A Multi-Method Approach to Survey Pretesting

Joe Murphy[1], Danielle Mayclin[2], Ashley Richards[1], David Roe[1]

[1]RTI International, [2]U.S. Energy Information Administration

Introduction

Prior to the collection of data, surveys typically undergo some form of pretesting. The methods used for survey pretesting vary depending on the specific purposes, available resources, and schedule, but generally, pretesting is used to assure that survey questions will collect the information for which they were designed. Pretesting helps to identify sources of measurement error in the survey instrument which can be ameliorated before the start of survey data collection, thereby assuring quality.

Common methods of pretesting include review of the draft questionnaire by subject matter experts and/or survey methodologists, cognitive interviews with individuals who represent the target population for the survey, and, more recently, rapid self-administered online tests of draft survey items. Each of these methods has particular advantages, disadvantages, and costs. Used in conjunction, the multiple methods allow for a well-rounded review of a draft survey, providing designers with actionable evidence for use in making updates.

This paper describes traditional and newer approaches to pretesting that were used prior to administration of the 2015 Residential Energy Consumption Survey (RECS), a national survey about home energy use sponsored by the U.S. Energy Information Administration. The 2015 RECS main study is conducted as a traditional computer-assisted personal interview (CAPI). Two pilots of self-administered RECS surveys are also being conducted via mail and web. Although the 2015 RECS and RECS Pilots are separate projects, pretesting and questionnaire design activities are conducted in coordination to meet the goal of designing web and mail instruments to collect data comparable to the CAPI instrument. The pretesting started with an expert review, followed by two rounds each of cognitive interviews and online pretesting. In this paper, we present an overview of the pretesting methods available, our considerations in designing pretests for the RECS, the results of the pretesting, and thoughts about combining multiple methods for pretesting for future studies. The RECS experience and insights may serve as a resource for other studies contemplating one or more methods of pretesting before launching a survey.

Traditional Pretesting Approaches

Two traditional and common pretesting methods are expert review of draft questionnaires and cognitive interviews with participants representing the survey's target population. Expert review can be conducted with varying levels of organization and rigor. On the low (and quick) end of the spectrum, a read-through of a draft questionnaire by an experienced subject matter export or survey methodologist can identify issues with question wording or administration that may lead to measurement error. On the more formal and rigorous end of the spectrum are methods such as RTI's Question Appraisal System (QAS). The QAS is a structured, standardized instrument review methodology that assists a survey design expert in evaluating questions relative to the tasks they require of respondents, specifically with regard to how respondents understand and respond to survey questions (Dean et al., 2004; Willis and Lessler, 1999). The QAS allows the reviewer to evaluate the structure and effectiveness of the questionnaire and identify question features that are likely to lead to response error. The reviewer examines proposed questions by considering specific categories of question characteristics in a step-wise fashion and, at each step, decides whether the question exhibits features that are likely to cause problems. The reviewer then develops recommendations for correcting each potential problem.

Expert review using approaches like the QAS can be very effective in identifying and correcting problematic questionnaire items before administration. It does not involve any interaction with actual research subjects and is a relatively efficient process. One limitation of expert review is that the perspective of methodologists or substantive researchers may not always match those of the target population. In other words, it may be difficult for the expert to put him- or herself "in the shoes" of the participant and anticipate all of the sources of confusion or error that an actual survey respondent may encounter.

Cognitive interviews are another traditional pretesting approach. Cognitive interviews are one-on-one interviews used to assess a questionnaire in terms of general understanding, question and response wording, skip logic, and visual aids (Willis, 2014). The goal is to gain an understanding of how well the questions work when administered to a sample of the survey's target population. The interviews follow a protocol with pre-scripted probes. Cognitive interviewers also use spontaneous probes to gain a better understanding of questions.

Cognitive interviewing allows the researcher to delve into depth on a particular topic and gain detailed insights into the thought processes of a small number of participants. This can be valuable information for use in refining survey questions before data collection begins. There are some limitations to cognitive interviewing, however. Given the cost of conducting interviews and schedule constraints, it is typically only feasible to conduct a small number of cognitive interviews in advance of a survey. Requiring the participant to physically travel to a particular location also limits the geographic diversity one can obtain with a cognitive interview sample. The cognitive interview process, like some surveys, is also subject to effects of the interviewer who follows a protocol but typically administers additional spontaneous probes during the course of each interview.

