Measuring Success in Midstream Programs: Design and Evaluation Recommendations from a Television Program

Marti Frank, Research Into Action, Inc., Portland, OR
Ty Stober, Northwest Energy Efficiency Alliance, Portland, OR
Jane Peters, Joe Van Clock, and Alex Dunn, Research Into Action, Inc., Portland, OR
Nicole DeHoratius, University of Chicago, Chicago, IL

ABSTRACT

Consumer electronics are an important but challenging target for U.S. energy efficiency programs. To date, most utility programs focus on televisions, and most are "midstream" programs that pay financial incentives to retailers, rather than end users, in an effort to effectively utilize limited funds. Although funders and implementers have lauded the success of these programs in increasing adoption of efficient televisions, evaluations of their impact are inconclusive. Previous evaluations found that programs struggled to quantify influence on the complex and fast-changing television market. Consequently, regulatory bodies direct major funders to withdraw their financial support putting the future of these programs, and the midstream program design, at risk.

The Northwest Energy Efficiency Alliance's (NEEA) Consumer Electronics Initiative, operated in coordination with several California electric utilities, had a preponderance of anecdotal success and yet quantifying its impact has proven extremely difficult. The most recent evaluation of the program, however, was able to measure the impact of some program activities by taking a quasi-experimental design approach to the program's abundance of television sales data. The paper describes two cases in which evaluators were able to quantify the impact of program activities on television sales, as well as several analytical approaches with potential in future evaluations. This is followed by a discussion of the elements of this midstream program that are unquantifiable, yet in which impact was conclusively demonstrated using qualitative evidence. The paper concludes with a summary of what NEEA learned from the study and recommendations for future midstream program design and evaluation.

Introduction

"Midstream" energy efficiency programs take their name from the point in a product's supply chain they seek to influence – the middle ground between the product's manufacturers and suppliers (the "upstream" players) and the end user. For most consumer goods, including appliances and major electronics, this middle ground is dominated by a handful of retailers. For example, four retailers together account for two-thirds of refrigerator sales (Sears, Lowe's, Home Depot, and Best Buy) and two retailers account for more than half of television sales (Best Buy and Walmart) (Research Into Action, Inc., 2011 and 2012).

The consolidated nature of the retail sector increases the attractiveness of midstream programs because of their potential for operational efficiency, and with it a positive benefit/cost ratio. Midstream programs can aspire to several desirable impacts on retailers. Midstream programs may influence product manufacturers and consumers (via the retailers), two groups that are difficult and expensive to reach otherwise. Midstream programs also facilitate access to top corporate decision-makers and thus greater potential for influence over product assortment and access to product sales data. Along with these advantages, there are also drawbacks. It is difficult to evaluate the level of impact midstream programs have on retailers' complex assortment and promotion decisions, manufacturers' product

design decisions, and ultimately product sales. There are few evaluation precedents for midstream programs.[1]

This paper uses a selective review of the second evaluation of the television program operated by the Northwest Energy Efficiency Alliance (NEEA), covering 2011, as a case study in evaluating midstream energy efficiency programs.[2] We provide an overview of the program's activities, expected outcomes, and NEEA's research questions. We discuss the challenges to answering some of these questions - and how the study design attempted to overcome them. We generalize from our experience to offer recommendations for midstream program design and evaluation more generally. In conclusion, we examine what NEEA program managers learned from the evaluation.

NEEA's Television Program

NEEA is a non-profit organization that works to accelerate the innovation and adoption of energy-efficient products, services, and practices in the Pacific Northwest. In 2009, NEEA launched its Consumer Electronics (Television) Initiative (the program) with the goal of increasing the availability of energy efficient televisions at retail stores in its four-state region. In operating this program, NEEA is a member of the "Business and Consumer Electronics (BCE) Alliance," a cooperative effort between Pacific Gas & Electric, Sacramento Municipal Utility District, San Diego Gas & Electric, and Nevada Energy. BCE Alliance sponsors utilize the same implementation contractor and qualified product criteria, but other elements of their implementation differ, including their per-unit incentive payments, total incentive payments, and point-of-purchase (POP) materials.

