efficiency, renewable energy, and energy policy development.

- The budget for the longstanding State Energy Program (SEP) was increased by well over an order of magnitude, from \$33 million in 2008, to \$3.1 billion for the 2009-2012 ARRA period.
- The Energy Efficiency and Conservation Block Grant program (EECBG), established in 2007 but never previously funded, received \$3.2 billion, of which \$2.8 billion was distributed by formula to states, federally-recognized Indian tribes, and larger cities and counties.
- The Weatherization Assistance Program (WAP) provided nearly \$5 billion to scale up existing state programs for home energy efficiency retrofits for low-income households.
- The State Energy-Efficient Appliance Rebate Program (SEEARP) provided \$300 million for states to offer rebates to consumers replacing older appliances with energy-efficiency models.

The Washington State Energy Office (WSEO), a unit within the state Department of Commerce (Commerce), administered the state's allocation of SEP, EECBG, and SEEARP funds.<sup>2</sup> Within the SEP and EECBG allocations, a portfolio of distinct market titles (hereafter referred to as programs) was developed, which targeted various aspects of the state's energy strategy, including building energy efficiency, transportation efficiency, and renewable energy market development.

# **Developing a State-level Portfolio Evaluation Framework**

The WSEO engaged Commerce's Research Services unit (Research Services) to evaluate the portfolio of ARRA-funded energy programs.<sup>3</sup> Project objectives included assisting WSEO staff in documenting and reporting on program performance; developing estimates of the programs' energy and economic impacts using regionally-accepted methodologies where possible; and identifying lessons learned to assist policy makers and program staff in future program design and delivery.

In developing an evaluation protocol to address these objectives within available resources, Research Services drew upon USDOE's evaluation guidelines and various impact evaluation protocols, consulted with experienced evaluators at Washington State University's Energy Extension Program<sup>4</sup>, and reviewed relevant literature including past IEPEC's proceedings.

## Traditional Impact Evaluation, Monitoring and Verification (EM&V)

Traditional energy impact evaluation, which employs various techniques to verify energy savings for a statistically valid sample of projects and then calculates overall savings for a program or portfolio, offers a high degree of confidence and precision. Rigorous, program-specific EM&V is critical to ensuring that evaluation remains methodologically sound and returns accurate information that is useful to program managers in improving service delivery (Hall & McCarthy 2009). However, this approach also entails technically complex estimation and attribution procedures that require significant investment

<sup>&</sup>lt;sup>2</sup> The WSEO managed the state's EECBG award, which funded both state-level activities and sub-grants to smaller cities and counties. Large cities and counties, and federally-recognized tribes, received direct allocations of EECBG funds from USDOE, which were managed at the local level. WAP funding was administered by the housing division of Commerce.

<sup>&</sup>lt;sup>3</sup> SEP and EECBG funds also supported state-level emergency planning, energy policy development and program administration. These activities were not included in the portfolio of evaluated programs.

<sup>&</sup>lt;sup>4</sup> The WSEO subcontracted administration of several of the ARRA-funded market titles to the WSU-EEP.

<sup>2013</sup> International Energy Program Evaluation Conference, Chicago

of resources and professional capacity. Typical EM&V protocols can require between 2 to 10 percent of program funds to implement (Ryan 2011), and work best when built into program design from the very beginning. Once a program has been developed and implemented, needed information can be difficult or exceedingly costly to collect for even a *post hoc* evaluation, and fully rigorous methodologies such as pre- and post-testing, or random assignment of treatment and control groups, are entirely precluded.

Beyond these logistical constraints, the energy evaluation community has also identified concerns about unintended consequences resulting from overreliance on traditional EM&V (Friedmann 2011; Mahone 2011), which raise questions about the ability of such approaches to meet the needs of all the players in an expanding arena of energy programs. For example, attribution generally, and more specifically free-ridership, is a pivotal consideration in typical EM&V frameworks. Significant resources are often dedicated to determining the extent to which an individual program induced desired behavior. But with decades of market interventions by various players, an evaluator's own assumptions may factor more heavily than empirical "proof" of free-ridership levels in a given impact attribution formula. As Mahone notes, this is problematic: "With free-ridership measurements, the devil is in the details.... If conservative measurement rules are adopted, [the] resulting high levels of free-ridership can come into conflict with other policy objectives."

