

# Portfolio Evaluation Versus Program Evaluation: Is There A Balance?

*Nick Hall, and Patrick McCarthy, Ph.D., J.D.  
TecMarket Works, Oregon, WI*

## ABSTRACT

This paper examines two different frameworks for establishing the energy impact evaluation function. These are the portfolio evaluation framework and the program evaluation framework. The paper discusses impact evaluations and does not address other types of evaluations, such as process evaluations, market effects studies or other studies. These other types of evaluations are important and provide information that program and portfolio managers need to design and implement effective programs. However, the focus of this paper is limited to the impact evaluation framework and the choice between a portfolio framework and a program framework. The paper focuses on the need for policy makers and evaluation professionals to establish their energy impact evaluation functions within a framework that plans for a wide range of information needs, all of which are dependent on well planned, managed, timed and budgeted program-level evaluations. The paper demonstrates that the benefits of a portfolio framework can be achieved within a program framework if that framework is well designed and implemented. The paper also demonstrates that a portfolio framework cannot meet the needs of a number of information consumers.

## Introduction

An evaluation's ability to document critical program-level achievements and benchmark key program metrics necessary for planning and implementing effective programs and portfolios should not be underestimated. The evaluation framework, in addition to an accomplishment documentation function, is a critical part of the portfolio's program operations and management function. Well-funded evaluations that are also well-timed, focused and managed can substantially improve program performance and the performance of the portfolio. At the same time, these studies can reliably document savings, help to update deemed or stipulated database metrics, help to better understand market potentials, and inform public policy. The evaluation effort has to provide both policy makers as well as program managers with the information they need in time to be effectively used.

Energy impact evaluation frameworks that focus on only one or two key objectives (such as the need to calculate incentive payments) at the exclusion of other objectives (such as quantifying program specific accomplishments) harm not only the evaluation efforts, but damage the efficiency and effectiveness of the portfolio. The evaluation framework essentially becomes a roadblock to maintaining effective programs. In a nation moving toward aggressive national energy and carbon reduction objectives, program planners and policy makers must be able to strategically position programs within a portfolio in a way that captures significant energy efficiency potentials and reduces carbon emissions. Evaluation efforts that are not structured to be effective within these new environments can reduce the energy and environmental achievements from the portfolio, decrease the ability to design and manage effective programs, increase the potential for fraud, waste and abuse, and decrease our country's ability to meet our energy and environmental objectives.

Fortunately, well planned and implemented evaluations that serve everyone's needs are possible if the evaluation efforts are effectively managed and structured within a framework designed to serve all key stakeholder needs. This paper discusses key issues pertaining to setting up evaluation frameworks

and evaluation systems that enable the impact evaluation effort to be effective at meeting all critical information needs.

## Key Concepts

As many experienced evaluation professionals know, structuring a statewide energy impact evaluation effort for an energy efficiency portfolio is a complex task. It is a task that requires a great deal of evaluation expertise, as well as an understanding of the types of evaluation-supplied information necessary for program design and redesign efforts. Recent initiatives have tried to move evaluation from an effective program and portfolio oversight and management tool to a regulatory tool that focuses only on calculating an incentive payment or updating deemed measure databases.<sup>1</sup> However, this not only harms the energy efficiency portfolio, but also forces program designers and portfolio management to operate under greater risk, without a clear path to increased efficiency or effectiveness.

This caution is not intended to suggest that focusing on updating deemed values is unimportant. In fact, that should be a focus of every impact evaluation. It should be done in a way that the evaluation helps to establish not only more reliable deemed values, but expands the deemed value measure categories to include a much wider range of program influenced, market and delivery approach specific values. In addition, the use of deemed data to support the program planning process is an expedient approach for estimating ex-ante savings when more specific estimates are unavailable or cannot be developed. They also provide an easy way for policy makers to estimate the range of what a program or portfolio might achieve in the absence of more detailed estimates.

