Observing Where the Rubber Meets the Road: Integrating Mystery Shopping and Ride-Alongs into Process Evaluations

Doré Mangan, Research Into Action, Portland, OR
Zac Hathaway, Research Into Action, Portland, OR
Jun Suzuki, Research Into Action, Portland, OR
Joe Van Clock, Research Into Action, Portland, OR
Erika Kociolek, Energy Trust of Oregon, Portland, OR

ABSTRACT

Many energy efficient retail product programs, whether using upstream, midstream, or downstream incentives, rely on partnerships with retailers to meet program objectives. Evaluators often use program data, staff interviews, and consumer surveys to gain insights into program processes. While these data provide a great deal of value for programs, they sometimes fail to capture the consumer experience on the sales floor, including interactions with retail sales staff. The sales floor is where the “rubber meets the road” and important consumer interactions occur. Mystery shopping and ride-along methodologies provide a glimpse into these critical interactions. This paper discusses experiences integrating these methods as a part of process evaluations.

Drawing from process evaluations of energy efficient retail product programs, the authors present details on:

- When and why to use mystery shopping and ride-along approaches in energy efficiency process evaluations;
- The challenges of conducting these activities; and
- Lessons learned during evaluation design and data collection.

The mystery shopping and ride-along methodologies discussed in this paper helped identify several important opportunities for improvement. For example, mystery shopper observations led to recommendations on more informative point-of-purchase (POP) promotional materials. Additionally, observations revealed a lack of knowledge about program POP materials and the efficiency of incented products among sales associates, indicating a need for additional focused training.

Introduction

This paper pulls from the authors’ experiences conducting mystery shopping and ride-along activities and using these types of observational data in process evaluations, particularly the 2013 process evaluation of Energy Trust of Oregon’s Products program and the 2014 market process evaluation for the Northwest Energy Efficiency Alliance’s (NEEA) Consumer Electronics Television Initiative. Using findings from these two process evaluations as case studies, this paper presents the benefits and challenges of integrating this type of observational data into process evaluations of energy efficiency programs. This paper begins by providing background on mystery shopping and ride-alongs, followed by a description of the mystery shopping and ride-alongs conducted for Energy Trust and NEEA, and ends with a summary of the value of this observational data, along with the benefits and challenges of performing these activities.
Background: Use of Observational Data in Market Research and Beyond

Mystery shopping is a qualitative, observational data collection method in which evaluators act as typical customers without revealing their identity and report on those experiences. Market researchers use this approach to evaluate customer service and determine if services or policies are being carried out as intended (Szwarc 2005). For process evaluations of energy efficiency programs, researchers typically observe how employees of retail stores present energy efficient products to customers, as well as the placement accuracy of marketing materials. In addition to in-store data collection, mystery shopping could include calling a helpline or visiting a website (Hamersveld 2008).

The ride-along is a similar method of observational data collection in which evaluators shadow individuals as they perform their regular work duties. Unlike mystery shopping, during ride-alongs the observed individual is aware of the evaluator collecting information and the observation typically extends beyond a single interaction. Evaluators use the term “ride-along” because the individuals observed are often program field staff visiting multiple stores or service locations, hence the observer literally rides in the vehicle along with field staff to directly observe and inquire on a typical workday. Evaluators may observe field staff as they visit retailers to place and ensure accuracy of POP materials and interact with sales associates. Ride-alongs are not limited to the retail environment. For example, evaluators may also observe audit or upgrade contractors as they conduct assessments or inspections and interact with customers.

Literature first documenting mystery shopping used among market researchers goes back more than 80 years. Market researchers developed the mystery shopping technique from methods used by private investigators to assess staff misconduct. The term “mystery shopping” was coined in the 1940s when the method began to be used more frequently (Hamersveld 2008). Today, researchers use mystery shopping worldwide, especially in such industries as retail, leisure, and finance; however, there has also been increased use by local and central government as well as regulatory bodies. For example, many local government bodies in the UK have used this technique to monitor service delivery, particularly by helplines (Hamersveld 2008). Although ride-alongs are common in certain situations (for example, interested or concerned citizens can ride along with law enforcement), there is less literature on the use of ride-alongs as a data collection method in market research. Organizations may use this method more often as an internal training or quality control exercise rather than a research activity. Nonetheless, both ride-alongs and mystery shopping fall into the broader research area of direct observation, which is a widely accepted data collection method in academic literature.