New Pretesting Approach: Online pretesting

Online pretesting is an approach used to quickly and efficiently identify potential issues in a survey. Unlike traditional pretesting, which is often conducted face-to-face using members of the target population, online pretesting is done entirely online, and often with a convenience sample that may or may not be part of the target population. This makes it a fast and inexpensive process to implement.

Online pretesting surveys can be programmed using free or low-cost software. Many of these software options offer built-in integration with online panels, providing seamless access to pretesting participants. Another option is using a crowdsourcing platform, such as Amazon Mechanical Turk (MTurk), which provides on-demand access to "workers" who are available to do a variety of tasks, including taking a survey.

If participants with very specific characteristics are desired, using an online panel or platform such as MTurk may be less efficient because a high number of participants may need to be screened to find enough participants with the desired characteristics. In these situations, recruiting through more common current methods, such as via Craigslist, advertisements, or flyers, may be more effective (Murphy et al., 2007). Once recruited, participants can then be directed to the online pretest, e.g., by emailing them a link to an online survey. Although this approach is more time-consuming, it allows for better targeting of specific characteristics.

Online pretesting can be set up to function similarly to a cognitive interview, but one that is self-administered rather than interviewer-administered. Traditional cognitive interviews use a combination of scripted and spontaneous probes. The latter cannot be used in online pretests, so these pretests rely entirely on scripted probes. A common approach is to ask participants to answer a survey question, then follow up with additional questions that assess the participants' comprehension, recall, interpretation, reaction, etc. to the survey question. Just like in an in-person cognitive interview, participants might be asked to describe the thought process they used to answer the previous question, or to describe anything about the question that they found confusing or unclear. In addition to asking open-ended probes following survey questions, online pretests can also include closed-ended questions to elicit other types of important feedback. Examples include asking participants about their level of confidence in their answer's accuracy, their familiarity with specific terminology, or their willingness to provide certain types of information in a survey. Online pretests typically include many more participants than cognitive interviews, so closed-ended questions can work particularly well in the online format.

Another advantage of the larger pool of participants feasible with online pretesting that sets it apart from cognitive interviewing is the ability to examine of response distributions to better detect issues with the survey questions. For example, participants can be randomly assigned to answer alternate versions of a question, then response distributions can be compared to see which versions of the question reduce the likelihood that respondents will provide improbable, irrelevant, or out of range responses. This technique is less practical in cognitive interviews or other in-person pretests because the number of participants is likely too small for detecting differences.

Online pretesting examples

Some of the first reported online pretesting was done by Edgar (2012; 2013). This was followed with a more extensive comparison of multiple online approaches done by Murphy, Keating, and Edgar (2014). They compared three approaches for recruiting for and conducting cognitive interviews in an online format using MTurk, Facebook, and TryMyUI—an online usability testing service. Murphy et al. concluded that TryMyUI may be best for getting straightforward verbal feedback more quickly than is possible in in-person cognitive interviews; MTurk may be best when large samples, rapid turnaround, and limited resources are drivers; and Facebook may be best when looking for reactions from the "real" population rather than people who seek out these types of opportunities.

An online pretest by Lee et al. (2015) took a slightly different approach. Because of the specific target population – U.S. adults who had used an e-cigarette in the past week – they thought it would be more efficient to recruit their own participants rather than using an existing panel or platform. Participants were recruited via Craigslist ads posted to targeted cities around the U.S. The ads directed potential participants to an online screening survey and participants were selected for the pretest from among the screener responses. This allowed for careful selection of participants, ensuring the sample was a demographically diverse group, included participants living in specific parts of the country, and a variety of types of e-cigarettes were used by the participants. Once participants were selected, they were emailed a link to the online pretesting survey. Participants were notified they had 24 hours to complete the survey and they would be emailed a $5 Amazon.com gift card automatically upon completion of the survey.