However, limiting the evaluation function to effectively achieve only part of the evaluation-driven information needs (i.e., deemed value updates) harms the portfolio as a whole and the individual programs operating within that portfolio. Without effective and accurate program impact evaluation, bad programs that appear to be good on paper can remain ineffective for years, wasting large amounts of resources that could have gone to effective programs. Moreover, good programs that are out-performing their energy or carbon targets do not get sufficiently recognized as such; instead, they become viewed as typical programs, not destined for the broader utilization that they merit. Likewise, key conditions that impact effectiveness and that need to be communicated remain hidden from the state's program planners and approvers, eroding the effectiveness of the portfolio.

The trend of having energy impact evaluations serve only ex-ante deemed savings update efforts or support incentive calculations detracts from the effort of improving the programs and the portfolio. As a result, the narrowly defined evaluation framework becomes responsible for helping to lower the portfolio's performance, acquiring more expensive energy resources, and speeding carbon accumulation in the atmosphere.

Before a decision is made regarding how an evaluation framework should be structured, it is important to carefully consider a wide range of potential consequences of that decision. However, before addressing the primary focus of this paper, it is important for the reader, especially those new to the evaluation field, to understand some basic key evaluation-related concepts. We briefly review a few of these concepts to provide a foundation for this paper.

---

<sup>1</sup> Deemed databases are libraries of estimated measure-specific projections of energy impacts for measures installed in a specific operational environment. They are used by program planners and policymakers to estimate program and portfolio energy impacts.

## Portfolio, Program and Measure Level Evaluations Are Different

It is important to understand the differences between the different types of evaluations because there are different studies that serve different needs.

**Portfolio-level evaluation.** Portfolio-level evaluations are structured to provide reliable results at the portfolio level. These studies sample and assess savings at the portfolio level. This often reduces the evaluation budget and scope by treating the portfolio as if it were a single program. The portfolio evaluation typically focuses on high portfolio-level impact measures that dominate the portfolio's savings. The benefit of this effort is that the study can give an overall reliable assessment of the portfolio as a single energy efficiency effort. Nonetheless, this approach has significant limitations because the portfolio evaluation sampling and analysis frame are focused on the portfolio, instead of the program where those conditions play a role in capturing savings. Consequently, significant detriments of this approach include reducing the reliability of program-level findings, and reducing the reliability of program-specific measure-level or condition-induced savings.

**Program-level evaluation.** Program-level evaluations focus on the program rather than the portfolio. These studies sample and assess savings at the program level, and thus provide reliable savings estimates for programs at the expense of reliable program-specific measure-level savings estimates. Program-level evaluations do well for comparing one program with another within a portfolio. They also do well for assessing portfolios, because reliable program evaluations are rolled up to provide a more reliable analysis of the portfolio. The evaluation metaphor “*Reliability only runs uphill*” is based on this notion. That is, with program-focused evaluation frameworks, the portfolio assessment is more reliable (data are being rolled up). In contrast, with portfolio evaluations, the program assessment is less reliable (data are being rolled-down and split apart, decreasing reliability at the program level).

**Measure-level evaluation.** Measure-level evaluations focus on samples at the measure level, providing reliable savings estimates for the measures installed within a program. The goal of these studies is to obtain reliable findings about how a specific measure provides savings (e.g., overhead lighting using T-5s with high efficiency electronic ballasts). These studies can also zoom-in, using a higher level of magnification, and look at measures within specific operational and delivery environments (e.g., small commercial overhead lighting using T-5s and high efficiency electronic ballasts offered via direct install program delivery approaches in an urban area). Regardless of the level of magnification, these studies go beyond program-level evaluations by focusing on individual measures. Because measure savings are significantly influenced by individual program operations, these studies help to determine the typical level of savings that are obtained from a technology or the typical level of savings from a technology under specific customer and program-affiliated use conditions<sup>2</sup>. These studies are program design support studies, or studies that are used to update deemed databases. They focus on energy, demand or load shapes.

**Top savers evaluation.** Top savers evaluations focus only on the top savings measures or programs within a portfolio and use deemed savings estimates for small programs and smaller impact measures. These are conducted one of two ways. In a top savers portfolio evaluation, the evaluation

---

<sup>2</sup> The term “use conditions” refer to any aspect associated with the way a participant uses a technology that impacts the achieved or achievable savings.

effort segregates the measures within the portfolio into top savers and non-top savers. The evaluation efforts are then planned at the portfolio level for the top savers. In this study the sampling and analysis is at the portfolio level. In the top savers program evaluation, the same segregation occurs, except that the groupings are for top savers within the programs. Combining these two approaches is also possible: focusing on the top saving measures within the portfolio and also including the top saving measures within programs; using a double sampling and analysis approach (portfolio and programs) that in most cases defaults to the program sample as the highest level of disaggregation, then restructuring the sample so that it is representative of the portfolio.