Observational data collection methods provide a level of depth and actionable detail that customer and employee surveys may not provide (Patton 2015). Observational data collection offers an opportunity to gather data on the customer experience when doing so would not be practical, or possible, by other means. For example, programs providing incentives to midstream and upstream market actors such as retailers may not have lists of end-user participants available for evaluators to survey. Using data from mystery shopping and ride-alongs together allows for an understanding of the context in which key customer interactions occur. These methods allow for a broader view, which can lead to more effectively interpreted findings from other data collection activities such as customer surveys or staff interviews. By providing the evaluator with a unique perspective from both the customer and from program staff, observational data improves the reliability of energy efficiency program evaluations (Ridge 2000).
Evaluators of energy efficiency programs often rely on program data, interviews with staff, and consumer surveys to gain insights into program processes. Industry literature documents many examples that report findings using these techniques; however, observational methods are often underutilized (Peters 2009). Some leading market researchers predict that due to the rapidly increasing use of customer surveys in the last decade, the effectiveness of quantitative surveys as a method to gain customer insights will decrease; conversely, more qualitative techniques as well as advanced analytics will increase in importance (Temkin 2014, Patton 2014). Regardless of the effectiveness of quantitative surveys, integrating qualitative methods into energy efficiency evaluations provides additional context to quantitative data collected, which offers a more in-depth understanding of program challenges. The following sections provide two case studies that describe how evaluators have effectively integrated mystery shopping and ride-alongs into efficiency process evaluations.

**Case Study: Energy Trust of Oregon Products Program Process Evaluation**

In 2014, the authors performed mystery shopping and ride-alongs, together with in-depth interviews and analysis of program data, to evaluate Energy Trust of Oregon’s Products program (Research Into Action 2014). The Products program promotes awareness and purchase of energy-efficient home products through 1) the delivery of incentives to residential customers for qualified ENERGY STAR® appliances and 2) markdowns of efficient lighting and showerheads at retail stores.¹ For appliances, the program takes a downstream approach, providing customers with incentives for the purchase of qualified appliances, whereas incentives for lighting and showerheads are provided to the retailer, who then marks down the prices of efficient products. In addition to incentives, the program provides support to retailers in the form of POP and other materials, which are provided and placed by field staff members visiting the each store on a regular basis. Field staff also provide training to sales associates and are available to answer questions about the program and energy-efficient products.

Two key goals of this evaluation were to: 1) to assess sales associate knowledge and awareness of the program and 2) document how field staff conduct store visits. The authors primarily used mystery shopping to address the first goal and ride-along observations to address the second goal.

**Mystery Shopping Methodology**

Mystery shopping consisted of in-store visits to 14 out of about 450 participating appliance and lighting stores. Mystery shoppers spoke with 19 sales associates who worked in the lighting or appliance departments of these retail stores. Mystery shoppers completed three tasks during the in-store visits: 1) conduct the mystery shop, 2) debriefed the sales associate, and 3) gathered data about data about energy efficiency-related POP marketing materials (i.e., stickers, videos, signs, and brochures).

In order to ensure consistency across mystery shopping visits, the evaluation team provided each mystery shopper with a conversation guide that included detailed directions. Directions included: how to identify the sales associates that regularly worked in the department; how to ask

¹ Although the program is designed to target residential customers, some small commercial customers obtaining products through retail channels may be taking advantage of the program’s offerings.
sales associates specific questions about the targeted products; the debriefing process; and a data collection form for information about POP and other marketing materials.

During store visits, mystery shoppers waited for a sales associate to approach them in either the appliance or lighting department. If a sales associate did not arrive within a reasonable amount of time (within 5 minutes), the mystery shopper sought out a sales associate that regularly worked in the department. Mystery shoppers asked about products with specific characteristics and listened for mentions of either the program or the efficiency of the products. If the sales associate did not mention these key topics, the mystery shopper asked the sales associate about the importance of energy efficiency in purchasing an appliance or light bulb and if there were any ENERGY STAR or Energy Trust program qualified products available. If the sales associate did not point out or explain program POP materials, mystery shoppers inquired about the meaning of those materials. Throughout the visit, mystery shoppers discreetly took notes on a scrap of paper or small notebook to help capture sales associates’ responses as close to verbatim as possible.