As a midstream program, NEEA's television program pays incentives to the retailers who sell the products, rather than the end users who buy them. The program theory postulates payment of an incentive for each qualified product sale will lead to several outcomes, most important of which is that retailers will bring more qualified products into their television assortments than they would absent the program, resulting in increased sales of energy efficient televisions.[3]

The program's other activities include sustained engagement with retailers' corporate sustainability staff and/or television buyers; delivery of timely, detailed information about the program's energy specification to help retailers identify qualified models when making their assortment decisions; employment of retail store "detailers" who visit each participating store location three to 10 times per year to place POP marketing materials on qualified televisions; coordination of social media-driven marketing campaigns; and participation in the ENERGY STAR® television specification development process to encourage the strictest possible energy consumption targets.

NEEA expects these activities to result in several outcomes of varying importance to the program's success. The most significant expected outcome, described above, is a shift in retailers' television assortments to include a higher proportion of program-qualified products.[4] Other outcomes

[1] Previous evaluations include a theory white paper and a baseline study conducted in advance of the program launch (ODC 2009, 2010); and an evaluation of NEEA's 2010 television program and a program "review" of a similar program in California in 2011 (EMI 2011, 2012).

[2] The full evaluation report is available for download from NEEA's website, at http://neea.org/resource-center/market-research-and-evaluation-reports. The full study includes detailed findings and all data collection instruments. The study is the source of all quotes used in this paper, unless noted otherwise.

[3] In retail, "assortment" is the mix of products available for purchase at any given time. TV assortment is seasonal, with retail buyers selecting the majority of TV models in the Fall, for delivery the following Spring. A small number of models may be brought in at other times of the year for special promotions or to replace a model that is not selling well.

[4] The program criteria specify the minimum efficiency level a TV must meet to qualify for an incentive. The program identifies two tiers, paying a higher incentive for sales of more efficient products. The criteria changed annually from 2009 to 2012 and stipulated a consumption level below (more efficient than) that of the ENERGY STAR specification currently in

were expected to contribute to increased sales of energy efficient televisions. They included retailers taking actions to accelerate the sales of program-qualified televisions; retailers increasing their demand on television manufacturers for program-qualified televisions; improvement in consumers' and retail sales associates' ability to identify program-qualified televisions; and increasing stringency of the ENERGY STAR television specification.

The Study Design

The study design for the evaluation of NEEA's 2011 television program was driven by NEEA's explicit research questions – the evaluation kick-off meeting yielded more than 20 of them. NEEA's several research questions fell into four key areas of inquiry, three of which provided evaluators with methodological challenges (a fourth research area, regarding NEEA's influence on ENERGY STAR specifications, was more easily addressed): What was the program's influence on increasing sales of qualified televisions? What was the program's influence on television retailers? What was the program's influence on television manufacturers?

In order to answer these three questions, evaluators designed a multi-modal study with five distinct data collection activities, including qualitative and quantitative analytical methods. The study includes activities similar to those used in the evaluation of downstream programs: a review of program data, in-depth interviews, and a telephone survey. The study also includes activities that occur less often in downstream evaluation: "mystery shopping" visits to retailers, several in-depth interviews with participants, and a statistical analysis of sales data. Table 1 shows the evaluation activities, data sources, sample size, and the research areas addressed. The resulting successes and challenges are discussed below.

Table 1. NEEA Television Program Evaluation Data Collection Activities

Activity	Source of Data	Interviewees (Firms)	Research Topics
Program Data Review	Salesforce.com Database	N/A	Influence of program on retailer behavior Influence of program on ENERGY STAR specifications Characteristics of participating stores
	Store detailer reports (from monthly store visits)		
	ENERGY STAR specification documents		
	Comparison of findings with raw data from the 2010 evaluation		
In-Depth Interviews	Program Staff	9 (3)	Influence of program on retailer behavior, including assortment decisions Influence of program on product design Influence of program on ENERGY STAR specifications
	Other television Program Managers	2 (2)	
	EPA ENERGY STAR Staff	3 (1)	
	Retailer Corporate Staff	11 (5)	
	Manufacturers	4 (4)	

effect. For example, in 2012 the program criteria required products to exceed ENERGY STAR v5 by 20% (lower tier) or 35% (higher tier).

