Show Me Your Moves: Making Business as Usual the Focus of Large Customer (C/I&I) Self-Report Surveys

Marc Collins and Jeremy Eddy, Itron, Inc.

ABSTRACT

It goes without saying that separating the impacts of utility energy efficiency programs from the background adoption rates for those same technologies is a deeply challenging prospect. Given the relative absence of good quality, full-scale and longitudinal market characterization data relevant to many programs, it is still necessary to rely upon customer self-reporting techniques for at least a significant part of a net-to-gross assessment. However, there is important room for innovation to deploy self-reporting surveys that avoid many of the well-documented pitfalls and biases of these common surveys.

The authors will discuss a new technique developed in the context of a complex program serving large customers where multi-faceted and multi-party decision making is commonplace and usually recognized in the program design and program theory. The approach emphasizes and prioritizes unprompted responses from respondents where possible and puts emphasis on fully characterizing typical conditions as a point of comparison with program participation.

Noteworthy steps include avoiding the use of judgment-based scalar responses and sticking to fact-based questions wherever possible, deploying an adaptive survey structure that respects respondents' priorities and entirely skips less or irrelevant sections, tying core questions directly to the program's logic model, and, ensuring the granularity of the scoring algorithm is appropriate to the level of confidence in the information being collected.

The authors frame the new approaches within existing regulatory and professional guidance regarding best practices for self-reporting surveys.

Introduction

Assessing energy program attribution is an exercise fraught with complexities and scope issues at the best of times. When the efficiency program serves large customers with large-scale incentives and larger projects that are typically part of multi-faceted, multi-purpose and longer-term internal corporate "projects," things can get hairy. Add in that these complex scenarios rarely involve only a single decision-maker, but instead include more than one person from the facility itself plus often others from a head office or from corporate functions (or both), it is apparent that the current crop of self-reporting surveys get pushed to the limit.

With assistance from the Lawrence Berkeley National Laboratory,[1] a survey of a small collection of recognized experts in both the implementation and evaluation of continuous energy improvement programs was undertaken to assess the need for innovation in this area. The survey of hand-picked experts was hardly scientific or large enough for statistically-valid results, but areas of consensus or unanimity are worth reporting here.

[1] Aimee McKane, Deputy Group Leader, Industrial Energy Efficiency, Environmental Energy Technologies Division

After confirming the depth of their professional expertise, respondents were asked to comment on the challenges involved in evaluating continuous improvement program-generated savings. Responses confirmed the integral points made below, so will not be repeated. When asked if current evaluations and evaluation methodologies successfully address these challenges, the unanimous response was "no." In some cases, jurisdictions haven't even made an attempt yet, while in other situations evaluation parameters are too narrow (timing, focus only on specific equipment, definition of what constitutes a free rider or appropriate program attribution) leading to the perception that true or full savings are not being properly captured.

Lastly, respondents were asked the key questions, "Would you support further methodology research and testing to try to improve evaluation best practices in this area? Is more methodological work needed to adequately capture the savings from continuous energy management improvement programs aimed at large industrial and commercial customers?" The unanimous responses were "yes" and "yes."

With the need for new research confirmed, what follows is an initial contribution of a design of one new approach that purposefully attempts only incremental change from current self-report survey practices. Although more radical models are available, the authors perceive that regulatory comfort and approval will be more likely with a less dramatic departure from prevalent existing practices. Those tried and tested surveys work well for single-point consumer purchase decisions or situations where "packaged" installations with a limited range of defined choices exist, but can have more difficulty dealing with wide-open, ill-defined and sometimes very convoluted custom projects.

This paper discusses an attempt to focus on attribution analysis of projects at the unwieldy end of the spectrum because many efficiency program portfolios contain programs that serve these large customers and a significant proportion of the total portfolio savings can come from a few very large projects. The current prevalent current survey approach has been refined and adapted in multiple cycles over the past years and will continue to evolve and improve with more use and attention, including for this most complex of customer class. However, in this paper we describe an attempt to develop and test an alternative approach that may have some advantages over the existing prevalent approach. The design philosophy when developing this approach was to be open to rethinking the survey methodology approach to the extent necessary to be maximally respectful of the "custom" environment, but to live within the restriction that a consistent and systematic survey would be more likely to appeal to utility and regulatory decision-makers.

It was determined the design should also remain within the boundaries of a replicable survey, as opposed to primarily open-ended in-depth interviews. A very wide range of participant decision-making scenarios needed to be accommodated (the complexity). A full net-to-gross analysis should be accommodated, not just free ridership. Spillover analysis with a high comparative degree of rigor is now often required for programs with large-scale efficiency projects. Lastly, the attribution analysis technique should complement market effects analysis that may also occur in parallel.