**Metering and verification (M&V).** M&V is a part of most evaluations. Portfolio M&V focuses the M&V efforts on the measures of greatest risk of error in the estimate at the portfolio level. Samples are drawn from the portfolio using random selection or stratified random selection to represent the portfolio. Program M&V focuses the M&V effort on program-level measure installation and use conditions. The samples are structured at the program level to accurately reflect the M&V needed to provide reliable program savings estimates. In these cases, the M&V sampling is program-stratified, so that analysis is conducted on measures for a specific program. Measure M&V focuses on sampling of specific measures representing a set of specified operational conditions that need to be better understood, such as filling a gap in a deemed database or confirming savings estimates for non-deemed measures.

All of these approaches work well for specific purposes. Portfolio M&V most likely will also work for measure-level sampling when the analysis is focused on the same objective, typical measure applications. However, when the sample needs to become more focused than what is typical for the portfolio, the M&V sampling approach needs to be stratified for certain conditions, such as weather, application, etc. However, an approach designed for a portfolio assessment, or for improving a deemed database, may not be able to provide reliable program assessments. For example, there may be over-sampling for one program-related condition and under-sampling for other conditions that are not typically considered in the portfolio sample. Thus, sampling approaches targeted for portfolio-level analysis or measure-level analyses and that are *not* structured for program-level considerations will not be adequate for program-level evaluations. Programs are different and the use conditions for a program are not typical of the portfolio in most cases. Applications that use portfolio M&V to assess programs will not deliver accurate program-level evaluation results unless there are enough sample points at the program level to reliably conduct the analysis.

## Why Does All This Stuff Matter?

### Sampling

Within evaluation discussion circles, some individuals have postulated that if you do a reliable portfolio-level sampling of the top saving measures, then you can obtain a reliable portfolio impact assessment. This assumption is probably true for the portfolio evaluation when either: (a) the portfolio is viewed as a single program operating within an environment in which program-level variances that impact the assessment effort are not large enough to impact the overall accuracy of the findings, or (b) when the sample of measures at the portfolio level captures the typical installation and use conditions for the portfolio. However, to jump from a portfolio sample to estimating the impact of programs within that portfolio would require very large program-structured sample sizes to support the detailed program level analysis that is needed. Nonetheless, portfolio-focused approaches can be accurate and can be used to reliably assess programs, if the study is well managed by expert and experienced evaluation professionals. Specifically, this requires evaluation professionals who understand the program-specific,

climate-specific, delivery-specific, behavior/use-specific, market-sector-specific characteristics that need to drive the program level impact sampling and analysis efforts. In other words, the study has to use program-focused sampling strategies, essentially converting the study from a portfolio study to a program study.

### **Program Planners and Managers' Needs and Decision-Making**

If all that is needed from an evaluation is a reliable picture of the portfolio, then the portfolio-level approach works well. However, for program managers and evaluators that need reliable program-level impact data to build and maintain effective programs, then the portfolio-level evaluation is not satisfactory. When program administrators use portfolio-level evaluations to plan programs, the portfolio becomes less effective over time because the reliable program-specific information needed to maintain high performance is not provided. As a result, program administrators can inadvertently end up not growing good programs and maintaining bad ones. Furthermore, portfolio-level evaluations force program managers who need impact information for monitoring and improving programs to conduct their own impact evaluations, potentially doubling impact evaluation costs.

Program planning decisions are not made at the portfolio level; they are made at the program level. Most regulation and oversight agencies and program administrators want to know if their programs are effective and which programs need to be improved. For this reason, most states conduct evaluation planning at the program level. The program is the primary building block of the portfolio, and most oversight agencies and program administrators know that if they focus on the programs, they can build and maintain more effective programs and portfolios.

