“Sure You Do. Uh-Huh”: Improving the Accuracy of Self-Reported Efficiency Actions

Marjorie R. McRae, Research Into Action, Inc.

ABSTRACT

We often attempt to estimate program impacts by asking participants and nonparticipants whether or not they have taken any actions from a list of energy efficiency steps. Unfortunately, respondents may recognize the social desirability of actions reflecting a concern for the environment, or of actions that are, in some way, “efficient.” Respondents may give inaccurate answers to portray themselves as making efficient choices about energy use. That is, their answers may be compromised by social desirability bias

During repeated evaluations of a multi-year program aimed at architects, we refined a series of questions that reduced social desirability bias and other measurement error. The structuring of the questions enabled respondents to both “save face” and give accurate answers. Using these questions, we were able to uncover differences between participants and nonparticipants.

This paper explores the issue of social desirability bias, how it was revealed in a multi-phase program evaluation, and how it was finally addressed in a fruitful study. It concludes that it is possible to discuss narrowly defined, specific actions with respondents and get a good characterization of their activities.

Introduction

We often attempt to estimate program impacts by asking participants and nonparticipants whether or not they have taken any actions from a list of energy efficiency steps. Often our questions rely on terms that have a clear meaning for us, but not necessarily for our respondents; for example, “energy-efficient.” Respondents may interpret such terms to be consistent with their actions, because such an interpretation coincides with what they believe to be true or wish to be true about the type of person they are. Even when our terms are well defined, respondents may not answer truthfully when they want their answers to reflect that they have a concern for the environment or some other pro-social attitude.

We were studying developers of commercial facilities in the Pacific Northwest to create a baseline for a market transformation program. We were concerned that our results suffered from social desirability bias.

When the majority of developers we interviewed said that their buildings incorporated a day lighting strategy, we found ourselves thinking, “Uh-huh, sure they do.” Our suspicions were increased when a few of the developers elaborated, “I always make sure there are lots of windows.”

As part of this same study, we interviewed architects. We hoped that the architects’ responses might contain less social desirability bias than the developers’ data, due to their greater technical knowledge of building features. Our hope received some corroboration by the quality of the architects’ responses to our open-ended questions. But we had no way assessing, for either developers or architects, if social desirability bias was, indeed, present.

For several years, we had been engaged in multi-year evaluations of two other market transformation programs. One of the programs educated architects in energy-efficient design through the use of awards for exemplary designs and through inter-active and traditional workshops. The other program educated building operators in energy-efficient building operations. For both programs, the tacks that we initially took to estimate program impacts (or to estimate simply whether the program was having an impact) were unsatisfactory for several reasons.

Estimating the impacts of educational programs can prove to be a harder nut to crack than for resource acquisition programs where specific installed measures can be counted or where billing analysis can detect impacts that are anticipated to be large. Educational forums that deliver information for several hours or more cover many efficiency concepts and techniques. Participants have varied backgrounds and learn different things from the same course material.

Participants themselves may have a hard time identifying whether they had "applied" what they had learned, or whether their educational experience simply became part of the many experiences out of which their actions arose. In the case of architects, it can be hard characterize "typical" practices, since the applicability of specific efficiency measures and approaches can vary greatly among projects.

Although the two program evaluations differed considerably, they both relied heavily on modified "case study" approaches to estimating the presence and magnitude of a program impact. For both programs, we looked for participants who said that they had applied concepts they learned in the program and had saved energy. We asked these participants what they had done and tried to assign impacts to their actions. We had difficulty obtaining examples specific enough to support the attribution of impacts. And yet, most of the participants reported that *they* thought their behaviors had changed.

Our early assessments of these programs left us dissatisfied with our ability to characterize the programs' impacts or even whether they were having an impact. The potential for social desirability bias was one of our concerns.

Fortunately, the multi-year evaluation plans for both education programs gave us the time and experience necessary to develop more satisfactory evaluation approaches. We believe that our final evaluation designs were less subject to social desirability bias.