Another major challenge for attribution analysis in this context is that efficiency is a relatively low priority for many industrial companies. Most industrial facilities rank energy efficiency below safety, meeting environmental regulations, and productivity. Industrial customers are, not unexpectedly, very risk-averse and will not make changes to existing processes if these could jeopardize the flow of operations. On the other hand, they are already used to measuring everything, and they're used to seeing copious amounts of data. Large

commercial customers may be less extreme in terms of these characteristics than their industrial counterparts, but certainly share some of the same reservations. In addition, a deep understanding of the equipment at a facility may only be held by a few people, some with deep-seated resistance to change on the grounds of not wanting to risk a disruption to productivity. It is therefore critical to frame survey design with this total context in mind—that is, the program influence element that may be central to the purpose of the evaluation is nevertheless not likely to be central to the decision-making respondent since the entire energy efficiency aspect itself, let alone the program-specific angle, is likely to be a secondary consideration.

Due to the wide range of variables often at play, especially in industrial environments, energy savings can be quite irregular in large custom projects. Energy efficiency is most likely a business decision, competing for time, attention and resources with production-driven and other critical issues. For that reason, programs often supplement a core package of incentives with various other types of information provision and technical support (on-site engineering analysis, expertise, audits, etc.) that contribute to risk mitigation (of disrupting production) or allow options to be considered that otherwise wouldn't "make the cut." Capturing this potential program influence set in a self-reporting survey is challenging due to both the nature of the influence (somewhat indirect) and because of the substantial permutations possible at participant sites/situations.

Timing is also critical. If a customer is a repeat participant, they may be more likely to self-describe as a free rider for their second (or later) program-funded project if a consistent definition of free rider is utilized. But, repeat participation is encouraged in strategic energy management (SEM) initiatives like Superior Energy Performance (SEP), ISO 50001 and other continuous improvement energy management programs under the implicit assumption that either something new or improved is occurring on each repeat round of participation across program years. This suggests a moving bar or threshold for program influence and free ridership. Less overall effort is required to move the energy efficiency needle for a first-time, less sophisticated participant from zero than from say an eight to a nine for a repeat, more sophisticated customer that is already well aware of their options and has ongoing energy management activities as part of their business-as-usual operations. However, there is an argument that says a program should be given at least as much credit, if not more, if it manages to push (or pull) a highly functioning participant even higher up the ladder than for moving a laggard from the ground to the bottom rung. Traditional characterization of free ridership might often result in overly characterizing these high functioning participants as free riders since they are already aware of and taking premeditated action on energy efficiency opportunities.

The final straw can be that a custom program has economic development elements. In these cases, overall energy reductions are not the goal, but rather improvements in energy intensity per unit of production (or data processing or other appropriate metric). If overall energy use goes up due to greater competitiveness and strength of the participants, as long as energy intensity was also improved, success would be declared. This orientation is not unexpected for jurisdictions with mature, but often stressed or declining, industrial bases. Thus continuous improvement in energy-using processes and energy management sophistication needed to be measured along with other more traditional energy conservation measures. These management items, by definition, do not involve point-in-time decision making. Not only do they not fit neatly into traditional influence models and constructs, they are more likely to confound them given that premeditated thought and action in advance of program-related activity is both a

requirement of strategic energy management and simultaneously the classic definition of a free rider.

The conundrums presented in the confluence of all of these elements pointed to the need for an approach that would address the specific requirements at hand rather than attempt to add to or edit existing survey tools. This was done both for an intellectual rationale and due to the very real lack of available "survey real estate" in the prevalent standard survey. When inside and outside spillover batteries are added to the already extensive free ridership components, the survey length is at the limits of tolerability, or according to some, past those limits. Regardless, adding more components to the existing survey could not be entertained and a selective pruning and replacing would likely result in an awkward hybrid that also had the potential to compromise the integrity of the current survey's scoring algorithms.

Some of the key elements of the new approach (that will be elaborated upon in further sections of the paper) include: i) attribution-specific baseline development that arises from an emphasis on the past typical, or business-as-usual decision-making and management context; ii) program-induced influence grounded in the difference between the attribution-specific baseline and what happened/what changed during the program-sponsored project; iii) the use of an "adaptive" survey construction where key elements of the survey itself are molded by respondent input, and; iv) precision of scoring algorithm options that are aligned with the confidence of the input and analysis.