### **Cost Effectiveness and Deemed Databases**

Program cost effectiveness represents a program-based statistical metric that is built from measure level savings consistent with the installation and use conditions impacted by the program. Program cost effectiveness is typically based on program installed measures that are rolled up into a program assessment. To capture portfolio cost effectiveness, the program-specific test results are rolled up to a portfolio cost effectiveness assessment, reflecting the summation of the actual program-level efforts and conditions. If there are no reliable program evaluations, then a program's cost effectiveness test (needed to decide which programs to maintain or change) cannot be reliably determined.

In the last 20 years, evaluation professionals have learned that there is a wide distribution around the savings estimate for individual measures across different programs. These studies have allowed deemed database managers to understand that markets, market sectors, outreach and enrollment practices and other program service-related characteristics impact savings. Portfolio evaluations that are not structured to consider these conditions for conducting their sampling and analysis systems will not be able to provide reliable estimates of program impacts. A few of the program-specific conditions that impact savings are presented below:

- Programs targeting first-time technology users can save more than a program targeting repeat users;
- Programs that screen enrollment for free riders can save more than ones that do not;
- Programs targeting the early majority<sup>3</sup> can save more than programs targeting innovators;
- Direct install programs operated by well-trained professionals can outperform rebate programs covering the same measures;

---

<sup>3</sup> The terms "early majority" and "innovators" are associated with market adoption characteristics and are used to classify when and why technologies are adopted. For more information, see Rogers 1995.

- Programs targeted at motivated participants can outperform the same program that uses a different marketing and outreach strategy;
- Door-to-door screening programs can outperform open enrollment programs;
- Programs targeted to middle-income neighborhoods can have different savings from the same program targeting upper or lower income areas;
- Give-away programs can save less per measure than inspection and rebate programs;
- Programs that provide diffusion support<sup>4</sup> information can outperform programs that do not.

As a result, studies have been conducted to allow deemed savings estimates to incorporate adjustments that are reflective of specific programs, serving specific markets, via specific approaches. These changes have allowed well-managed deemed databases to become more accurate over time. As deemed databases continue to evolve, incorporating lessons learned from measure-level and program-level evaluations, they will continue to become more accurate when applied to specific types of programs. This fine-tuning of the evaluations to support deemed database improvements requires program-level evaluation planning. Deemed savings estimates are program environment and measure use metrics, not portfolio measure metrics. Many program managers already know this when they tailor their projected energy savings to reflect program-specific conditions. For this reason, program-focused measure saving estimates done well will typically be more accurate than deemed values built for a typical portfolio installed measure.

### **Measures Do Not Save Energy**

Energy-efficient measures do not save energy or reduce carbon emissions. Only when a measure is installed and used is there a “*potential*” for the measure to save energy. Energy savings are then achieved only if that measure is used in a way that provides the expected energy savings. Thus, energy savings are as much a function of the behavior and use environment of the participant as it is a function of the measures being installed. The participant and their associated operational environment is a function of not only the program’s geographical/climate area but also the program’s operational systems, particularly the program’s targeting, outreach, marketing and enrollment systems. To accurately evaluate energy impacts, it is important to assess those impacts around the systems that allow those savings to occur. This is best accomplished at the program level.

### **Small Programs Are Important**

One of the arguments used to support portfolio-level evaluation is that it can ignore small programs that do not contribute significant energy savings to the portfolio. The rationale is that the savings for these programs can be based on deemed savings without harming the overall accuracy of the portfolio estimate. Unfortunately, this approach will give the program managers inaccurate information for developing their future program plans or for adjusting their current operations to be more cost effective. For example, some of these small programs produce no savings or provide savings well below their deemed values. On the other hand, some of these small programs out-perform their deemed savings targets and obtain higher levels of savings for the measure mix used. If the portfolio-level evaluation framework only waits until the programs are large before they are evaluated, this framework essentially supports wasting energy efficiency dollars on ineffective programs that are not evaluated. On the other

---

<sup>4</sup> Diffusion support refers to any material or information provided to participants that allow them to network the benefits of the program or action within their social or business circles in a way that acts to spread the behavior in the market (see Rogers, 1995).

hand, if the program is a high performing program, it will not be recognized as a best practice and emulated in the market, thereby lowering the savings that could have been achieved by growing the small program.