Activity	Source of Data	Interviewees (Firms)	Research Topics
Store/Department Manager Surveys	Retail store and television department managers	30 (7)	Influence of program at store level Store-level awareness of program
Retail Store Visits	"Mystery shopping" visits to participating stores	59 (10)	Store-level awareness and attitudes toward efficiency Television promotional activities
Sales Data Analysis	Television sales data submitted to online program incentives database	N/A	Characteristics of Northwest television market Program influence on sales of qualified televisions

Quantifying Program Influence on Qualified Television Sales

NEEA's television program, like all its energy efficiency work, is designed and implemented as a market transformation program. Its aim is to catalyze lasting change to existing practices for television retailers and in the broader television marketplace. NEEA wanted to look at all possible approaches to measuring the program's impact on television sales. The evaluation team was able to make use of an abundance of sales data provided to the program by participating retailers, and although data were not available to carry out all the desired research approaches, the team did find that one of the program's in-store marketing tactics resulted in an increase in sales of qualified televisions.

One benefit of midstream programs is access to detailed unit sales data provided by the participating retailers. To some extent, these data can be obtained in other ways; for example, purchased from a data aggregator, such as NPD or tabulated based on submitted rebates (in an end-user focused program). Retailer-sourced data, however, has several advantages. Retailer data includes identifying information for each unit sold, such as retail chain and store, that is absent from aggregated third-party data and necessary for statistical analyses of program influence. Retailer-sourced data may also include fields of use to program managers and evaluators, such as the presence or absence of efficiency features. And retailer-sourced data, unlike unit sales data derived from end users' rebate submissions, has the potential to include "full category" data, meaning *all* unit sales not merely qualified unit sales.

Evaluators achieved two notable successes in the effort to use sales data to quantify the impact of this midstream program. There were also several challenges, some of which the evaluation team was able to overcome, and others that should be addressed in future studies.

Four Analytical Approaches to Quantitative Analysis of Program Impact

Evaluators identified four analytical approaches with potential for evaluating midstream programs, should the program design and data collection support them. These approaches are informed by similar analyses conducted in an academic context and published in business and economics journals (Cho, 2009; DeHoratius, 2007; DeHoratius, 2011; Raman, 2009).

Compare periods before and after a change in program criteria. Evaluators hypothesized that if, in fact, retailers were acting to boost sales of qualified television models, there would be a change in the proportion of sales of certain television models around periods in which the program

criteria changed. For example, evaluators speculated they would detect a decrease in sales of televisions that no longer qualified for an incentive in the "post" period. Two factors curtailed this approach. First, the program's sales data lacked information on non-energy related television features, such as unit sale price, that could confound this analysis and needed to be controlled for. Interviews revealed a second confounding factor. Program criteria changed on January 1, so the "before" period fell over the winter holidays, when retailers typically assort low-priced televisions, skewing the proportion of qualified television sales. Further, field staff data and store manager interviews showed store-level staff lacked direction from corporate, awareness of the program and its qualification criteria, and did not conduct the type of promotion necessary to support the underlying hypothesis. This type of analysis may be possible if program qualifying criteria shifted at any time in which other elements in the retail environment were relatively stable, sales data contain enough depth to allow evaluators to control for confounding factors, and program activities support the hypothesis that retailers are acting to increase sales of qualified products.

Compare similar participating and non-participating stores. This approach requires detailed characteristics for all stores in order to match participating and non-participating stores. The approach also requires sales data from non-participating stores at the same level of detail as those received from participating stores. As noted above, the data required to make this comparison were lacking, as were the program design elements required to show variation.

Compare major national chains and small chain and non-chain stores. The small data sample size and the lack of sales dates prevented this analysis. Although some minor and non-chain stores provided full category data, many provided only partial year data for 2011, precluding many potential statistical comparisons. Additionally, buying groups typically provided the sales data for minor and non-chain stores, rather than the stores themselves, and thus the date-of-sale included with each unit was the date the unit shipped from the buying group to the store, not the date-of-sale to the customer. Lacking a probable estimate for the length of turnover for these units, evaluators could not accurately compare sales between major chain and non-major chain stores.

Quantitatively assess influence of promotional activities. The data were sufficient to support analysis of one of the program's promotional activities: the airing of a 15-second promotional video on in-store video walls at one of the major retailers. For this approach, evaluators required the dates and locations of promotional activities, a description of the activity, and a non-promotional period with enough similarity to the promotional period that it could be judged to control for other possible influences on television sales. This approach has perhaps the greatest potential for use in evaluating midstream programs, if evaluators and program managers work together during program design to ensure that program activities and data collection will support experimental or quasi-experimental analysis.