Looking beyond traditional impact evaluation frameworks, Research Services found that IEPEC proceedings documenting past efforts to implement reliable, low-cost, portfolio-level evaluations – in particular, the approach employed by the Oak Ridge National Laboratory (ORNL) in conducting a national evaluation of the SEP program – proved better models for our evaluation project.

#### **ORNL Enumeration Indicator Approach**

As developed and documented by Schweitzer, Tonn and others in a series of papers beginning in the early 2000s, the enumeration indicator approach, in essence, comprised three steps. ORNL's team developed "energy savings coefficients" from the results of recent impact evaluations, and applied those coefficients to state-reported output metrics for a variety of SEP-funded activities, generating cost and emission saving estimates for over 75 percent of the portfolio of programs implemented by the states using SEP funds. This approach allowed the team to develop *outcome* estimates from *output* data, as "[s]tate-provided information on the number of activities undertaken can be multiplied by estimates of the amount of energy saved per activity, and the product will approximate energy savings for that program area" (Schweitzer et al. 2003).



#### **Figure 1: Enumeration Indicator Approach**

When first presented, the approach had employed 80 distinct indicators, and results were constrained by lack of participation from more than half of states (Schweitzer et al. 2003). A second round of evaluation later conducted in which both complications had been addressed resulted in near universal reporting, and a condensed list of 32 indicators. The team emphasized that its approach relied on using past impact evaluation findings conservatively in estimating program delivery rates, and that it returned results with a wider margin of error than a more costly methodology might allow. Still, the evaluators concluded they had found "an economical approach to estimating the outcomes achieved by an extremely broad range of energy efficiency and renewable energy activities undertaken at the state level under the umbrella of a single federal program," and recommended that the approach be deployed further and augmented with data from additional evaluations of the SEP and other programs (Schweitzer & Tonn 2005).

### Washington State's Evaluation Framework

Drawing particularly upon the ORNL team's approach and the work of Goepfrich et al. (2003), Research Services developed a desk-based evaluation framework that was centered on the economic and energy output performance metrics being collected through the various state and federal reporting structures that had been established under ARRA. This allowed our team to augment the WSEO's administrative resources, help implement changes in reporting requirements, and provide ongoing support related to monitoring and process improvement.

Principal reporting structures from which we collected economic- and energy-related output metrics included:

- **1512 reports.** Section 1512 of the act required recipients to report information on ARRA-funded employment and large vendor contracts quarterly throughout their projects. To administer this requirement, Commerce developed a sub-recipient reporting template based on guidelines established by the U.S. Office of Management and Budget (OMB) and the Washington State Office of Financial Management (OFM). Sub-recipients' self-reports were submitted quarterly, compiled by WSEO staff, and reported through state and federal systems.
- **Project invoicing and supporting materials.** WSEO's sub-recipients were required to submit detailed invoices and supporting materials in requesting disbursement of awards. Commerce contract management policies require that expenses, leveraged support, and matching funds be fully documented. State and federal policies regarding prevailing wage on construction projects, preferred purchasing, and cultural, historical and environmental review, resulted in additional layers of documentation.
- **PAGE-reported process metrics.** Quarterly progress reports were submitted through USDOE's *Performance and Accountability for Grants in Energy* (PAGE) system. These reports initially included recipients' estimates of energy impacts such as reduced energy used and greenhouse gas emissions, as well as employment figures. However, in June 2011, reporting of these "impact" metrics was discontinued, for reasons discussed below. Subsequently, metrics reported through the PAGE system consisted principally of the enumeration indicators employed by ORNL in the earlier SEP evaluations, referred to as "process metrics" during the ARRA period.
- **Ongoing narrative reports and project close-out reports.** Sub-recipients provided written reports of activities and progress quarterly, and close-out reports after completion of the ARRA-funded portions of their projects. These narratives, and correspondence between WSEO staff and sub-recipients, frequently provided additional metric data.