Another key element that permeates the whole exercise is an attempt to use fact/evidence-based queries whenever possible and to avoid hypothetical, opinion and judgment-based ones wherever possible. Essentially this means gathering "evidence" of norms and values during this initial mini market characterization section—before the respondent is even aware that program attribution is a topic of interest for the interview—that will be held up to the respondent later when evidence related to the program-sponsored project is the central topic. First, a contrast will be drawn between the historical norms and whatever change was necessary to accomplish the program-sponsored project. When later queried about possible program influence the respondent will have, unwittingly, already provided the vast majority of the evidence relevant to an understanding of whether anything substantive was different for and during the project versus BAU and will have a difficult time purposefully telling a tale regarding program influence. These tales and tendencies may exaggerate or minimize program influence based on a number of well-documented biases inherent in self-reporting surveys.

SEM-orientated programs try to instill a philosophy of continuous improvement and therefore present a dynamic, ever-changing "baseline" to evaluate. For instance, installing energy monitoring equipment on a process or major piece of energy-using equipment could be part of an efficiency program project. If afterwards, that monitoring reveals an opportunity that should be pursued, and the participant goes on to act, this might be construed as a classic example of spillover. But what if they applied to the program to do this action as a "phase 2" of the original project? Even if a standard eligible measure for others in the Program, due to the expectation of continuous improvement, it may now be deemed as something that should occur as standard practice and that program incentives and support should be reserved for measures above and beyond that example. This continuous raising of the bar is both the fundamental purpose of SEM-promoting programs—the rising tide lifts all boats concept—and a pronounced point of contention for attribution assessment. It provides a vivid example of why a thorough understanding of the context, or the "starting point," and the background to an efficiency program project is so critical to the veracity of an attribution assessment.

This approach also differs from the traditional free ridership strategy in that it does not directly compare program-related decision-making factors to non-program factors. As described earlier, the complexity of the facility and decision-making environment and the likelihood of a long-term relationship with the utility means that it is often incredibly difficult to distinguish decision-making factors according to program versus non-program influence for this market sector. For example, internal policy changes relating to energy efficiency may have been influenced by earlier interactions with the utility, yet the participant today may not recognize that—even if they are aware of it intellectually. Respondents should not be expected to think about decision-making and influences through the somewhat arcane prism of our program and evaluation-industry lens when these elements are real and tangible to them, but in other ways relevant to their everyday business. A simple analogy is experimental drug testing. Participants are only asked by researchers about the effect of the drug on their experience and health, before and after trying it, not whether they think the particular drug is worthwhile relative to other alternatives, most of which they wouldn't know about, or hypothetically what they think their response would have been had they not taken the medication. Since there are so many potential gray areas, the concept here is to avoid forcing a "scales of justice" analogy that pits the known versus the hypothetical. Examples could be the effects (or lack thereof) of the administration of a drug noticeable to its taker versus the potential effects of a range of alternate therapies, or, efficiency program influences versus a potentially very broad range of non-program influences.

In the sections that follow, we walk through both the conceptual reasoning and practical question structure for the elements in this approach. The final part of the paper will deal with the pilot and testing of this approach.

Attribution-Specific Baseline

The key mindset shift imbued in this alternate approach to attribution may be best characterized in terms of the development of the attribution-specific baseline. In effect, it suggests that it is not enough to just acknowledge the wide variety of contexts and situational differences inherent in the participant population and their various rationales for engaging in the program. That background is usually treated as an interesting and sometimes informative context for the central task of probing respondents' balancing of program- versus non-program-related influences. In this approach, the background and foreground get flipped in terms of importance and perspective. The background becomes the story in that by determining a credible business-as-usual (BAU) scenario, it is possible to both contrast what happened differently during the program participation episode/period and ground the extent of any reported influence within a context of a characterization of the respondent themselves and where on the "ladder" of energy efficiency sophistication is the starting point (e.g., are they already a sophisticated strategic energy management practitioner or more of a laggard relative to other companies in their industry?).

By accepting that complexity (of decision-making styles and other contexts) is the norm in this customer class, we sought to recognize that the combination of greater business sophistication and the general relative lack of emphasis towards energy management and energy efficiency relative to core business priorities represent an opportunity. Specifically, it suggests that the background or context could be very informative in terms of both the BAU baseline and as a method to ground the respondent in their own stated BAU context when probing about program influences. The survey could then avoid the generic and the hypothetical, focusing instead on the unique before and after scenario of each respondent.