Three Challenges

There were three key challenges to using sales data to quantify the impact of this midstream program using a quasi-experimental method following one of the four approaches described above. Most could not be overcome, primarily because the study was retrospective and most activities were not designed to support this type of evaluation. However, program managers who work closely with evaluators during the design of program activities can hope to alleviate most of the barriers described below through modifications to the program design and improved data collection.

Lack of a control group. NEEA's television program focused on building relationships with national retail chains and conducted limited regional activities, other than field staff visits to stores to place program point-of-purchase materials. National retailers have centralized decision-making and their stores are incredibly consistent in their execution of headquarters' plans. As a result, the program's influence affected the energy efficiency of televisions available nationally, not just in NEEA territory. This spillover effect benefits consumers, but prevented evaluators' from using sales data to quantify the program's impact, as there was no viable control group.

Insufficient data to define a comparison group. In order to create a comparison group the team required two types of information. The team needed sales data from non-participating stores (at retailers who *were* participating in the program) in areas outside NEEA territory and without any other television energy efficiency programs. The team also needed detailed store characteristics for all stores to support the use of propensity score analysis, a statistical technique for creating treatment and control groups. None of the retailer participants supplied store characteristics and only one retailer supplied sales data from non-participating stores.

Varying levels of data quality. Evaluators needed full category sales data in daily or at least weekly intervals to allow for the greatest possible statistical precision and flexibility. Of the four biggest retailers, only two provided data of this quality.

Two Successes

The evaluation team used statistical analyses of sales data to document the program's influence on television sales in two scenarios. A program-produced video showing the POP and tying it to energy efficiency, aired at all stores in one major retail chain, increased sales of top-tier televisions by three percentage points in months when the video played compared to months when it did not (11% of total sales compared to 8%). This analysis was only possible because of a lucky turn of events. Because of an agreement between NEEA and the retailer, unrelated to any future analysis, the video aired every-other month for most of 2011, providing both "control" and "treatment" periods. Evaluators used a repeated measures/paired samples t-test to compare the proportion of sales of top-tier televisions, by store, in months when the video did and did not play. Table 2 shows the mean, standard error, and t-test values for the shares of top-tier sales in control months and treatment months.

Table 2. Statistics for shares of top-tier television sales in control and treatment months

Condition	Months	Mean share of top-tier television sales	Standard error	T-test value
Control (No Video)	January, March, May, July, September	8%	.02%	t(49) = -8.1, p<.0001
Treatment (Video)	February, April, June, August, October	11%	.02%	

Evaluators were also able to quantify the relationship between television assortment and sales using a hierarchical linear model.[5] The model showed an increase in program-qualified models in a retailer's assortment is associated with an increase in sales of those products. For qualified televisions, a 1% increase in assortment is associated with a 1.3% increase in sales. This finding is statistically significant and demonstrates that the program's influence on television assortment had a positive significant impact on television sales. Table 3 lists inputs to the model and associated statistics.

Table 3. Inputs to the Hierarchical Linear Model

Input	Definition	Change in qualified TV models sold	Standard error	T-test value
Month	Month in which the sales occurred	0.07%	0	t(5206) = 74.7, p<.0001
Specification change	Whether sales occurred within one month of a change in program specifications (December-January)	-0.47%	0.02	t(5206) = -28.5, p<.0001
Month*Specification change	Interaction effect	-.03%	0	t(5206) = -26.7, p<.0001
Urban	Whether the store at which the sales occurred was in an urban zip code	-0.08%	0.07	t(241) = -1.2, p=.24
Population density	Population per square foot according to U.S. census data	0.05%	0.02	t(241) = 2.1, p=.03
Display	Proportion of program-qualified TVs on display at the store at which the sales occurred	1.3%	0.08	t(241) = 16.2, p<.0001
Turns	Retailer's annual inventory turnover, from publicly-available sources	0.04%	0	t(241) = 51.4, p<.0001
Income	Median income for the store's zip code	0	0	t(241) = -1.46, p=0.15

Quantifying Program Influence on Business Decision-Making

A key element of NEEA's television program is to build close relationships with market actors and to influence their business-as-usual practices. The evaluation showed, through a number of data collection activities, that NEEA's efforts on this front had a verifiable impact and that the program, to a

[5] Evaluators created a hierarchical linear regression model with two levels (retail chain and store) to explain variation in the outcome variable, the percent of qualified televisions sold.

large degree, accomplished what it set out to do. However, the evaluation also showed that the barriers to quantifying the program's influence on business decision-making are likely insurmountable, absent a radical openness on the part of the participants, which seems unlikely to occur.