Figure 2: Sources of economic and energy output data employed in Washington State's evaluation



Research Services made follow-up calls to sub-recipients where necessary to fill output metric data gaps from the above sources; compiled and validated the data; and employed state, regional and national modeling tools to estimate energy and economic impacts. Estimates were augmented where possible with sub-recipients' self-reported energy impacts (for those developed by competent energy service professionals in accordance with regionally-accepted methodologies), and with results of WSU-EEP's internal evaluations of the ARRA-funded programs it administered. Surveys and structured interviews with WSEO and WSU staff, sub-recipients, and other stakeholders rounded out our research approach, highlighting strengths, weaknesses and opportunities for improvement in program operations. Full results of the evaluation are estimated to be available in late-summer 2013.

# **Challenges Affecting the Reliability of Performance Metrics**

Research Services' evaluation of ARRA-funded energy programs in Washington State identified a series of challenges that affected the reliability of performance metrics collected through federal and state reporting systems, which in turn represent potential sources of error in impact estimates derived from those metrics.

Some of these sources of error have previously been recognized as outgrowths of inherent limitations of any non-EM&V evaluation framework, particularly a framework that relies on output metrics to estimate program impacts. Other challenges we discuss were less well documented in the energy evaluation literature, suggesting that ARRA represents a unique opportunity to field-test and refine enumeration-indicator-based evaluation frameworks. While some of these challenges may be particular to the circumstances faced by program staff during the ARRA period, they may also manifest to lesser degrees in traditional energy programs, as well as future implementation of the SEP program. Several types of challenges are discussed in turn.

### **Inherent Limitations**

Evaluations lacking strong on-site measurement and verification components are subject to inherent limitations. Past IEPEC contributors have articulated concerns with the reliability of impact estimates derived through desk-based portfolio evaluations. In presenting the enumeration-indicator-based evaluation approach for the SEP, for example, Schweitzer et al. (2003) acknowledged that no

energy savings coefficients could be developed for indicators in six of the 20 programmatic areas subject to the evaluation, as sufficient data were not available for some areas, and others were not intended to result in immediate energy impacts. Similarly, Hall and McCarthy (2009) noted that many evaluations reveal "a wide distribution around the savings estimate for individual measures across different programs." Caution is therefore warranted when considering adoption of coefficient-calculated savings estimates absent specific knowledge of how a program was implemented.

In addition to such potential sources of error, and other limitations noted by the ORNL team<sup>5</sup>, Gaffney (2012) reviewed other factors contributing to uncertainty around the earlier desk-based enumeration indicator evaluation of the SEP, and described steps being taken to improve the methodology of the current national SEP evaluation, which adopts a more traditional EM&V approach to evaluate SEP activities during program year 2008 and the ARRA period. This evaluation will provide a greater degree of accuracy and help to verify the impacts of the SEP program, complimenting the performance measurement reporting structure.

#### **Range and Variability of Program Design**

As McCarthy et al. noted, "[a] critical aspect of SEP funding is that it can be used by the states in those ways that best meet each individual state's policies and priorities" (2011). This flexibility is highly valued at the state and local level but also limits the effectiveness of evaluation efforts, particularly where compiling "apples-to-apples" output metrics across various states' programs from which to generate overall impact estimates. For an evaluator, such variability can be maddening.

A principal challenge in measuring the performance of a portfolio of ARRA-funded energy programs, therefore, is simply the incredible diversity of activities that each state had available to choose from in developing plans to expend SEP and EECBG awards. Even after states established high-level guidelines in plans submitted to USDOE, they retained significant flexibility in the types of projects and activities that would be funded.

In the case of Washington, the Legislature provided the WSEO with additional statutory directives to implement or expand specific programs within the overall allocations of ARRA and state dollars. Furthermore, Commerce and the WSEO conducted significant stakeholder outreach and market research to design programs, including hosting a conference where hundreds of people and businesses shared their priorities for the use of ARRA energy funds, as well as meetings, surveys and webinars targeted to individual program constituents.