The cost to conduct impact evaluations of small programs ranges from a low of about 3 percent of the program budget to a high of about 7 percent. If a program is allowed to stay in the market one year and it achieves 75 percent of the projected savings, in most cases it would have wasted more money than the evaluation would have cost. If the program is more effective than projected and not allowed to grow because the portfolio evaluation applied typical values for the program's savings, the program will not be seen as a champion, and the designers and managers of that program will receive information that does not reflect their program's contributions. The program remains small, and the customers paying the price of the portfolio are harmed by over-paying for energy efficiency.

### **Poor and Great Managers Look the Same**

If program managers are told that their programs will not be evaluated or will be evaluated by applying typical saving values from the portfolio evaluation or by applying deemed values, there is little incentive to maintain a high performing program. Program managers will quickly see that no matter how well they perform, the evaluation will not explicitly examine their program except via an average or typical portfolio installation approach. Essentially, it will not matter if they do a good job or a poor job managing their program. In addition, they will not receive recognition for their work (good or bad). As a result, high performing managers who are made to look average in the portfolio evaluation will look for employment that recognizes their contributions and excellence. Program and portfolio effectiveness will further erode until there is a need to improve the effectiveness of the portfolio by identifying low and high performing programs.

### **Program Evaluation Provides the Pressure for Sustained Performance**

Policy makers need to understand that program evaluation places pressure on managers to maintain a high level of performance. At the end of the day, it is the program energy impact evaluation that top managers and designers rely upon to give them the information that they need to capture energy savings effectively. Program evaluation builds and maintains good programs and portfolios.

### **Without Thorough Evaluation, Regulatory Oversight Is Harmed**

Evaluation is one of the most important regulatory and oversight functions. When the regulatory oversight function is weakened by an evaluation framework that is focused on the portfolio, it may not detect a problem that would show up in the program evaluation. In the last couple of years, we have seen cases that suggest that evaluation as an oversight function is needed, even on small programs. For example, we have seen: (1) air conditioning tune-up programs that have resulted in increased loads on the units after they have been "tuned-up" while other tune-up programs reduced consumption; (2) multi-family programs that installed measures in areas that were not occupied, while other programs installed measures in ways that obtained high savings; (3) program measures left in participants' facilities that were never installed, while other programs were able to achieve the installations; and (4) program funds spent on measures that provided no savings from the majority of the installation's use conditions, while others achieved their savings. Some may argue that program evaluation as an oversight function is not needed on every program. However, what program is so small that oversight efforts on impact achievements are not needed?

## **Portfolio Evaluation Encourages Waste, Fraud and Abuse**

Typically, energy efficiency funding is tax-based, public funds-based, rate-based, or recovery-based. When funding is provided because of a state or federal policy decision, the public has a right to assurances that the programs are prudently using public funds to support programs. The U.S. Department of Energy (USDOE) and many states have indicated that assessing fraud, waste and abuse is important. If impact evaluations are only done at the portfolio level, there is less chance that program-specific waste, fraud and abuse will be uncovered in the evaluation effort. The evaluations are not just documentation efforts, but they are also a part of the management and oversight responsibility platform. Public policy should address how well their oversight and accountability responsibility is served via a portfolio evaluation approach. If an evaluation framework is established that requires added audits and financial examinations to limit waste, fraud and abuse, what savings are gained?

## **Portfolio Evaluation May Slow the Use and Application of Information**

One frequently heard complaint from the consumers of evaluation information is that the evaluation was delayed to the point where the information could not be effectively used to manage the program or the portfolio. Portfolio evaluations are large, cumbersome efforts that have to address the entire portfolio. A portfolio evaluation is not complete until all the coordinated evaluation research efforts across the portfolio have been completed, compiled, analyzed and reported. This acts to slow the information pipeline to not only program managers who need the information, but also to policy makers who must review and approve future programs and portfolios. Program evaluations are smaller and more focused studies. Not only can they be staggered to reflect information needs, they are smaller efforts that can be completed more rapidly. Evaluation needs to be responsive to the needs of these stakeholders. Accordingly, the evaluation framework should be housed within an organizational and operational environment that is capable of handling the process in a timely fashion. Program and portfolio managers need information in real time. They need to know if their programs are implemented effectively in order to capture the energy savings. If they do not have the energy savings information at the relevant decision-making points, then it is difficult to know if program adjustments are needed or if the results of previous programs adjustments have been effective.