Success in Documenting Impact through Qualitative Findings

This second evaluation of NEEA's television program found the qualitative and anecdotal evidence of NEEA's influence on its retailer participants to be strong and convincing. Below we describe two of the evaluation's key findings regarding the program's influence on retailers.

The program improved retailers' awareness of, and attitude towards, energy efficiency and midstream efficiency programs. In-depth interviews, and the review of correspondence between the program implementer and the participants, showed that energy efficiency has become part of retailers' business-as-usual practices, in concert with an attention to sustainability at the corporate level. Major chain retailers reported increased knowledge of energy efficiency as a result of their participation in the program. As one of the first and still the largest television efficiency program, the television program operated by NEEA and the co-sponsors in California and Nevada piqued retailers' interest in midstream incentive programs, and all major chain retailers reported that their companies were participating in ten to 15 television incentive programs in 2011. In reference to the program, one major chain retailer reported that there was "lots of mindshare on this program" within the company, and that messages encouraging key employees to consider energy efficiency were "a regular drumbeat." Retailers' use of staff resources also demonstrates their commitment to midstream efficiency programs. Retailers reported creating new in-house positions and hiring consultants to facilitate their participation in utility programs. Retailers' willingness to make exceptions to regular policies on behalf of the program further demonstrates their engagement.

The program influenced retailers' assortment decisions. Abundant evidence obtained through in-depth interviews and a review of program correspondence confirmed that the program succeeded in influencing retailers to consider its qualification criteria during their assortment decision-making. Everyone interviewed for this evaluation was well informed of the program's efficiency specification, expressed cordial familiarity with the program implementer, and appeared to understand NEEA's objectives. Retailers reported receiving the qualification criteria in a timely manner and in a clear format that they were able to use within their organization. Interviews also confirmed that the program's relationship with retailers resulted in influence on the assortment decision that would not have occurred otherwise, and provided specific examples. From a financial perspective, three major chain retailers allocated whole or partial credit for program incentive dollars to the merchants' business unit, making television assortment decision-makers the direct beneficiaries of the program's incentive dollars. These findings contrasted to some extent with the results of the first evaluation of the program, published in 2011 and focused on the program's 2010 activities (Energy Market Innovations, Inc., 2011). Both evaluations found the program had established strong relationships with retailers. But the second evaluation (the subject of this paper) differed from the first in its assessment of the program's influence on retailers' business decision-making. The first evaluation argued the program had not influenced retailers' assortment decisions in a substantial or quantifiable way, while the second asserted that the program *did* influence these decisions in a way that was meaningful, if not quantifiable. One possible explanation for the difference in findings between the two evaluations is that program effectiveness increased over time. Another possibility is that the different evaluation approaches resulted in correspondingly different findings.

Challenges in Quantifying Influence on Business Decision-Making

The context of strong, dissenting opinions about the program's success was a prelude to the second evaluation, and to the challenges inherent in assessing, and particularly in *quantifying*, influence on supply chain decision-making, some of which only became evident during the course of the study. The barriers evaluators faced were remarkably similar in both the retailer and manufacturer populations, and may be encountered by evaluators of midstream programs focused on other product areas or other market actors.

Limited access to the decision-makers. Retail product buyers or "merchants," the ultimate decision-makers who the program sought to influence, were usually unwilling to speak with evaluators. Evaluators were only allowed to speak with buyers at one retail organization, even after repeated requests to different individuals. Questions submitted by email went unanswered. Among major television manufacturers, the targeted interviewees (product managers and design decision-makers) were also unwilling to be interviewed. Most manufacturers' design staffs are located in Asia, and evaluators were told repeatedly that these individuals might not feel comfortable conversing in English.

The confidential nature of the decision-making process. The decisions the team was tasked with evaluating - retailers' assortment selection and manufacturers' product designs - are held in utmost secrecy. As areas of competitive intelligence, interviewees were willing to speak only in generalities about their organization's processes and considerations. No one was willing to share evidence that would have allowed the team to quantify or even just verify program influence, for example, a product selection algorithm or product matrix. As one retailer interviewee commented about his company's television selection algorithm, "Pursuit of confirmation of what that algorithm is would be futile. I work [there], and they won't even tell me. . . It's like asking for the secret combination for how to make Coca-Cola. They're just not going to tell you."