The combined result of these factors within Washington State was a multiplicity of market titles that spanned much of the authorized range of eligible activities, targeted interventions across almost all of the stages of Rogers' "diffusion of innovation," and involved a diversity of stakeholders, including individuals, households, businesses, industry, governmental entities and utilities. Some programs had tightly limited foci and were geared principally towards effecting transformation in discrete market sectors.<sup>6</sup> Others, including one where an applicant could choose from a list of 14 distinct energy efficiency and conservation activity categories (the 14<sup>th</sup> of which comprised "Any Other Appropriate Activity"), supported a wide variety of project types and/or could be accessed by participants from various points on the innovation curve.<sup>7</sup>

<sup>&</sup>lt;sup>5</sup> The ORNL team identified the following four factors limiting the certainty of energy impact estimates developed through the enumeration indicator approach: "(1) imprecision of the energy-savings multipliers used; (2) incomplete coverage of state activities; (3) lack of attribution of savings; and (4) the exclusion of certain benefits from the analysis." (Schweitzer and Tonn 2005)

<sup>&</sup>lt;sup>6</sup> An illustrative example was WSU's Farm Energy Program, which focused on developing a tool for conducting investmentgrade audits of dairies to expand their ability to access the energy efficiency incentive market

<sup>&</sup>lt;sup>7</sup> Washington's EECBG Grants for Smaller Cities and Counties program, which allowed applicants to select any of the 14

Washington's experience with the range and variability of energy program design serves as a microcosm of that of the nation. The national evaluation of the SEP, currently underway, identified over 1,000 programmatic activities, including 546 during the ARRA period, deployed nationwide (Gaffney 2012).

#### **Methodological Difficulties in Estimating Energy Impacts**

DOE's performance reporting requirements for both the SEP and EECBG programs initially included the following energy-related impact metrics: energy savings, energy cost savings, renewableenergy generation, and emissions reductions. These outcomes, particularly energy savings, were viewed as central performance measures for the programs, but evidence of methodological difficulties in generating reliable estimates appeared early in the ARRA period.

The wide variety of project types, substantial caseloads confronting staff at the state and national levels, and significant pressure to rapidly deploy funding to achieve maximal simulative effect, contributed to an environment where participants struggled to develop *any* programmatic impact estimates. Validating or verifying those estimates was also a challenge, as evidenced in a report from the USDOE Inspector General's office which noted that, "the sum of the states' estimates for anticipated energy savings was 88 billion MBtus based on their initial proposed SEP projects," however this estimate, "was not realistic or achievable since the United States' total energy consumption is estimated at 100 billion MBtus."<sup>8</sup>

As programs were implemented, states were initially required to submit quarterly impact estimates of the energy metrics, and reports of cumulative impacts attributable to each project were also required as individual projects closed out. While states were given wide latitude in how to calculate these metrics, methodological difficulties and limited staff capacity proved significant obstacles to fulfilling this reporting requirement. An April 2011 report from the United States Government Accountability Office (GAO)<sup>9</sup> noted several factors that reduced the reliability of recipients' reported energy impacts, including the variety of calculation methods used to estimate impacts, and the timing relative to project completion in which reporting was required.

**Recipients employed a wide variety of calculation methods. USDOE** provided recipients with guidance on various approaches to calculating these impacts, including standards for recipient-led EM&V-based evaluations and a dedicated impact estimation tool<sup>10</sup> for recipients' use, but it did not *require* specific calculation protocols. This opened several avenues of potential inaccuracy into the impact reporting. For example, while USDOE made substantial improvements to its estimate tool, some recipients continued to use the earlier version. As the GAO report notes, "[w]ithout knowing the methods being used by recipients to estimate energy-related impacts, DOE cannot identify instances where the method along with the associated assumptions being used in calculating estimates may need to be more carefully reviewed." This in turn limited the ability to verify programmatic energy savings, or compare results across recipients.