A two-stage process is required to pursue this approach. The first stage involves determining a facility-level BAU characterization—in a neutral manner divorced from the program. This BAU context may already be affected by a long-term relationship with the utility, which is another potentially confounding variable when using a traditional survey approach. The second stage involves determining a project-specific characterization that will illustrate differences from BAU that are potential program influences. If there are no consequential differences between the historical BAU characterization and the project characterization then the participant has a predisposition towards free ridership.

Information in three key areas is collected to develop a facility-level characterization or baseline. The first of those areas is Market Barriers—those things that generally prevent the facility or firm from pursuing higher degrees of energy efficiency or greater degrees of strategic energy management. Second is Decision-Making— a characterization of all the corporate policies and practices, especially related to investments, daily routine habits, formal committees or management chain of command that drive decision-making in general and specifically, if there is any unique approach, towards energy use. Last, energy management practices—which could range from almost nothing to a plethora of monitoring and management of equipment, processes, facilities, decisions, and policies towards energy use that variously could be voluntary or mandatory.

The purpose is to determine the BAU decision making approach for projects that would be of a similar type and scope to those in the program. This is done with little to no reference to the program or the program administrator or sponsor. The goal is to get as comprehensive as possible documentation of the pre-program BAU situation and history. Since it is possible for current program participants to have participated in past DSM programs offered by their utility, that utility relationship needs to be clear because influence from prior programs and contact with the utility could have influenced both their BAU situation and the current participation under review. This is also a customer class that typically has utility account representatives (or equivalents) who are in regular contact. For that reason, the opening section of the survey begins with an examination of that prior utility relationship and history. A gauge of the length and depth (or seriousness) of that pre-existing relationship is important context. As such, its self-reported assessment is collected in a neutral manner well before any program details are discussed. Some evidence of the history may be available in a program and/or contacts tracking database, but it is still important to collect the respondent's direct impressions and feedback.

The three core areas of baseline information (Figure 1), within the context of the pre-existing relationship with the utility, form the basis of a "mini" market characterization and are designed to dovetail with the more comprehensive market characterizations that occur separately from this evaluation. If planned and executed in a coordinated and optimal manner, the mini market characterization should occur on a cyclical basis and can therefore provide an alternate year/cycle opportunity to reinforce the main ongoing market characterization process. Market effects assessment relies upon multiple and consistent data points over time for its usefulness, so the opportunity for the program attribution assessment exercise to contribute to the exercise is clear. Conversely, program-specific attribution assessment, often forced to rely heavily on self-reporting surveys, can benefit from better market characterization, especially if data points already exist/existed before the assessment takes place. The three core areas are elaborated next.

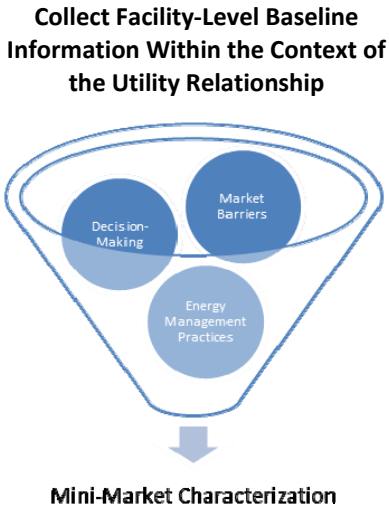

Figure 1. Initial characterization data collection.

Market Barriers: Large customer programs are often planned to address a wide and extensive variety of market barriers to energy efficiency. For this section of the survey, a list of known and expected barriers may be articulated in the program's logic model or could be developed from a wellspring of long-term experience and awareness of efficiency program best practices. This section of the survey is designed to determine the respondent's market barriers that exist and that would affect a project of a similar type and scope as the program-sponsored project.

Asking respondents to self-identify, unprompted, the key market barriers to additional investment in energy efficiency measures and actions that have traditionally affected their facility/firm—prior to their involvement with the program—provides the following: 1) market characterization information about the prevalence of various market barriers on a general level; 2) identification of the barriers that are known, possibly seen as intractable and most relevant to the respondent; 3) (potentially) confirmation whether program-identified (in the logic model most likely) barriers are a true reflection of participant's perspectives, and; 4) the foundation that will be needed later in the survey to determine if the program had a significant impact on one or more of the market barriers that are important to the respondent (without requiring the respondent to rank or rate them).

Interviewers would have a list of barriers such as the following: lack of time and competing priorities; volatility and risk related to energy prices and the business environment; lack of information to support EE investment; risk that EE-driven process changes will negatively impact production (rates, quality, reliability) or tenant comfort; concern for confidentiality of process-related proprietary information or intellectual property; lack of funding to support detailed engineering analysis; competing needs for capital and other resources; lack of awareness, knowledge and understanding of EE features/benefits; uncertainty about savings; high incremental or first costs.