Social Desirability Bias and Measurement Error

Social desirability bias occurs when respondents portray themselves in a good light, responding in a way that conforms to the expectations of the group to which the respondent belongs or identifies (Dillman 1978). The survey research literature typically illustrates the concept by exploring question design for topics that respondents may find highly embarrassing or sensitive, or for illegal activities. Common examples include mental illness, sexually transmitted diseases, and alcoholism (Fowler 1993). However, as Dillman (2000) points out, "even very ordinary questions that seem, on the surface, to have little social desirability" can generate inaccurate answers. Dillman continues, "Social desirability operates at a threshold far below what one thinks of as anti-social behavior."

Social desirability bias can occur in data when the possible responses to a question vary in their connotations to the respondents. When some response options carry positive connotations and other carry either neutral or negative connotations, respondents may be

inclined to give the response that they feel has positive connotations or that they assume has such connotations for the interviewer.

Fowler (1995) discusses the need in survey design to manage the meaning of answers to questions. Questions and response categories need to be viewed from the respondent's perspective. Respondents are concerned about how their answers will be interpreted and what their answers may mean about the type of person they are. "A goal is to permit respondents to present themselves in a positive way at the same time they provide the information needed" (Fowler 1995).

People give socially desirable responses because, at a minimum, we want to avoid experiences of others thinking badly of us. In addition, we function with some degree of confidence by holding positive images of ourselves. If we fully understood the consequences of our actions, there would be are many actions we would change, with examples ranging from our use of nonrenewable natural resources to the frequency with which we get angry to how we spend our time. Yet we cannot change many actions at once, so our awareness functions as a filter. Change occurs as we become aware of some of the undesirable consequences of our actions, aware of alternative actions and their benefits, and motivated to embrace the alternatives (Miller & Rollnick 1991; Prochaska & DiClemente 1986).

When one asks developers, or architects, or building operators about their work behaviors, there is a natural inclination for these professionals to believe and to report that they follow the best practices of their professions. When some of the respondents think that a question explores their use of best practices, they may report taking actions that they do not take or they may interpret the question to include their behavior. That is, they may over-report desirable actions.

Only a portion of respondents will give the desirable response instead of a less desirable, more accurate response. Other respondents will respond accurately. Perhaps they don't perceive the desirability issue, or they have already come to terms with the fact that their actions fall short of their ideals, or they are not influenced by the assumed opinions of interviewers, even though they may care what their colleagues think. Finally, some respondents may under-report an action. Perhaps their assessment of their actions is erroneous or their cynicism or insecurity leads them to think that things are worse off than they are.

As true for other fields and what we might expect from human nature, research in energy efficiency behavior encounters more over-reporting of desirable actions than under-reporting (Peters 2002). The degree of social desirability bias can differ from question to question and from study to study.

A recent survey of appliance purchase behaviors in Wisconsin that we analyzed found that the proportion of respondents reporting that they had purchased high-efficiency appliances was nearly double the proportion that actually had purchased high-efficiency appliances (Energy Center of Wisconsin 2002). Ninety-one percent of respondents purchasing a major appliance in the previous twelve months reported that their appliance was high-efficiency. (Nine percent of respondents said their appliance was not high-efficiency; "don't know" responses are excluded from these percentages.) Among the respondents who were aware of the ENERGY STAR® logo and gave an adequate explanation of its meaning (a group one might expect to be well-informed about efficiency), 81% said that their appliance was ENERGY STAR® compliant. In contrast to these high percentages, we were able to verify that 44% of the purchased appliances were high-efficiency.

The over-reporting of high-efficiency among recent appliance purchasers in Wisconsin no doubt reflects some degree of social desirability bias. Other factors may also be at play. Respondents might genuinely have a misunderstanding of their specific appliance or of the appliances available in stores. For example, some respondents may believe that "the government has taken care of that" through the imposition of energy-efficiency standards that all appliances must meet. Similarly, some respondents may believe that "everything on the market today is high efficiency."