USDOE-established categories of activities detailed in Program Notice 10-021, "Guidance for Eligibility of Activities," <a href="http://www1.eere.energy.gov/wip/pdfs/eecbg\_10-021\_eigibility\_guidance\_010411.pdf">http://www1.eere.energy.gov/wip/pdfs/eecbg\_10-021\_eigibility\_guidance\_010411.pdf</a>

<sup>&</sup>lt;sup>8</sup> "Status Report: The Department of Energy's State Energy Program Formula Grants Awarded under the American Recovery and Reinvestment Act," OAS-RA-10-17, September 2010. <u>http://energy.gov/sites/prod/files/igprod/documents/OAS-RA-10-17.pdf</u>

<sup>&</sup>lt;sup>9</sup> "Energy Efficiency and Conservation Block Grant Recipients Face Challenges Meeting Legislative and Program Goals and Requirements," USGAO, April 2011. <u>http://www.gao.gov/products/GAO-11-379</u>. Similar findings regarding the SEP program were contained in other GAO reports.

<sup>&</sup>lt;sup>10</sup> The Recovery Act Benefits Calculator, accessible at <u>http://www1.eere.energy.gov/wip/docs/owip\_energy\_calculator.xls</u>

The reporting timeframe limited availability of actual impact data. Impact metrics were reported quarterly, with a cumulative total provided at the close of each project. This timing placed states in a difficult position, as many efficiency and generation projects will not have actual energy outcomes calculated until a year or more after completion. As a result, USDOE officials indicated to the GAO that "instead of collecting actual energy savings data, most recipients report estimates to comply with program reporting requirements." In many cases, these estimates were repetitions of sub-recipients' initial savings projections, though changes in the scopes of projects may have occurred during the intervening time period. Furthermore, as we found in Washington, ARRA funds sometimes supported one phase of a large, multi-year project, meaning that no results exist to be documented until well after the end of the ARRA period.

We note that, ironically, these projects may actually tend to be among the most successful and impactful. For example, a SEP-funded grant helped Washington recruit SGL Automotive, a BMW-affiliate that was developing carbon fiber production lines for use in the i3 line of vehicles. This project created 80 ongoing jobs, and in 2012 the plant announced it was doubling its production line and employment. Materials produced at this plant will result in lighter cars, reducing vehicle fuel consumption. Over time, these impacts will become substantial, perhaps exceeding those of any other single project in Washington's ARRA-funded portfolio. While attribution becomes quite complicated and the SEP program can only lay claim to a small portion of these energy savings, *no energy impacts* can be documented within the ARRA period.

Ultimately, methodological difficulties faced by recipients in calculating energy impacts were deemed too severe to rectify. In June 2011, quarterly impact metric reporting requirements were removed, and USDOE stated its intention to "use grantees' reported process metrics to calculate impact metrics for all grantees using a standard methodology."<sup>11</sup> USDOE explained, "[i]n the past, grantees used their own separate methodology to compute impact metrics. Without any insight into this process, DOE has had difficulty comparing the accuracy of impact metrics across grantees. In the future, DOE will use consistent conversion factors to convert process metric data into energy market impacts."<sup>12</sup>

#### Accuracy and Consistency of Process Metric Reports

DOE noted additional benefits of the reporting requirement change, in that the administrative burden would be reduced, and estimates of program results would be more comparable across grantees. However, as with estimating impacts, recipients had significant discretion in choosing programmatic and project activities and categorizing the resulting process metrics. This complicated using the resulting output measures in a portfolio-based evaluation approach, as different perspectives and assumptions made by various reporters resulted in similar activities being categorized differently.