Nevertheless, the preference is to seek unprompted feedback from respondents in this section and also in the other areas of the mini market characterization. The rationale for this is two-fold. First, from a research perspective unprompted feedback is simply worth much more than prompted. Unprompted identification of market barriers, as in this case, indicates that not only does the respondent correctly understand the concept that is being discussed, but they are also self-aware of something that is more likely to be related to program influence assuming their suggested barrier is one that the program explicitly attempts to address. For example, if "uncertainty about savings" has been an historical barrier in that facility or firm, there is an obvious need to (later in the survey) connect the dots on whether those elements of the program that address uncertainty of savings were utilized, relevant, satisfactory, etc.

In addition to unprompted feedback being more visceral and tangible, this part of the process feeds the adaptive aspect of the survey construction (described in detail in a later section). The self-identified priorities of the respondent unearthed during this purely exploratory characterization component of the interview are used to narrow and focus the program influence-related questions that will come later. The interviewer will be able to use prompts, e.g., from a list like the one above, but only needs to do so if the respondent is unable to produce their own relevant feedback.

Decision-Making: The core purpose of most large-customer efficiency programs is to influence decision-making positively toward higher/better efficiency options. Inevitably, this occurs within the somewhat limited reality that energy use is often not a top priority for participating customers, and so by extension neither is energy efficiency. Asking respondents (again, mostly unprompted) to self-identify and describe decision-making protocols, structures, investment thresholds and practices for capital projects in general in their facility—and then differentiating for energy efficiency-related projects—provides the following: 1) market characterization information about capital project investment practices and policies and decision-making in general; 2) the relative importance of energy efficiency-related investments and how they are treated compared to other capital and resource priorities, including who makes the decisions, why, and where are the decision-makers (if they are subject to broader corporate/head-office influence outside the facility itself), and; 3) the foundation needed later in the survey to determine if the program-sponsored project "moved the needle" in terms of, e.g., pushing an investment into acceptable territory, or, possibly, changed investment review practices themselves.

As with market barriers, interviewers are equipped with lists of archetypal examples related to investment decision-making, but the preference is to let respondents guide the way. If necessary, they can be prompted related to the investment decision-making process regarding: the replacement of old or outdated equipment on some sort of schedule; how much production needs drive choices; planned remodeling, build-outs, or expansions; available program or tax incentives; protecting the environment; reducing energy costs; reducing power outages; updating to the latest technology and so forth. External (competitive environment; economic cycle; safety and environmental regulations; etc.) factors that might influence decisions are also available as are internal (inter-plant competition; capital investment policies; technology policies; facility planning time scale; etc.) factors and a prompt to think about any recent issues, if any. There is also interest in how energy plans are generated in the facility/company should they exist.

Energy Management Practices: Improving strategic energy management (SEM) practices (both management-related and technical ones) is often a long-term, market transformational goal of large-customer efficiency programs. A challenge with respect to attribution analysis is that facilities and firms will have a wide variety of "starting points" regarding their strategic energy management "position" or sophistication. As discussed above, attribution assessment must be careful not to misconstrue a sophisticated participant as a free rider just because they were already aware of and paying careful attention to energy issues in their facilities. Asking respondents to self-identify the extent to which they were already deploying various SEM best practices provides the following: 1) market characterization information about SEM practices in general; 2) the relative "status" of the participating facility/firm, pre-program, on the "market adoption curve" of SEM ("laggard" at one extreme and ISO 50001/Superior Energy Performance-certifiable at the other), and; 3) the foundation

needed later in the survey to determine if the program helped to "move the needle" in terms of a facility moving up the adoption curve or "ladder" of SEM best practices.

There are two categories of practices, management best practices (MBP) and technical best practices (TBP) that are treated differently. MBP are a fairly well-defined universe, so they are prompted because it is important to build a characterization for the individual case and to contribute to the wider utility- or state-wide picture in terms of: a) documented corporate energy policies; b) specific energy plans; c) energy as a discrete element in financing decisions; d) presence of staff with responsibility for energy and that have accountability in the organization; e) project management and implementation; f) energy management information system(s) (EMIS); g) energy-related training and capacity development, and; h) formal reporting and communication.

Technical Best Practices (TBP) are comprised of a much more fluid list, particularly in the industrial sector where there are so many different equipment and process variations. Here, the process is to get input from the respondent as to the technical specifics, but also to examine how, when and where learning about and applying TBP occurs.