The over-reporting of socially desirable responses constitutes measurement error. What we would really like to measure are the characteristics of the buildings, or the actions, or the appliances. But it is prohibitively costly to go to a representative sample of buildings and count or analyze the efficiency measures incorporated; or to follow a representative sample of building operators around for six months and count the number of efficiency actions they take. So instead of direct observation, we need to rely on the reports of the market actors—their claims to have done something. It is important that the questions that we ask and the way we ask them be carefully thought through to minimize the likelihood of measurement error.

Measurement error is not limited to socially desirable responses. It covers all situations where one cannot, for whatever reason, accurately measure or ascribe a value to the concept of interest. One type of measurement error occurs when respondents do not understand the terms we use in our questions or when they interpret terms differently than we intended. In energy efficiency research, inaccurate or unclear terms often conspire with people's desire to portray themselves favorably. Thus, the large discrepancy between the proportion of Wisconsin respondents who said they had purchased high-efficiency appliances and the proportion verified to have purchased such appliances may owe, in part, to respondents' genuine confusion about just what constitutes a high-efficiency appliance (Energy Center of Wisconsin 2002).

Structuring Questions to Reduce Measurement Error

When we felt unsatisfied with our modified case study approach to determining whether the educational program for architects had impacts on participants' design practices, we set about to develop another approach.

For the architects' education program, we asked the educator to provide a list of specific actions, methods, and concepts covered in the training. We received a three-page, single-spaced, small-font list containing over 150 elements!

We arranged the elements by similarity and came up with 11 groups that covered most of the items. We developed an introductory question for each of these groups, asking respondents whether, in the past year, they had engaged in the general type of activity or not. For example, we asked, "In the past year, have you engaged in pre-design activities to address energy and resource savings project-wide" and "For any of the projects was the building's site or orientation selected because of solar access, shading or other resource considerations."

Under each of the 11 broad questions, we asked from two to six questions probing the specific actions taken. For example, under the broad question about pre-design activities, we asked: "Have you talked with the client about efficiency and resources in the pre-design stage," "Have you set energy efficiency goals or performance benchmarks," "Have you had

discussions with your in-house design team about how the designs under consideration affects the building's energy use," and "Have you had discussions with project consultants and contractors about the energy use implications of the alternatives under consideration". We concluded each set of questions with a request for "other," e.g., "please describe any other steps you took during the pre-design phase to address energy savings."

Most of the 150 elements listed by the educator were called out in the subquestions. Often, we grouped common elements together in a single question, such as this question following a general question on the use of a day lighting strategy: "Have you incorporated shading strategies such as louvers, projections, and light shelves." In total, we asked architects 70 questions about their design practices in the prior year.

We asked respondents two types of questions about the activities. The first question was Yes/No: have you done this on any project in the last year. The next question was the frequency with which they had taken the design action. Was it on one project out of the 20 they had worked on during the year, or six of eight projects? Through the frequency question, we could gauge the extent to which the action characterized their architectural practices.

The broad questions were open to a much wider range of interpretation than the subquestions, and the final question in each section was open-ended. This structuring of the questionnaire provided two opportunities in each series of questions for respondents to convey how they would like to be viewed, as well as sufficient detail and specificity for us to get the information we sought. We reduced the likelihood of measurement error from unclear terms and social desirability bias.

Results of the Structured Questions

We found that the detailed subquestions gave us a much better sense of the energy efficiency practices of architects than did the responses to the general questions. Occasionally, respondents would endorse the general statement, responding yes, and yet answer no to each of the subquestions. This pattern suggests that these architects may have been giving a socially desirable response to the lead question, perhaps due in part to an uncertainty about the meaning of our terms.