Advantages and disadvantages of online pretesting

Online pretesting has many benefits. One of the key benefits is the speed at which it can be completed. While in-person pretesting requires recruiting participants individually and meeting with them one-on-one, online pretesting is simpler because it requires no interviewers and recruitment is often built into the crowdsourcing platform. Furthermore, because online pretests are self-administered, all participants can complete the pretest simultaneously, regardless of staff size. Also, participants who drop out before completing the task can be replaced more quickly than in a face-to-face pretest that must be scheduled in advance. When using an online panel or a platform like MTurk, this replacement happens seamlessly with no additional effort or delay. Because of these considerations, an entire online pretest can be completed in less than one hour from start to finish.

Online pretesting allows for quick recruitment of participants, as well as a number of other recruitment benefits. The online setting eliminates geographic barriers and opens up the pool of prospective participants to literally any willing participant in the world who has Internet access. This allows for greater diversity of participants compared to a traditional pretest, with geographic diversity being especially important. Greater diversity among participants can lead to richer findings. For example, in Murphy et al.'s online pretest of a survey, participants from different regions of the United States identified region-specific issues with the wording of response options (2014). If the pretest had been conducted in-person in only one or two locations, these regional differences would have likely been missed. Regional differences were also critical in Lee et al.'s online pretest (2015). A convenience sample was carefully selected to include a roughly even mix of participants who live in states that have and have not legalized marijuana because some of the questions asked about familiarity with devices that can be used for smoking marijuana, and the researchers predicted this would differ by legalization status. In both of these studies, regional differences in responding may have been missed if pretesting had not been conducted online and was restricted to one or two cities due to cost and time considerations.

It follows from the benefits mentioned above that an added benefit of online pretesting is the low cost compared to traditional pretesting. There are no travel costs for staff and there is no need to provide travel or parking reimbursements for participants. Because there is less staff involvement, staffing costs are lower. Incentives paid to respondents are also typically lower because of the nature of the task and conventions for similar online tasks. Whereas traditional, in-person pretesting may provide incentives of $40 or $50 per respondent, the norm in online pretesting is much lower, ranging from $0.75 for a short task to $5 or $10 for a longer task.

Aside from the practical considerations of time, recruitment, and cost, it is also important to consider the quality of responses produced from online pretesting. An advantage of online pretesting is that participants may be more willing to share criticism of the questionnaire. Sometimes participants hold back negative opinions in face-to-face pretests so as not to offend the person administering the pretest. Because online pretesting is self-administered,

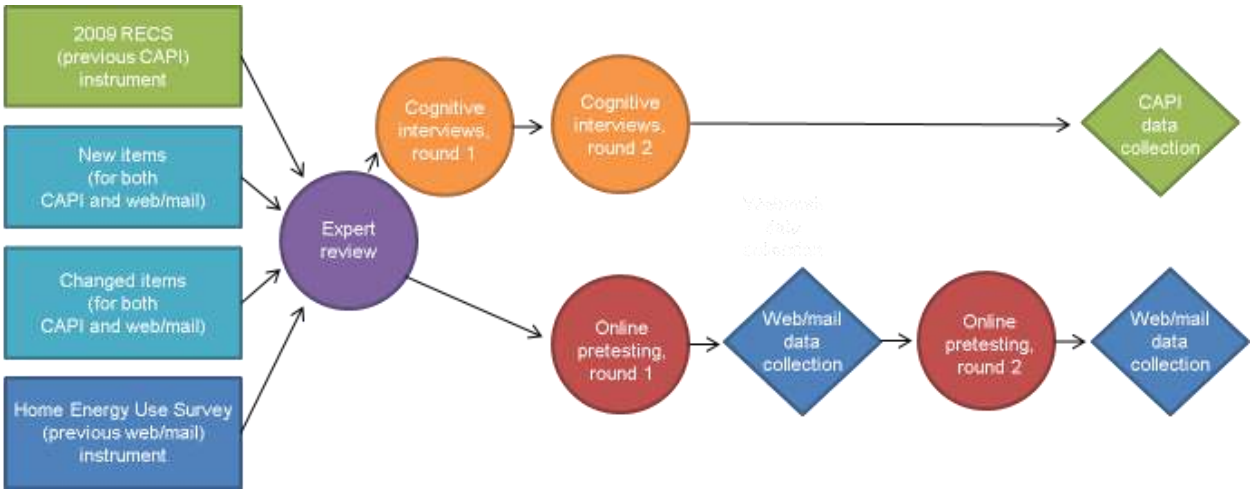
participants may be more likely to share their full opinion of the weaknesses of a questionnaire. However, even though participants in an online pretest may be willing to openly share their feedback, that does not guarantee they will. Many online participants aim to complete the task efficiently, so they may not provide as thorough or thoughtful feedback as they likely would in a face-to-face setting. This is more common when using participants from MTurk or an online panel, because these participants tend to complete multiple, similar tasks and earn more money as they complete more tasks. Another data quality limitation of online pretesting is that all probes must be scripted. The data collected may not be as rich as data collected face-to-face, because that setting allows for follow-up questions to address unique situations or feedback. Because online pretesting has both advantages and disadvantages that set it apart from traditional pretesting, the best approach to pretesting may be a combination of multiple approaches.