After a mystery shop concluded, shoppers thanked the sales associates, disclosed their identity, and debriefed the associates. During the debriefing, mystery shoppers asked the sales associate a series of questions regarding work history, training, and program experience. Finally, mystery shoppers observed and documented in-store promotional materials and information on each appliance or light bulb shown during the mystery shop. Information included details on the products shown to the mystery shopper such as: brand, configuration, price, and whether the product was ENERGY STAR qualified or qualified for an Energy Trust rebate. Mystery shoppers would also collect data on any type of efficiency-related POP marketing materials on or around other products in the department.

**Ride-Along Methodology**

For the ride-alongs, evaluators observed 21 store visits conducted by four program field staff, two field staff servicing lighting and showerheads and two servicing appliances. In order to gather data on field staff experience in a wider range of participating stores, evaluators conducted one day of ride-along observations in the Portland, Oregon metro area and one day of observations outside the Portland metro area with field staff servicing each product type. Evaluators filled out an observation guide for each store visit that consisted of three sections: 1) observations of interactions with sales associates, 2) observation of POP placement, and 3) questions for field staff. When possible, evaluators took photos to illustrate issues field staff encountered.

When observing field staff interactions with sales associates, evaluators documented ways in which the associates demonstrated knowledge of the program. Additionally, evaluators documented how sales associates interacted with field staff. For instance, whether sales associates had prior knowledge about the program or if there was any resistance to the field staff activities. Evaluators also observed the placement of POP materials on qualified products, specifically noting the number of products with missing or incorrect materials as well as any challenges encountered during placement of POP materials.

Evaluators asked field staff about their experiences during a typical day. Questions asked during each store visit were designed to help the evaluators determine whether the observed interactions had been typical, and consequently how generalizable any findings derived from the interactions might be. Topics included how the visit to a given store differed from a typical visit,
any differences between the observed store and others in the same chain, and any in-store changes since the field staff member’s last visit. In addition, the evaluators asked field staff questions throughout the day about their experiences with the program, placement of POP materials, and interacting with sales associates.

**Key Findings**

Mystery shopping activities uncovered two key findings that would have been difficult to determine through surveys with sales associates. First, mystery shoppers found that sales associates’ positive attitudes towards ENERGY STAR did not necessarily translate to recommendations of ENERGY STAR products to customers. During the debriefing portion of mystery shops, many sales associates self-reported positive attitudes toward ENERGY STAR, and the importance of recommending ENERGY STAR products to customers. However, no sales associates mentioned ENERGY STAR unprompted. In this case, interviews alone likely would have captured only sales associates’ reported affinity for ENERGY STAR products, leaving evaluators unaware that their behavior did not reflect these attitudes. It would have also been difficult to uncover this finding through customer surveys, as customers may not remember whether a sales associate mentioned ENERGY STAR products.

Second, mystery shoppers found that fewer sales associates were typically available in a store’s lighting departments than in the appliance department, which increased the significance of POP materials in informing customers of energy efficient lighting options. Additionally, mystery shoppers found the POP materials were not always clear or easily visible when looking for products. Mystery shopping also identified opportunities to improve the POP materials. When a shopper could find a lighting sales associate during the mystery shop, most sales associates could not correctly explain what the lighting POP material signified. The POP material, which displayed only the word “save” and the name of the program, did not provide enough information for customers to understand that labeled products were the most energy efficient options on their own.

Consistent with mystery shopping findings, ride-alongs also found that there were fewer sales associates in the lighting department than in the appliance department. Interactions between field staff and sales associates differed between appliance departments, which had dedicated sales associates with detailed product knowledge, and lighting departments, which were typically staffed by sales associates responsible for a much larger portion of the store. Evaluators also found that although there was little error in POP placement on qualified products or resistance from sales associates to field staff activities, field staff were limited in where they could place some POP material due to the placement of POP from other sources (manufacturers, retailers, etc.).