For example, Washington State allowed smaller cities and counties applying for grants from the dedicated portion of EECBG funding<sup>13</sup> to choose from any of the 14 established categories of eligible activities, rather than limiting sub-recipients to specified project types. As a result, the entire program was categorized for PAGE reporting purposes under Activity 4, Financial Incentive Programs, and process metrics attributable to any of these projects were reported within this category. Data provided by USDOE from the PAGE reporting system shows, along with Washington, seven other states categorized all activities as financial incentives. However, they also show that states made a variety of choices in

<sup>&</sup>lt;sup>11</sup> <u>http://www1.eere.energy.gov/wip/pdfs/sep\_%20reporting\_pre-guidance\_announcement\_06022011.pdf</u>

<sup>&</sup>lt;sup>12</sup> http://www1.eere.energy.gov/wip/pdfs/sep\_eecbg\_reporting\_guidance\_faqs\_06242011.pdf

<sup>&</sup>lt;sup>13</sup> Of the portion of EECBG funding directly received by states, 60 percent was dedicated for sub-grants to smaller cities and counties.

categorizing EECBG program activities. While the modal number of activity categories was one (15 of 50 states), on average states reported activities in three of the 14 categories.<sup>14</sup>

Under the process metric reporting guidelines for financial incentives, recipients were required to report the number and value of incentives provided, by sector. Initially, Washington reported only that high level of information. As projects reached completion, Research Services conducted additional data collection to report process metrics for all applicable categories. For example, a specific project in Washington's small cities and counties grant program that funded audits of retrofits of government building within the sub-recipient city's own jurisdiction reported process metrics such as the number and square footage of buildings audited; auditors' reported estimation of energy savings, and the number and square footage of buildings actually retrofitted; and procurement of specified measures such as HVAC systems. In total, we reported *process metrics* within 11 of 14 categories, though all were reported as outputs of the financial incentive activity category. With each state left to choose its own assumptions as to the categorization and level of detail reported, the variances within each category become significant.

# Conclusion

Methodologically-rigorous EM&V-style evaluations represent the most accurate and complete evaluation protocols available. When portfolio-based, output indicator approaches employing energy saving coefficients are used to estimate energy impacts, they are often criticized as failing to replicate the precision of traditional evaluation methods. These avenues of concern, while valid, do not obviate the need within the energy policy and program community for low-cost, desk-based evaluation protocols that may be "good enough" to meet the needs of a pool of program implementers expanding beyond utility rate-payer incentive programs.

For many state and local entities, the primary objective of evaluation is, as Hall and McCarthy note, "to provide information to policy makers that documents the effects that [programs] are having in the energy efficiency and renewable energy markets in which they operate." Policy makers are arguably more likely to understand program effects in terms of outputs, and so may be more interested in verifying the successful implementation of their programs, than they are in understanding the intricacies of impact calculation methodologies.

Desk-based portfolio evaluations remain a promising option for these entities, so long as they remain rooted in more rigorous approaches. Further research is needed to understand the degree to which desk-based evaluations return less accurate results than traditional EM&V-style evaluations, as well as to quantify the range of cost savings typically associated with these approaches.

The ARRA period provided an opportunity to examine one such desk-based portfolio evaluation system, as USDOE's performance reporting structure evolved to focus on output, rather than outcome, metrics. Challenges observed in the implementation of this reporting system may serve to improve the accuracy and reliability of metrics reported for future programs. Ensuring consistency of output metrics across project types and recipients is key to the viability of metric-based estimations. Leveraging the existing reporting system with the results of the national EM&V protocol-based impact evaluations of ARRA-funded energy programs that are currently underway can be used to further improve the accuracy and reliability of output-enumeration-based impact evaluations.

<sup>&</sup>lt;sup>14</sup> Data provided by USDOE EECBG program lead in personal correspondence with author, 3/29/13.

## Acknowledgements

The authors appreciate the support of their colleagues within Commerce Research Services, as well as the assistance of Washington State Energy Office and Washington State University Energy Extension Program staff. Past IEPEC contributors referenced in this paper provided helpful direction in developing evaluation protocols for this project, which was completed using ARRA funds provided through USDOE.<sup>15</sup>

## References

Donat, Ted. (USDOE EECBG Program Lead). 2013. Personal communication to author, including spreadsheet with PAGE reporting data for EECBG program. March 29.