Knowing the "starting point" in terms of SEM is critically important context and input for attribution for several reasons. Unlike a traditional energy conservation measure that can be installed at a point in time and then potentially forgotten, SEM is a philosophy and way of doing business that permeates decision-making and affects (or should affect) all other energy conservation measures.

Program-Induced Influence

Once the pre-program era, project-specific, attribution-related baseline characterization is complete, the next step is to make a second pass through the same research territory, but this time aimed at assessing if the decision-making related to the program-sponsored project differed from the "norms" that the respondent had just provided in the prior section of the survey—and if so, how. It involves utilizing the information provided in the initial characterization to adapt a series of "question branches." Once again, the queries can remain primarily evidence-based because the key indicators in each of the three areas (described above) have already been self-identified in the first two steps.

Since respondents are more likely to self-identify factors in each of the three topic areas that they thought were most important or relevant, there is no need to move though a pre-set list of factors except in the occasion that the respondent is unable to think of anything relevant. This is unlikely because, the respondent has already been prepared to opine on these same topics based on the first section of the survey. By relying on simple, factual, verifiable and likely to be documented (somewhere) variances from BAU allows for a parallel characterization of the program-sponsored scenario that can be compared with the BAU characterization gathered in the earlier part of the survey. The respondent is not required to postulate regarding what might have happened in the absence of the Program, even though that is the core question for evaluators.

At this point in the process, the interviewer and respondent should have a shared understanding of how and where the program-sponsored project deviated from standard practice for the facility (and the issue of when this occurred is already known). Focusing on the possible variance between BAU and the project-specific baseline improves the ability to identify program impacts because there is a significant likelihood that some of the issues will be directly related to the program or program requirements.

This “second pass” section of the survey probes for any significant reasons for the deviation using primarily open-ended and unprompted mechanisms, but with the aid of reference to common and plausible scenarios if the respondent is unable to articulate a response.

The tone of the approach is intended to be extremely respectful of the customer and their unique context. It should be perceived as a conversation, not an interrogation. Despite that, a helpful analogy is to imagine interviewers akin to investigators gathering evidence until a threshold is reached that causes a suspect to be released or charged. But to extend the analogy, the system is based on a “guilty until proven innocent” premise, or in this case, the way this approach is structured, there is effectively a default assumption of no program influence until evidence is presented to counter that status.

It is readily possible to gather conclusive evidence that little or nothing was done (i.e., there was little to no deviation from BAU caused by the program) because the explicit, intentional “actions” of the program are known and by definition limited in range. If nothing that the program offered stimulated any noticeable movement, the participant is presumably a free rider.

It is not possible to be so definitive about which influences caused certain actions. Influences that are traditionally considered “program influences” or “non-program influences” likely intermingle. The approach described here moves away from seeking to score and weight a pre-determined set of program and non-program influences and toward a model of putting the onus on the respondent to demonstrate their unique actions surpassed a reasonable threshold.

To address this conundrum, we return to the threshold concept. How this is implemented in a pragmatic manner is the subject of the next section.

Adaptive Survey Construction

The gathering of, or more precisely, the preponderance of evidence-based approach described in the prior section can be illustrated with a relatively simple-to-understand utilization of the Ishikawa (or Fishbone) Diagram (Ishikawa, 1976). Use of this diagram is often considered the best of the seven core Japanese quality management Kaizen tools for cause-and-effect analysis to avoid missing the real problem while being distracted by symptoms. It also has the advantage that it may be familiar to many industrial-class respondents because it is used in the "analyze" phase of SixSigma's DMAIC (define, measure, analyze, improve, control) approach to problem solving.

The diagram below uses only two levels of analysis for illustration purposes. More could be used as necessary. The top level consists of the three categories/branches of potential deviation from company norms. In this particular sample/example, there are three items the respondent identified in the market barriers category that further probing confirmed could be considered significant—and all three of them are items that in theory at least could be successfully addressed by the efficiency program offerings. Two changes in the decision-making category and three more in energy management practices makes a total of eight factors to be explored and probed by the interviewer. In each case, whether it is placing a higher priority on efficiency in the context of other, production-driven factors or modifying company policy to require high efficiency equipment for projects similar to the project at hand, an opportunity for program influence exists. Once opportunities have been defined, any specific activities or provisions related to the program, e.g., to address and overcome market barriers to move the company toward greater efficiency, can be assessed in detail.