Pretesting Methods for RECS CAPI and Pilots

The RECS CAPI and Pilots pretesting involved testing of questionnaire content and materials over multiple stages and modes. Pretesting leveraged the unique benefits of multiple approaches, employing a combination of expert review (QAS), cognitive interviewing, and online pretesting via crowdsourcing during a three-phase pretesting period.

When taken together, pretesting methods were critical in receiving user feedback and making adjustments including improvements to question wording, respondent assistance (definitions, instructions, and clarification), and show cards. These improvements covered a variety of key topical areas including lighting, heating and air conditioning usage and behavior, cooking behavior, appliance ownership and usage, and television equipment and usage. **Figure 1** presents the sequence of pretesting methods employed for the RECS CAPI and Pilots, beginning with the input material that underwent expert review using the QAS, followed by two rounds each of cognitive interviewing and online pretesting prior to data collection. Below, we describe each component of the three-step approach employed.

Figure 1. Coordinated pretesting schematic for the 2015 Residential Energy Consumption Survey and Pilots



Expert review via Question Appraisal System

Using the QAS, expert reviewers examined proposed questions by considering specific categories of question characteristics in a step-wise fashion and, at each step, decided whether the questions exhibited features that were likely to cause problems. The reviewer then developed recommendations for correcting each potential problem. The QAS was conducted in stages as follows:

1. review of the 2009 RECS CAPI instrument, which served as the base for the 2015 RECS CAPI,

2. review of the Home Energy Use Survey (HEUS) instrument, which served as the base for the RECS self-administered web and mail instruments, and

3. review of items slated for addition or modification across all instrument modes.

At each stage, the QAS took into account the planned mode of data collection (in-person, web, or paper), which is an important consideration when identifying potential problems in a questionnaire. The QAS efforts were split between two reviewers.

Cognitive interviews

Cognitive interviews for the RECS CAPI and Pilots focused on how to collect information about new technologies that are now common in homes, potential revisions to questions that were outdated or known to be problematic, necessary adjustments to items based on data collection mode (interviewer- or self-administered) and comparison of alternate response options that may improve respondent understanding and data quality.

Cognitive interviewing took place over two rounds. The first tested the questionnaires with 15 participants in three U.S. cities and probed their comprehension and response strategies while providing respondents with an opportunity to discuss areas of confusion. Following Round 1 testing, we updated the questionnaire with accepted cognitive interview recommendations. We then conducted a second round of cognitive interviews focused on testing the questionnaire updates, again with 15 participants across three cities

Online pretesting

The online pretesting used for the RECS CAPI and Pilots recruited participants using MTurk. We conducted two rounds of online pretesting. The first round included 57 participants and the second round included 95 participants. With MTurk, we were able to set a quota and ensure that the appropriate exclusion criteria (e.g. U.S. adults only) were applied.

Results

In this section we illustrate some questionnaire issues discovered at each stage of pretesting, and then highlight some of the questionnaire improvements made to specific items over multiple pretesting methods.