Friedman, Raphael, Dr. "A Fresh Look at Evaluation to Support Energy Efficiency in the 21st Century," International Energy Program Evaluation Conference, Boston, MA, 2011

Frischmann, Michael; and Kroll, Ryan. "On-site Measurement and Verification versus Project File Desk Review," International Energy Program Evaluation Conference, Rome, Italy, 2012

Gaffney, Kathleen. "National Evaluation of the US Department of Energy State Energy Program," International Energy Program Evaluation Conference, Rome, Italy, 2012

Goepfrich, Valy; Dyson, Christopher; Goldberg, Miriam; Erickson, Jeff, PA.; and Bloch, Oscar. "Getting the Best Evaluation for the Buck," International Energy Program Evaluation Conference, Seattle, WA., 2003

Hall, Nick; De Cotis, Paul; Kushler, Marty; Megdal, Lori; and Vine, Edward. An Evaluation Approach for Assessing Program Performance from the State Energy Program, TecMarket Works, October 2007 (http://www.tecmarket.net/documents/Final%20SEP%20Evaluation%20White%20Paper%2010-18.pdf)

Hall, Nick; and McCarthy, Patrick Ph.D., J.D. "Portfolio Evaluation Versus Program Evaluation: Is There A Balance?", International Energy Program Evaluation Conference, Portland, OR., 2009

Mahone, Douglas. "Free-Ridership as a Way to Kill Programs - How Evaluation Policies Can Frustrate Efficiency Goals," International Energy Program Evaluation Conference, Boston, MA, 2011

2013 International Energy Program Evaluation Conference, Chicago

<sup>&</sup>lt;sup>15</sup> The following disclaimer is required for reports prepared in part with ARRA funds: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

McCarthy, Patrick Ph.D., J.D.; Hall, Nick; Schweitzer, Martin; and Lambert, Faith. "Building Long-Term Sustained Capacity and Influencing Decisions: The U.S. DOE SEP Capacity Model," accessible at (<u>http://www.naseo.org/programs/sep/documents/SEP Building EE and RE Capacity in the States.pdf</u>)

Rogers, Everett M. 2003. Diffusion of Innovations, Fifth Edition. New York: Free Press.

Ryan, Barb; and Ignelzi, Patrice. "Doing More with Less: Getting What's Needed Most from Evaluations," International Energy Program Evaluation Conference, Boston, MA, 2011

Saxonis, Bill. "A New Evaluation Paradigm for a New Generation of Energy Efficiency Programs," International Energy Program Evaluation Conference, Portland, OR., 2009

Schweitzer, Martin; Berry, Linda; Jones, Donald; and Tonn, Bruce. (2002) "Quantifying the Outputs of the State Energy Program: the Enumeration Indicators Approach," *Energy Studies Review:* Vol. 10: Iss.2, Article 8. (http://digitalcommons.mcmaster.ca/esr/vol10/iss2/8)

Schweitzer, Martin; Jones, Donald; Berry, Linda; and Tonn, Bruce. 2003. "Estimating Energy and Cost Savings and Emissions Reductions for the State Energy Program Based on Enumeration Indicators." ORNL/CON-487. Oak Ridge, Tenn: Oak Ridge National Laboratory.

Schweitzer, Martin; Tonn, Bruce; Berry, Linda; and Jones, Donald. "Using Performance Metrics to Quantify Energy Savings and Emissions Reductions for the U.S. Department of Energy's State Energy Program," International Energy Program Evaluation Conference, Seattle, WA., 2003

Schweitzer, Martin; and Tonn, Bruce. "Painting a Big Picture: Using Activity Counts and Savings Multipliers To Estimate State Energy Program Accomplishments," International Energy Program Evaluation Conference, Brooklyn, NY., 2005

Schweitzer, Martin; and Tonn, Bruce. 2005. "An Evaluation Of State Energy Program Accomplishments: 2002 Program Year." ORNL/CON-492. Oak Ridge, Tenn: Oak Ridge National Laboratory.

Vreuls, Harry. "Comparable energy savings: how to ensure that singers form a harmonious chorus?" International Energy Program Evaluation Conference, Rome, Italy, 2012