## **Not All Programs Need the Same Level of Evaluation**

While program evaluation is important, not all programs need to be evaluated at a high level of reliability, but all programs should, to some degree, be evaluated at least to the extent that portfolio managers and program designers can use the data to improve programs and to assure a reasonable level of oversight and impact reporting accuracy. Thus, program evaluation planning efforts need to identify the program evaluation needs. They must also determine the rigor and methodology of approaches that meet the wide range of information needs that the evaluation effort must support. The list of considerations that must be incorporated into the evaluation planning effort is long. Detailed explanations of these considerations would require a separate paper to address them and, therefore, are beyond the limits of this paper. However, these considerations include the following:

- Size of the program in dollars
- Size of the program's projected energy impacts
- Newness of the program or the delivery approaches
- Newness of the measures covered
- Standard operations of the market and market sector being served
- Measures included in the program



- Marketing, enrollment and delivery approaches for the program and the program's measures
- Participants' energy use conditions and factors affecting the variation in energy use (weather, operations, timing, etc.)
- Estimated risk factors associated with the measure and the program-specific use conditions
- Stability of the savings expected
- Rigor of past program evaluations and confidence in the results
- Changes to the market, the program or the measures
- Baseline conditions associated with the measures and their use conditions
- Market operations changes that impact measure selection and use decisions and options
- Expected accuracy and comprehensiveness of program-influenced load shape distributions
- Expected accuracy and comprehensiveness of program-influenced deemed savings metrics
- Cost of the various evaluation approaches
- Needs for special studies to support policy decisions
- Needs for special studies to support program or portfolio construction
- Past program management performance or impact-related program operational concerns
- Familiarity with the program administrator and service contactors
- Experience of the service provider
- Potential for waste, fraud or abuse

## Do It All by Planning Well

The above parts of this paper provide a lot of information to consider. In some cases, the very purpose of the evaluation effort is discussed, so that policy makers can determine the context of their considerations and the implications of their decisions. However, this paper needs to end by pointing the way forward rather than looking backward. Effective frameworks need to be constructed, and defective frameworks improved or removed. In planning the evaluation framework, it is important to consider the lessons that this field has learned over the last 30 years, and particularly the notion that measures do not produce savings. Understanding this single concept can help guide the selection of the appropriate evaluation framework. For the experienced evaluation professional, once this simple underlying concept is well understood, the rest of the pieces fall into place. As noted earlier, savings from a measure are based on use conditions, and use conditions are significantly impacted by the type and operations of the program. To establish a good evaluation framework, we recommend that the following steps be incorporated into the evaluation process:

1. **Identify Use Conditions.** Identify the specific use conditions that are expected to affect the ability of the program to achieve savings. This should include assessments of the program's operational approaches, an understanding of customers that the program is trying to reach, an understanding of the use conditions within the participant's environment that can impact savings, knowledge of the technologies and services offered by the program, and an understanding of the program's marketing, enrollment and outreach approaches that will influence achieved savings<sup>5</sup>.

---

<sup>5</sup> Experienced evaluation professionals understand these concepts and how program conditions can impact energy impacts, however, these concepts are difficult to for evaluation professionals just entering the field to grasp. This information can be found via reviews of past process and impact evaluations and interviews with program managers and designers, and they are embedded within program theories and logic models, and within the market operations and market effects literature.