Data availability and market conditions confound approaches to quantify impact on decision-making. Although the evaluation team believed it would be *possible* to quantify the impact of the program on retailers' assortment decisions and manufacturers' product design decisions, the challenges to obtaining the data required were judged too onerous to include in this study, as they would have required, for example, a list of every television model manufacturers made available to retailers during the course of several model years, and their features and energy efficiency characteristics. The strategies the team used to get around these barriers, at least as far as was possible, may be useful to other evaluators who have a need to go beyond the midstream participants' own perspective on a program in order to understand its broader context. They may also help explain why the findings of the second evaluation differed from the first. As a result of the first evaluation of the program, the research team knew their go-to data collection activity for deeply qualitative information, the in-depth interview, would need to be supplemented by additional data sources. Interviews with corporate decision-makers at the major national retailers, conducted for that evaluation, showed the decision-makers were unable to quantify the influence of the program on their television assortments. The 2010 evaluation paraphrased one major retailer interviewee as referring to the program as "a tie-breaker when deciding between two very similar models." The responses obtained during the 2010 evaluation to questions like "What impact do the current incentives have on encouraging your organization to *purchase* and *stock* more energy-efficient televisions in the Northwest market?", also showed retailers did not attribute the program as a primary influence over their assortment decisions. The 2011 evaluation team's tactics for overcoming these challenges were to:

Get familiar with the decision-maker's world. Evaluators consulted with a business professor who is an expert in the study of the retail field and read widely in the academic and popular press about the participants' organizations and industry. When speaking with retailers and manufacturers, evaluators inquired into the conditions in which they operate. For example, in seeking to understand the program's potential influence over the assortment decision, interviewees were asked about how their television assortment decisions are made, who makes them, what influences them, and how they have changed over time.

Use multiple informants at each organization. Business decisions may ultimately be made by a single individual, but typically many people contribute to them. Evaluators spoke with multiple respondents in varying positions within each retail organization, all of whom attempted to influence the television buyers' decisions. For example, at one major retailer, evaluators interviewed the sustainability lead, the utility coordinator, and a senior director focused on energy management. Evaluators had several conversations with many of the interviewees, some of which occurred in person at industry events.

Ask indirect questions and capture the nuance. In addition to asking a direct question like, "Do you have a sense of how often the fact that a television is qualified for this program weighs into your decision to buy it?" evaluators also asked many indirect questions like, "What, if anything, would you have to *stop* doing if this program ended?" Another example of this approach was in the series of questions about the implementer and the information they provided. Evaluators chose not to use typical satisfaction questions and instead asked both about respondents' familiarity ("Would you say that you feel well-informed about upcoming changes to the program's specifications?"), usefulness ("How do you use the information from the program in your conversations with manufacturers?") and suggestions for improvement ("The implementer told us that they check in with you throughout the year. Is there anything they could do to better support you?").

Hypothesize about what might have happened if the program achieved its desired influence, and test these hypotheses. This tactic was particularly important and guided nearly every aspect of the data collection process, from interview questions to the inclusion of new data collection methods. Mystery shopping visits to participating stores and interviews with store managers, for example, were included in the study to test some of the teams' hypotheses, and resulted in findings that were some of the most helpful to program managers in improving the program. A few examples:

- Evaluators speculated that if the program influenced buying decisions, retail buyers would demonstrate familiarity with the qualifying product criteria, have received them well in advance of their decision-making, and have discussed the specifications with manufacturers. Evaluators sought to document whether this was the case in their questions to retailers and manufacturers and close examination of the content and chronology of the program implementer's correspondence.
- Evaluators theorized that if the program influenced retailers to work to increase sales of qualified products, corporate executives (the program's primary contacts within the retail organizations) might have taken several possible actions, including: communicating program details to store managers and/or sales associates; providing incentives to store managers and/or sales associates for increased sales of qualified products; and have increased promotion of qualified products. Evaluators interviewed store managers and conducted "mystery shopping" visits to participating stores to document whether influence of this type was observable.

In the above approach to assessing program influence on retailers' business practices, and assortment decision-making specifically, evaluators contextualized their findings based on what they learned about the retailer's *milieu*. As in the first evaluation, market actors told this study's authors that, although they did consider whether a particular television model qualified for a program incentive, this was never the primary factor in their decision to include the product in their assortment. But in considering the broader context of the assortment process, evaluators recognized just how complex and critically important those decisions are to retailers' financial success. Supply chain security, product features, value to the customer, and negotiations with manufacturers take priority over energy efficiency program qualifications. One interviewee from a major U.S. retailer said balancing all of the considerations is like putting together a jigsaw puzzle with "ten or twelve really big pieces."