These observational data supported a conclusion that there are important differences between lighting and appliances in the retail environment. These differences suggested the need for increased training for sales associates on efficient lighting and lighting POP that could better communicate to customers without the assistance of sales associates.

**Case Study: Northwest Energy Efficiency Alliance Consumer Electronics Television Initiative Market Process Evaluation**

The authors conducted a market progress evaluation of NEEA’s Consumer Electronics Television Initiative (Research Into Action 2014). This midstream incentive program had a goal
of increasing the availability of highly energy-efficient televisions in participating consumer electronics retailers’ product assortments. The program offered financial incentives to retailers for each television sold that met program-defined efficiency specifications, which were typically more stringent than ENERGY STAR. Additionally, program field staff placed POP signage on qualified television models in participating retail stores to help consumers identify the most energy efficient televisions.

Two key goals of this evaluation were to: 1) determine sales associates’ awareness of the program and their knowledge of the energy efficiency of televisions, and 2) identify opportunities for NEEA to more effectively draw on field staff to help support current and future programs. The authors primarily used mystery shopping to address the first goal and ride-along observations to address the second goal.

**Mystery Shopping Methodology**

The methodology for mystery shopping was in line with the methods used for the Energy Trust evaluation. Shoppers visited 59 of 190 participating consumer electronics stores located in NEEA’s service territory to conduct the mystery shopping, debrief the sales associates about the purpose of the visit, and gather data about energy efficiency related POP marketing materials.

**Ride-Along Methodology**

The methodology for ride-alongs was in line with the methods used for the Energy Trust evaluation. Evaluators conducted two days of observations in NEEA’s service area to observe two program field staff who visited ten participating retailers to place POP materials on qualified televisions and inform sales associates about the Initiative. As in the Energy Trust evaluation, evaluators conducted one day of ride-along observations in the Portland metro area and one day outside the Portland area.

**Key Findings**

Mystery shopping revealed a disconnect between sales associates’ self-reported attitudes and the behavior evaluators observed. During the debriefing portion of the mystery shopping activity, mystery shoppers found sales associates had a positive attitude towards energy efficiency in general. Nonetheless, during mystery shopping most sales associates did not specifically promote energy efficiency when discussing television features, although they were able to locate energy efficient televisions once prompted by the shopper. In fact, some sales associates discouraged shoppers from considering the energy efficiency of televisions because of the relatively small differences in energy usage between models.

An additional key finding uncovered during mystery shopping was that none of the 59 sales associates that shoppers interacted with mentioned the program or pointed out the program POP materials when discussing televisions with shoppers. When shoppers asked sales associates to provide an explanation of the program POP materials, about half were able to provide an explanation, and fewer were able to provide accurate information.

During ride-alongs, evaluators observed that the field staff could effectively introduce the Initiative to sales associates, but noted the field staff had little new information to provide sales associates to build the associates’ understanding of TV energy efficiency in subsequent visits. In addition, evaluators found that field staff had difficulty providing details about the program’s specification levels and the process for determining which televisions qualified.
Another key finding from ride-along observations was related to the limited time field staff were able to spend interacting with sales associates. Timing each store visit, evaluators found that field staff spent an average of five minutes interacting with sales associates at each store. Understanding this time limitation allowed the evaluation team to provide more realistic recommendations about how to best use field staff in delivering training to sales associates.

Finally, evaluators observed sales associates informing field staff that few customers asked about program POP materials. These sales associates mentioned to field staff and the researchers that the relatively limited yearly energy cost savings potential for efficient TVs was a barrier to increased consumer uptake of efficient televisions.

Overall, combining mystery shopping and ride-alongs provided complementary findings. From mystery shopping, evaluators found sales associates do not promote televisions based on their energy efficiency, which made sense because findings from the ride-alongs suggested that sales associates found the cost-savings benefits to be trivial. Additionally, evaluators found that sales associates who reported interacting with program field staff had increased knowledge and awareness of both the program and ENERGY STAR. This finding suggested that interactions between field staff and sales associates were important to increasing awareness of the program. Additional sales associate training and information could potentially increase sales of energy efficient televisions, although field staff had limited time to deliver that training. Together, findings from mystery shopping and ride-along data, along with other data collection activities, led to two key recommendations: 1) develop online training for sales associates similar to online training already used by retailers and 2) develop alternative program POP materials with clearer, more effective messaging.