The multi-method, multi-stage pretesting approach used for RECS pretesting led to a multitude of questionnaire improvements across the CAPI, paper self-administered, and web self-administered instruments. The pretesting started with the very thorough QAS review, in which multiple experts evaluated potential new questionnaire items, potential changes to previous questionnaire items, and previous survey instruments. The QAS reports totaled over fifty pages, and included global wording recommendations as well as recommendations and comments on individual items. Global wording recommendations coming out of the QAS included ensuring consistency in terminology (for example: home vs. house vs. housing unit), adding or modifying respondent assistance (definitions, instructions, etc.), clarifying reference periods, ensuring response options are mutually exclusive and exhaustive, and for the web and mail instruments, tailoring the wording and format of items previously used in the CAPI instrument to perform in the new self-administered modes. An example of the expert review leading to improved question wording is shown below with a newly drafted item about vacancy in the housing unit, which had not previously been collected. The goal of the question was to identify housing units with periods of vacancy, which can have a major impact on overall energy consumption (usage would likely be quite low or zero during a vacant period).

Question reviewed in QAS: Before your household moved in, was your home occupied by a different household, vacant, or do you not know?

Occupied by a different household
Vacant
Do not know if occupied or vacant

Comments from report: The question is a bit vague. How long does a home need to be empty before it counts as vacant? If a rental is empty for 2 weeks between when the old tenant moves out and the new tenant moves in, does that count? How about 1 month? 2 months? Etc.

Recommendation: Provide specifications on how long different properties must be empty for it to be vacant.

Question in final CAPI, web, and mail instruments: During the past year, was your home vacant for one or more months? – *Web and mail instruments included "Don't know" as an explicit response option.*

Yes
No

After most of the QAS recommendations were implemented in the updated CAPI, web and paper questionnaire specifications, some of these revised items underwent further testing in subsequent cognitive interviews and online pretesting. As the questionnaire specifications included hundreds of individual items, this content needed to be prioritized prior to cognitive interviewing and online pretesting to make the best use of participant and project staff time. When prioritizing content, the focus was on new items, changed items, and items for which data quality issues had been observed in previous rounds. Additionally, as the survey sponsor was piloting web and paper instruments for the first time, attention was paid to items identified as "mode sensitive." For example, in the CAPI instrument, the interviewer determines the housing unit type, which is used to drive much of the survey skip logic; we were concerned about how respondent identification of housing unit type would compare and thus included it as a pretesting priority. After about thirty priority items or topics were identified, the survey sponsor and data collection contractor collaborated to determine the most appropriate pretesting method(s) for each. The pros and cons of each method described earlier were considered when making this determination, leading to the final identification of a handful of topics for inclusion in cognitive interviewing and slightly more topics for inclusion in online pretesting. A few select high priority areas were identified for both cognitive interviewing and online pretesting.

As mentioned earlier, the two rounds of cognitive interviews each included 15 participants across 3 U.S. cities. Each round of interviews focused on about five topical areas, such as behavior related to heating and cooling the home, and followed a protocol with pre-scripted probes. The interviewers also used spontaneous probes as they deemed necessary. The first round of cognitive interviews was designed to be more exploratory in nature, as demonstrated by the example lighting question below. The recommendations from the first round report led to drafting, modifying, or removing specific items from the CAPI instrument, with some of these items undergoing further testing in the second round of cognitive interviews or online pretesting. The protocol for the second round of cognitive interviews more closely mimicked the CAPI interview, with closed response sets for most items and the interviewer presenting show cards to the participants where necessary. The recommendations from the second round report typically led to minor refinements in question wording. Additionally, the assessment for many items that had been changed following the first round of cognitive interviews was that no further testing or changes of these items were required.

One key topic the RECS project team prioritized for cognitive interviewing was lighting. The primary items used in previous RECS rounds asked about the number of lights used in specific ranges, for example between 1 and 4 hours per day, or between 4 and 12 hours per day. The responses to these self-reported items were considered to be of poor quality (a severe undercount) compared to external data collected through physical measures such as home audits and light logging studies, and thus were not used for internal analysis or included in data tabulations. A series of exploratory questions about lighting for the first round of cognitive interviews collected rich detail about the types of

information respondents could provide about the lighting in their home, which informed the key question in the newly-designed lighting section of the CAPI, web, and mail instruments (shown below)

Question in cognitive interviews, round 1: How many indoor lights do you have in your house?