Other times, respondents would answer yes to the general question and yes to only one of the subquestions. For example, when answering about predesign activities, the architect may have recalled that he or she said to a client something to the effect of, "And of course the building will be efficient; the state's energy code governs that." The architect then construed that as having discussed energy efficiency during the predesign phase.

Still other times, respondents would endorse both the general question and several subquestions, but report that they had done each of these on only one project, or had done one thing on one project and another thing on another project. In this case, their endorsement of the lead question is accurate, but in the absence of the follow-up questions the researchers would draw erroneous conclusions about the prevalence of the practice.

We had insight into the accuracy of the responses from respondents' open-ended comments. These interviews were in-depth phone conversations, often lasting 20 to 30 minutes or more. Frequently, the architects would describe in some detail their projects and their activities. These discussions provided clarification for and confirmation of our close-ended coding.

When we compared the yes/no responses of program participants and nonparticipants, the participants were somewhat more likely to have taken the various actions, but the differences were not statistically significant. When we estimated the proportion of each respondent's past year's work that reflected the actions, the participants differed significantly from the nonparticipants. The more detailed the data that we examined about their actions, the more apparent the differences between the two groups.

Using the structured questions, we found that architects who had multiple exposures to the education (i.e., attended more than one event or worked with colleagues who had also attended an event) had a higher proportion of design work that incorporated the actions than did single-exposure participants. Single-exposure participants incorporated efficiency actions into their designs only a little more likely, and not significantly so, than nonparticipants.

The patterns that we found in the data about who was taking what efficiency actions are consistent with marketing theory that a message is best conveyed through multiple exposures from multiple sources (Goldberg, Fishbein & Middlestadt 1997; Hiam & Schewe 1992). Participating architects who have gone to more than one training event or who work with colleagues who have attended training engage in the efficiency actions more frequently than single-exposure participants or nonparticipants.

Of course, these participants in the voluntary training program are those who are most interested in energy-efficient design; it is likely that they were using efficient design principles prior to participating in the educational events, as Janda (1996) found for architects and engineers participating in an Energy Edge program. Our study did not compare the design behaviors of participating and nonparticipating architects prior to the training.

To gauge the influence of the program, we asked participants to what they attributed their use of energy efficiency design actions. Almost all (86%) of the multiple-exposure participants said that they suggest these energy-efficiency ideas to clients more frequently since, and *because of,* the training than they did prior to the training. Only a handful (18%) of single-exposure participants said this (Research Into Action 2001).

Half of all multiple-exposure participants also said that client acceptance of these efficiency ideas had increased since the event due, in part or fully, to the influence of the training on their (the architects') knowledge, enthusiasm, and persuasiveness. Only 11% on single-exposure participants said this.

Thus, the group that engaged in these efficiency activities most frequently—the multiple-exposure participants—attributed their activities to the influence of the training.

Understanding the Structured Questions in Light of a Thought Experiment

I can understand the architects' responses to the structured questions in light of a thought experiment I engaged in. Were you to ask me if I ate a healthy diet, I would report that I do. If you then asked me about specifics from the published dietary guidelines, such as whether I ate lots of whole grains, eight servings of fruit and vegetables per day, little fat, little salt, and little sugar, I would have to say "no," "no," "not really," "maybe," and "probably not." If you asked me about "other" actions, I would say that I consume very little sodas, fast food, and junk food.

If you asked me about the frequency or magnitude with which I consume fat, salt, and sugar, I would be able to refine my "not really" and "maybe" responses. Alternatively, based on my responses about frequency, the analyst could code "not really" and "maybe" into

accurate responses based on the concept to be measured and could come up with a measurement of the degree to which my diet is healthy.

In this thought experiment, social desirability bias enters into my first answer. I think of myself as someone who eats a healthy diet, because I think of sodas, fast food, and junk food as unhealthy and I avoid these. I also know that a good diet is important to one's health and I believe that I am taking adequate care of myself, so it follows (through faulty logic, of course) that my diet is okay. Yet were you to ask me about the specific elements that constitute a good diet according to the U.S. medical profession, I would give a thoughtful, honest appraisal of these elements. And my answers may not add up to "healthy."