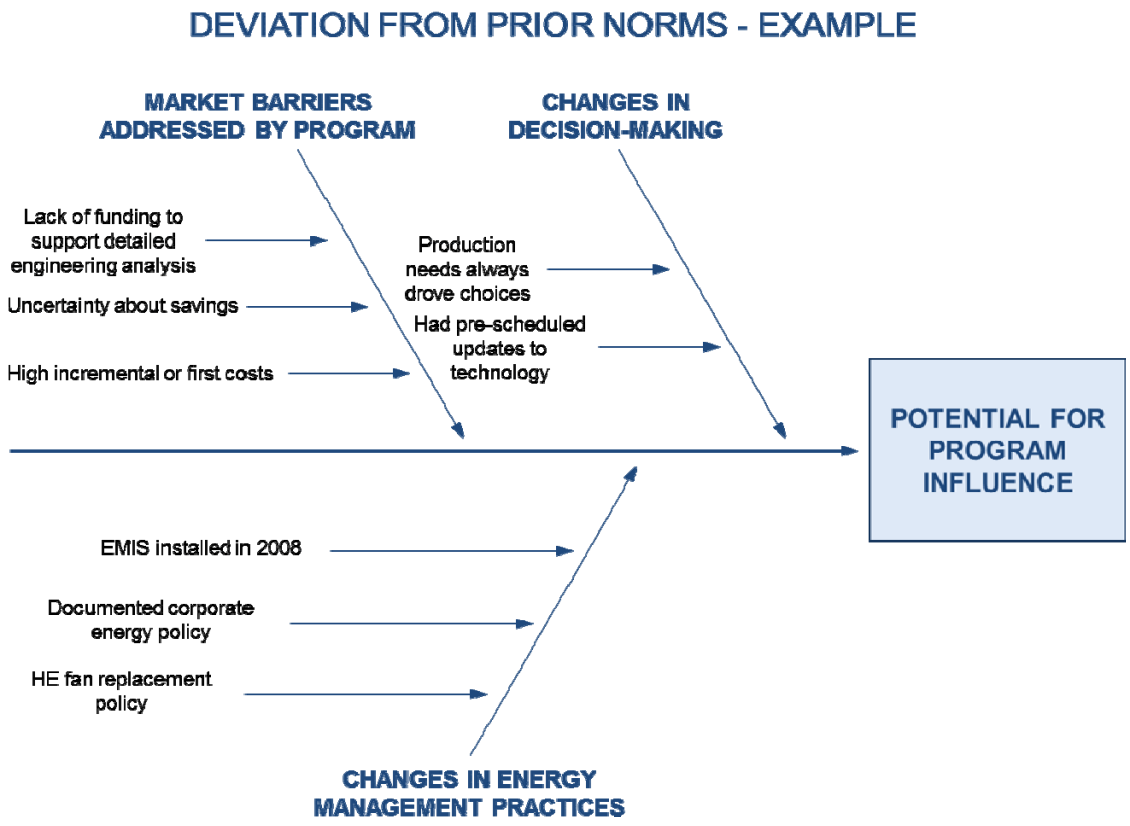

Figure 2. Illustrative Ishikawa (Fishbone) diagram of deviation from prior norms.

This deviation from norms diagram also illustrates the key adaptive approach to questioning. The eight factors illustrated here were self-identified by the respondent and selected because differences between the program-sponsored project and the BAU norm were indicated. The goal for the interviewer is to pursue evidence related to the most likely sources of program influence (here, eight of them) and attempt to gather compelling and conclusive evidence (of status of free rider, or not free rider) in each. There could of course be other factors that influenced the project, but the interviewer/investigator need not be exhaustive or comprehensive. The need is simply to find some...enough…compelling evidence, but not try to define or collect all the evidence in the universe. The interviewer can provide a neutral receptacle for the respondent's feedback, with no need to apply pressure to rank or prioritize the evidence. Some evidence will naturally be stronger and clearer; other less so. In addition, there is a deliberate opportunity provided for the respondent to add one or more "Other" factors if appropriate. (By this time in the survey, the respondent's memory may have been jogged after passing through the complete characterization section and a new factor may spring to mind).

The selection of the factors themselves is "adaptive" in that only those items emerging from the specific interview process feed the creation of the interview structure. Then, each line of questioning within (eight in the provided example) is also adaptive in that it only requires pursuit until evidence of program influence is revealed. This may occur after only a first-level of questioning, or, it may take a second, third, or fourth level of ever-tightening questioning. The rationale behind this approach is three-fold. First, due to the respondent context-centered nature of the line of questioning, a gentle, open-ended approach can be taken at the outset that will often

produce sufficient evidence without need for, or the perception of, further interrogation. Second, the minimum number of questions overall are asked, serving to reduce survey fatigue. Third, time is saved for the less clear areas where further questioning would be necessary to make a declaration about program influence, one way or the other. Once each line of questions is exhausted, a score (weight) must be assigned for that item—those details are discussed in the next section.