In this context, it became clear to evaluators that energy efficiency would never trump factors like the retailers' need to minimize the potential for disruptions to product availability. Nor would retailers bring products into the assortment if doing so jeopardizes their relationship with manufacturers, who provide retailers with "market development funds" as an incentive for retailers to assort particular models or a particular number of models, or promote their products in certain ways. Evaluators concluded that any energy efficiency program's potential for influence would necessarily be limited by the predominance of other considerations more central to a retailer's business needs.

These broader insights were used to contextualize the evaluation finding that the NEEA television program was never a retailer's *primary reason* for assorting a particular television model, but that retailers *did consider* whether a product met the program's efficiency qualifications during their assortment decision process. Judged from this frame of reference, the program's influence appeared both substantial and meaningful, although in ways not historically used to evaluate end-user focused assessments of free ridership or net-to-gross calculations.

What NEEA Learned and Integrated

The study confirmed NEEA's observations that its activities had an influence on the television supply chain and that its midstream program design was having an impact on retailers' assortment decisions and business practices around energy efficiency. NEEA also learned it could further increase qualified unit sales and improve the program's evaluability by increasing activities at the store level.

In response to the study's findings, NEEA is making several changes to its 2013 program design. First, NEEA is revising the program's logic model to reflect new information about the outcomes it can expect as a result of its activities. For example, NEEA learned that per-unit sales incentives alone are unlikely to result in retailer-driven promotional activities. NEEA is also refining promotional activities at the store level, including incorporating current adult education methodologies into its store associate interaction scripts and working to secure additional time on retailers' in-store video walls. NEEA is designing these activities in close cooperation with the evaluation team to employ an experimental design approach, in an effort to address and minimize the challenges to quantifying impact described above. NEEA is also exploring additional ways to leverage its incentive dollars with its retailer participants.

Finally, the study reinforced NEEA's perspective on its relationship with its retailer participants. NEEA knew the retailers' valued the television program, with its educational outreach and incentive dollars. As a result of the findings, NEEA's understanding of the importance of long-term partnerships with retailers was reinforced and NEEA learned additional methods for capitalizing on that long-term relationship.

References

Cho, S., McCardle, K., and C. Tang. 2009. "Optimal Pricing and Rebate Strategies in a Two-level Supply Chain." Production and Operations Management 18 (4): 426-446.

DeHoratius, N. and A. Raman. 2007. "Store Manager Incentive Design and Retail Performance: An Exploratory Investigation." Manufacturing & Service Operations Management 9 (4): 518-534.

Energy Market Innovations, Inc. 2011. *Consumer Electronics Television Initiative Market Progress Evaluation Report.* Report #E11-230. Portland, Ore.: Northwest Energy Efficiency Alliance.

Opinion Dynamics Corporation. 2009. *Statewide Business and Consumer Electronics Baseline Study.* Study ID: PGE0283.01. Pacific Gas & Electric Company, Southern California Edison, and San Diego Gas & Electric.

Opinion Dynamics Corporation. 2010. *The Market for Energy Efficient Electronics: Pre-Program Findings on Consumer Perceptions and Retail Shelf Stocking Practices*. November 1, 2010.

Raman, A., DeHoratius, N., and Z. Kanji. 2009. *Supply Chain Optimization at Hugo Boss (A).* Harvard Business School Case Study # 9-609-029. Boston, Mass.: President and Fellows of Harvard College.

Research Into Action, Inc. and Nicole DeHoratius. 2013. *Consumer Electronics Television Initiative Market Progress Evaluation Report #2.* Report #E13-255. Portland, Ore.: Northwest Energy Efficiency Alliance.

Research Into Action, Inc. and Energy Market Innovations, Inc. 2012. *Program & Technology Review of Two Residential Product Programs: Home Energy Efficiency Rebate (HEER) / Business & Consumer Electronics (BCE) (Final).* Study #SCE0306. Rosemead, Calif.: Southern California Edison.

Research Into Action, Inc. 2011. *Energy Savings Opportunities and Market Descriptions for Four Residential Consumer Electronics Products (Final).* Report #11-227. Portland, Ore.: Northwest Energy Efficiency Alliance.