"Energy efficient" is a term very much like "healthy." To the energy professionals and medical professionals, respectively, each term may be clear. But to the people we speak with, the terms may be only suggestive of good conditions. Respondents who do not perceive themselves as wasteful or negligent could in good conscience report that their actions are efficient, or healthy!

Conclusions

It is possible to explore narrowly defined, specific actions with respondents and get a good characterization of their activities.

Inaccurate socially desirable responses can be elicited from any question when there is the possibility that a respondent thinks his or her accurate answer might convey a negative impression. However, socially desirable responses particularly plague questions that use undefined terms. For these questions, it is easy for respondents to assume definitions of the concepts that are most favorable for them.

Thus, social desirability bias rears its ugly head when we ask respondents such questions as whether their appliance is energy-efficient, or their building designs incorporate day lighting, or they have applied what they learned from a training event. The more narrowly we can identify the exact action indicative of an energy-efficient behavior, the less the respondents will interpret what we mean and the less likelihood of misinterpretation.

In addition to using clearly defined terms, it works well to introduce and conclude each set of detailed questions with a more general one. The general questions enable respondents to present themselves in a positive way, as Fowler (1995) recommends.

Finally, it is useful to ask respondents to describe the extent of their actions: the frequency, or the number of buildings or pieces of equipment effected. As energy-efficiency concepts become widespread, more and more people take some energy efficiency actions. People are distinguished by the degree to which they incorporate energy efficient behaviors into their activities, not merely by whether or not they do anything at all.

References

Dillman, Don A. 1978. Mail and telephone surveys: *The Total Design Method*. New York, N.Y.: John Wiley & Sons.

Dillman, Don A. 2000. Mail and Internet surveys: *The Tailored Design Method*. New York, N.Y.: John Wiley & Sons.

Energy Center of Wisconsin. 2002. *Appliance Sales Tracking—2001 Residential Survey*. Madison, Wisc.: Energy Center of Wisconsin.

Fowler, Floyd J. Jr. 1993. *Survey Research Methods. Second Edition*. Newbury Park, Calif.: Sage Publications.

Fowler, Floyd J. Jr. 1995. *Improving Survey Questions: Design and Evaluation*. Newbury Park, Calif.: Sage Publications.

Goldberg, Marvin E., Martin Fishbein, and Susan E. Middlestadt, eds. 1997. *Social Marketing: Theoretical and Practical Perspectives*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Hiam, Alexander, and Charles D. Schewe. 1992. *The Portable MBA in Marketing*. New York, N.Y.: John Wiley & Sons.

Janda, Katheryn. 1996. "Designing from Experience: The Effects of Research on Practice." In *Proceedings of the ACEEE 1996 Summer Study on Energy Efficiency in Buildings*, 6:89-97. Washington, D.C.: American Council for an Energy-Efficient Economy.

Miller, William R., and Rollnick, Stephen. 1991. *Motivational Interviewing: Preparing People to Change Addictive Behavior*. New York, N.Y.: The Guilford Press.

Peters, Jane S. 2002. "Effects of Using WebTV Graphics on the Accuracy of Self-Reports." In *Proceedings of the ACEEE 2002 Summer Study on Energy Efficiency in Buildings*, Forthcoming. Washington, D.C.: American Council for an Energy-Efficient Economy.

Prochaska, James O., and Carlo C. DiClemente. 1986. "Toward a Comprehensive Model of Change." In W. R. Miller & N. Heather (Eds.), *Treating Addictive Behaviors: Processes of Change,* 3-27. New York, N.Y.: Putnam Press.

Research Into Action, Inc. 2001. *Third Market Progress Evaluation Report Architecture + Energy Program*. Portland, Oreg.: Northwest Energy Efficiency Alliance.