Scoring Algorithm

The end result of pursuing all the lines of questioning is a series of status indicators (eight in the example), or scores. Best practices advice reminds us not to concoct arbitrary formulae or scoring models[2], although some form of scoring is beneficial/necessary for evaluation purposes. However, it is not reasonable to assume that the respondent's identified program influences are necessarily equivalent to each other in strength; nor could one branch of questioning be considered equivalent to another or all others for purposes of assigning scores or weights. It is also clear that the limbs of the branches themselves are subject to respondent selection bias—albeit purposeful, it still results in differing sets of limbs that make comparing one respondent to another inconsistent. Rather than the program versus non-program influences "scales of justice" analogy, the conceptual framework of the weighting and relationships of the eight example factors looks more like that shown in Figure 3.

Referring back to the top level, it is also possible to assign weights and values to the three main branches of the fishbone diagram/construct, but any elaborate scoring scheme would, again, have significant arbitrary aspects and could be perceived by many as a "black box." Instead, remaining true to the preponderance of evidence approach by assigning a score to the evidence at the end of each line of questioning and then asking the respondent to also assign their score fulfills two needs: a confirmation that the respondent and the interviewer did not misunderstand each other (reflected in similar or the same scores for each item), and, a weighted or relative value for each of the relevant issues that can later be plugged-into an algorithm(s).

[2] For example, two major states have specific self-report survey guidance. New York Evaluation Plan Guidance for EEPS Program Administrators (Appendix G: Guidelines for Estimating Net-To-Gross Ratios Using the Self-Report Approach, updated August 2013). California DSM Measurement Advisory Council, "Appendix J: Quality Assurance Guidelines For Statistical, Engineering, and Self-Report Methods for Estimating DSM Program Impacts."

It is possible to use a very simple scoring scheme such as the following[3]:

Zero – the program played no role in the deviation from BAU (therefore non-program-related factors were influential)

Low – the program played a likely minor role in the deviation from BAU

Medium – the program played a significant, supporting role in the deviation from BAU

Full – the deviation from BAU would not have occurred without the contribution from or fulfilling requirement(s) of the program

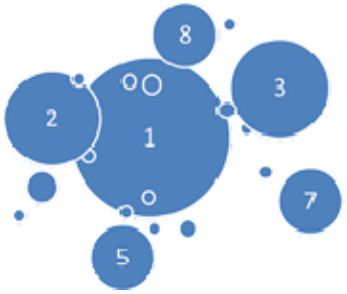

Figure 3. Conceptual framework of decision-making factors.

The selection of the number of response categories for scoring (four here) should be driven by the following question: how frequently would respondent and interviewer agree on the meaning of the category if asked (e.g., to defend their rating)? Four gradations seems reasonable if that test is applied.

Note that the simplicity of these ratings, including that all ratings have some degree of equivalency due to their assignment at the farthest/deepest level required, keeps open a variety of scoring algorithm options. For example, at one extreme, a project could either be rated as a free rider or not (i.e., a one-zero toggle), with the program's free ridership rate being derived from the average of the sum of the project scores. This is not as far-fetched as it may seem on the surface due to the often smaller number of participants in large-customer custom programs. A census for attribution purposes may be possible and could reinforce the validity of a binary scoring approach. Other partial free-ridership algorithms are easily feasible using the four categories. Multiple algorithms could be applied and sensitivity testing utilized to determine what is most appropriate for the particular jurisdiction and program circumstances. The important point is that this overall approach has options for calculating and reporting program influence at varying levels of granularity.

Like all others, this new survey approach cannot prevent someone from wanting to "manage" their responses to a certain end, but it can make that task extremely difficult when all of the history and context has already been presented and discussed by the time the respondent perspective is sought on the matter of influence. Typically, the interviewer should be able to correctly predict the respondent's suggested score regarding influence because sufficient specific evidence, set in context, will have been collected during the earlier sections of the survey and in the factor-specific line of questioning.

The authors look forward to pilot testing this approach soon and reporting on those results.

[3] The definition of "program" might be defined to include program administrator and past program influences, but the short-form is used here for simplicity's sake.

References

Ishikawa, Kaoru (1976). *Guide to Quality Control*. Asian Productivity Organization. ISBN 92-833-1036-5.