2. **Identify Deemed Gaps.** Identify gaps (energy, demand and load shape) that need to be filled, approved or strengthened, paying particular attention to not only the types of measures, but also the program and market conditions that impact the savings for those measures. Structure the deemed database so that it can be populated with program and market-associated influence metrics, so that when they are collected, they can be put to use. (For example, when the evaluation documents the savings difference between a T-5 lighting fixture installed in a commercial facility via a rebate program, and one installed via a direct install program or a program that audits for best opportunities, both of which would provide substantially different savings for the that measure.) Identify the M&V requirements needed to support deemed database needs that can be embedded within the program evaluations without overly harming the ability of the evaluation to assess program impacts.
3. **Identify Uncertainty and Risk.** Identify areas of uncertainty and risk associated with each of the key program, market, operations, implementation and technology use factors that can influence the accuracy of the savings estimates. This will help identify areas of greatest need for assigning resources and rigor levels. This is especially important for program evaluations because the uncertainty and risk can be different for different types of programs.
4. **Identify Reliability Needs.** Identify the reliability needs for each program in the portfolio and the measures within the portfolio that are in need of a more intensive focus. In conducting this assessment, understand the information needs of all key consumers and commit to an evaluation framework that places value on the information needs of these consumers, but particularly the needs of the people who must keep the programs and the portfolio cost effective. This assessment will identify which programs and measure-performance aspects need rigorous evaluations and which need lower rigor studies.
5. **Plan and Budget Carefully.** Plan evaluation budgets so that the needs of all consumers are maintained throughout the study period. Plan the program-specific analysis needs as well as the data collection efforts that need to be accomplished to perform all analyses. Allocate the evaluation resources so that the evaluations accomplish the greatest fulfillment of the information needs. This may mean delaying some studies, so that the studies done early can be reliable enough to meet the consumers' needs.
6. **Engage Program Managers and Other Stakeholders.** Engage program managers and other key stakeholders early and often. Plan for the tracking and reporting of early feedback information that managers need to both maintain high performance and decide on program corrections and course changes.
7. **Conduct Evaluation Studies Well.** Conduct the evaluation studies and manage them well. Maintain a focus on the important information needs, and avoid distractions that may reduce the effectiveness of the evaluation.
8. **Report Conclusions.** Report conclusions of the analysis so that the key consumers have the information that they need in time to effectively use them. This may mean multiple deliverables from the evaluation effort for multiple consumers. Roll up the program savings to provide accurate portfolio-level savings.

9. **Assess Use of Results.** Conduct ongoing assessments of how well evaluation findings are being incorporated into programs to enhance operations and cost effectiveness.
10. **Assess Success of the Evaluation Framework.** Do not always assume that the evaluation function is effective. Conduct on-going assessments of how well the evaluation effort is being conducted and its results communicated. Additionally, explore ways to improve the evaluation process with the key information consumers.

## Conclusions

While portfolio evaluation can be valuable for particular purposes (i.e., average deemed value updates, portfolio-wide typical impacts), broader reliance on this approach can harm the portfolio as a whole and the individual programs operating within that portfolio. In contrast, this paper shows how the benefits of a portfolio framework can be achieved within a program evaluation framework, and without such harm. Moreover, the benefits of a program evaluation framework include program-focused savings estimates that are normally more accurate than deemed values built for a typical portfolio installed measure. Additionally, a bottom-line benefit of the program evaluation framework is that most program planning decisions are made at the program level. These are among several fundamental advantages that have led most states to conduct evaluation planning at the program level.

Depending solely on a portfolio-based evaluation approach raises several significant concerns. Moving to evaluation frameworks that focus on the portfolio as a single energy efficiency entity not only impedes the ability of the portfolio to acquire savings, it also reduces the ability of the portfolio to be well-designed, well-managed and cost effective. Such a framework lowers the probability of identifying and removing poor performing programs, in large part because a portfolio framework that does not reliably assess programs instead rewards poor management at the expense of shining a light on excellence. This also delays other corrective actions necessary to keep the portfolio cost effective, and can end up encouraging waste, fraud and abuse. In the long run, an overly exclusive reliance on the portfolio framework will reduce the desire to innovate and build better programs, because the information needed to do so will not be collected. A counterproductive effect is that participating customers will not receive the most effective energy services needed to reach their own energy resource and climate change objectives and will get less for their energy portfolio dollar.

In closing, the evaluation framework selected has a significant influence on the ability of the portfolio to meet its energy savings objectives, especially for future portfolios that need to be built with well-managed and cost effective programs. The foundation for a strong portfolio is good, effective programs. Effective programs are built from reliable program evaluations. Ultimately, program evaluation frameworks provide essential advantages and alleviate critical risks relative to more exclusively portfolio-based alternatives.

## Reference

Rogers, Everett. 1995. *Diffusion of Innovation*. New York, NY: The Free Press.