Values of Observational Data in Process Evaluations

The findings from the two case studies above highlight the unique contributions of observational methods, such as mystery shopping and ride-alongs, and the actionable insights they provided. Direct observational methods can contribute to process evaluations in three key ways: 1) identification of discrepancies in implementation allows for unanticipated findings; 2) identification of breakdowns in program design logic, leading to actionable recommendations; 3) and providing a unique perspective that supplements other research methods.

Identification of discrepancies in implementation allows for unanticipated findings. Documenting implementation details such as POP placement and field staff activities helps in examining whether program activities are being carried out as expected, and their effectiveness. For example, the Energy Trust case study found that although POP placement for qualified products was accurate, some program materials could be difficult to find among the vast array of promotional materials. Direct observational methods easily uncover these types of findings, which may not be anticipated by program staff.

Identification of logic breakdowns in design, leading to actionable recommendations. The case studies provide examples of how observational methods can identify logic breakdowns in program design assumptions. By observing sales associate and customer interactions first hand, evaluators can understand how the program works in the field, uncovering gaps in training, sales associate attitudes, and behaviors from program’s assumptions. For example, The Energy Trust case study found the lighting promotion strategies did not reflect the reality that many stores typically did not have dedicated lighting staff and lighting POP was insufficient to inform the
appropriate audience. Additionally, both case studies found that although sales associates’ held positive attitudes towards energy efficiency, in real-life scenarios they often did not promote energy efficiency in their interactions with customers. These findings ran against the program logic assumptions that staff trainings and provisions of program POP would sufficiently prompt sales associates to promote energy efficient choices to customers. Uncovering these findings enabled evaluators to make practical recommendations regarding the content of POP materials. In the NEEA case study, data from store manager surveys and mystery shopping suggested a need for more sales associate training. However, with information collected from the ride-alongs, evaluators found that the amount of time field staff have to talk with sales associates was limited, and therefore training by field staff would have been impractical. Instead, this context provided constructive recommendations about how the program might be able to supplement online training already used by retailers.

Provides a unique perspective that supplements other research methods. Observational methods such as mystery shopping and ride-alongs are one of many techniques that evaluators can use to assess energy efficiency programs, and evaluators should integrate these findings with other forms of measurement to provide the most value. Mystery shopping and ride-along observations can complement each other to provide a more detailed and unique view of the retail environment, and bolster the confidence of the findings. While mystery shopping provides direct insights into customers’ experiences and interactions with sales associates, ride-alongs provide information about implementation of in-store interventions. Additionally, observational methods can provide insights into program processes when other data collection methods are impractical.

The case studies provide examples of programs with midstream incentive elements, where lists of customers were not readily available because customers did not directly receive incentives from the program. Therefore, customer surveys would have not been possible.

Challenges and Tips for Using Mystery Shopping and Ride-Along Data in Process Evaluations

Despite the value that mystery shopping and ride-alongs can add to energy efficiency program evaluations, there are a number of limitations and considerations. These issues center on methodology, observer training, and implicit biases. The sections below describe key challenges to conducting mystery shopping and ride-alongs, and strategies to overcome these challenges.

Methodology

Observational data collection requires thoughtful methodological considerations, like any other method. The section below describes insights for developing an effective sampling design, writing a relevant data collection instrument, and documenting important process details.

Due to the amount of time involved in observational data collection, evaluators often use a purposive, rather than a representative, sampling method. Purposive sampling can allow evaluators to gain meaningful insights from a smaller sample, which can limit costs. By carefully

---

2 While cost is an additional aspect to consider, the difference compared to other methods is dependent on a number of variables and could be a challenge or benefit based on the scope of the research.

3 Purposive sampling is a non-probability sampling technique in which researchers strategically select information-rich cases to study. (Patton 2015)
selecting observations to obtain the greatest diversity, or to focus on cases of particular interest to the program, purposive samples can provide an informative view of the program, although they may not lend themselves to statistical generalizations.

In the Energy Trust case study, for example, mystery shopping visits focused on retailers that accounted for a large portion of program savings, and mystery shoppers visited multiple store locations for each retailer. Other factors the sampling strategy considered included whether a store’s location was in the Portland area and the number of field staff visits the store had received. While this was not a completely representative sample, it was purposeful and well balanced given the goals of the evaluation and the practical constraints.

Ride-along visits in both case studies also sought to assess the representativeness of the store visits researchers observed to help evaluators avoid putting undue weight on anomalous observations in their analysis. After each visit, the researcher asked the field staff member how the observed visit compared to a typical visit to that store and stores in that chain in order to identify any elements that may not have been representative. Researchers also asked field staff questions to obtain contextual information about the store visits they observed, including the number of times they had visited the store previously, both in general and in the past year, and how long it had been since their most recent visit.

During both mystery shopping visits and ride-alongs, it is important for observers to ask consistent and salient questions. A thoughtfully developed data collection guide both helps standardize the data collection process and allows for flexibility in the field. The guide should highlight key topics observers should be listening and looking for and develop contingencies based on foreseeable scenarios. Additionally, a guide should provide reasonable parameters of what the mystery shopper is requesting and consider what a typical customer would know about the product. For example, when shopping for a refrigerator or television, most customers would have basic preferences – size, type, or price range.

While many mystery shopping activities do not require observers to debrief associates, the authors have found doing so to be valuable. During the debriefing process, mystery shoppers can obtain further information from the sales associate that can be valuable to an evaluation. During this process, it is important to assure the sales associates that information provided during the visit is confidential and that the purpose of the research is not to evaluate their individual performance.

To maintain data quality, observers should document visits in detail as soon as they are able. This is especially important for mystery shopping because, in their role as shoppers, it is not possible for researchers to write down everything that sales associates say. The authors recommend that observers have a notebook with a few key reminders where they can take notes to help document details of their interactions immediately after the store visit.

**Observer Training**

Standardization of observations stems from good instruments and methodology, but additional training beyond how to develop and use guides is also important. The non-standardized nature of qualitative inquiry means the quality of data is more dependent on the experience and training of the observer, therefore it is important to invest in observer training. This training should include emphasis on a number of key topics, including:

- Learning what to pay attention to, and how to separate out important details
• How to write descriptively
• Discipline in recording field notes
• Acknowledging strengths and limitations of one’s own perspective, assumptions and biases (Patton 2014)
• How to act as a customer

Evaluators should coach observers to act naturally. During mystery shopping, observers should act in a way that does not arouse suspicion in the sales associate. During a ride-along, observers should minimize the impact they have on the observed interactions. Researchers may have opportunities to ask minimal questions during ride-alongs while the field staff are doing aspects of the work that do not involve interaction with others. However, the researcher’s primary role is as an observer, so it is important not to interrupt field staff members’ natural behaviors, as much as possible.

Implicit Biases

Due to the qualitative nature of observational research, evaluators should be aware of some inherent biases. Observers’ subjectivity may unconsciously effect the behavior of observed individuals or the quality of data collected. However, developing a strategic methodology and conducting observer training can help maintain consistency across visits and observers, thus mitigating observer bias.

Potential biases also exist from the perspective of the observed individuals. Since these data collection activities are in person, there are biases often not seen in other data collection activities. For example, the mystery shoppers’ gender, race, age, or dress could influence staff interactions with them. For the ride-alongs, the in-person nature of the activity may increase social desirability bias, by making field staff act in ways they perceive they are expected to act or respond, rather than how they normally act. While it may be difficult to combat these biases, acknowledging them during analysis can strengthen an evaluation.

Conclusions

Evaluations of energy efficiency programs would benefit from a greater integration of observational data collection activities in their process evaluations. These methods have been widely used in traditional market research for many years, but are often underutilized in the evaluation of energy efficiency programs. These methods are particularly useful for evaluating midstream incentive programs or other programs where a list of end-users would be difficult to obtain. Although these techniques come with certain challenges and limitations, they provide a valuable alternative approach to developing an enriched perspective to questions being evaluated. These methods can provide a level of detail and context that customer or sales associate surveys likely would not reveal. Mystery shopping and ride-alongs not only provide greater depth, but also can be used together to triangulate and support other data sources to provide for a more rigorous evaluation.
References