Selected comments from report: Many participants "reported confusion when asked to report how many lights they have in their home. Confusion related to what was to be included in count…" Some respondents asked for clarification, for example: "Do you mean in every room?" "Should I include lamps?" and "Are these only lights in the ceiling?" Additionally, many "forgot about the lights that are rarely used."

Recommendations: Provide more instructions on what to include in the count (fixtures, bulbs, etc.). Remind participants to include rarely used lights such as lights in closets, attics and basements, as well as under cabinet lighting and lights above cooktops. Specify if participants should count light bulbs or fixtures.

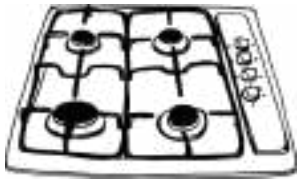
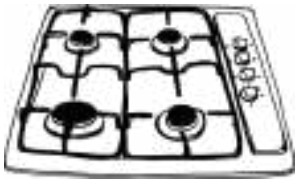
Revised question in final CAPI, web, and mail instruments: About how many light bulbs are installed inside your home? Include light bulbs in ceiling fixtures and fans, table and floor lamps, as well as those used infrequently, such as in hallways, closets, and garages. For fixtures with multiple bulbs, count each bulb separately. – *CAPI instrument included a show card with the following response options.*

Fewer than 20 light bulbs
20 to 39 light bulbs
40 to 59 light bulbs
60 to 79 light bulbs
80 or more light bulbs

As mentioned earlier, two rounds of online pretesting were conducted using MTurk, with a total of 152 participants. The first round of online pretesting was divided across three "Human Intelligence Tasks" or HITs, which each had a topical theme, such as water heating. The relatively large number of participants engaged in the online pretesting allowed for quantitative feedback from the observed response distributions, in addition to qualitative feedback collected through pre-scripted follow-up questions following many items (which is more similar to the feedback from cognitive interviews). Some HITs included items previously tested and improved upon by the cognitive interview process, notably the series of new items about lighting. The second round of online pretesting occurred between the two web/mail pilots and was designed to allow for testing problematic items identified in the initial instruments. In an attempt to fix observed issues from the first pilot, the participants for the second round of online pretesting were split across three HITs that covered the same content, but with different versions of selected items.

Figure 2 shows three versions of an item tested against each other in the second round of online pretesting. In the previous CAPI instrument this item, which collects the number of separate cooktops in the home, had not been problematic, but in the first web/mail pilot it appeared that many respondents misinterpreted the item (seen in version A) to ask for the number of separate **burners** instead. This led to approximately 20% of respondents providing answers of 4 or higher for this item. Both a different graphic and alternate instructional text were tested, with one-third of the participants shown version A, one-third shown version B, and one-third shown version C. As version C, and to a lesser extent version B, reduced the number of respondents with answers of 4 or higher, the final web and mail instruments for the second pilot used the wording from version C along with the picture from version B.

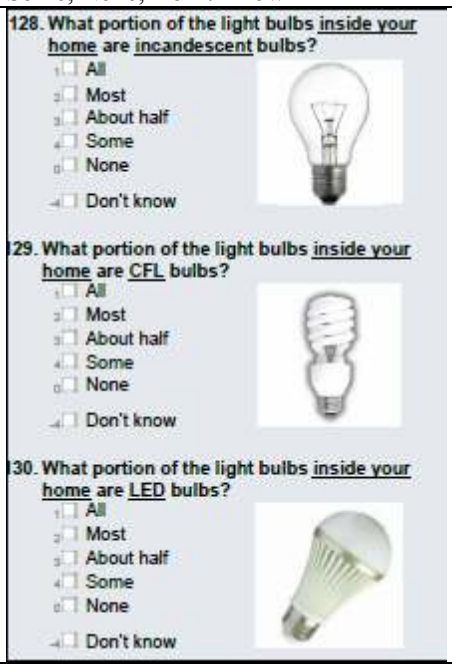
Figure 2. Three Versions of Cooktop Questions Presented in Online Pretests

Version A	Version B	Version C
		
The next question asks about **separate cooktops**. An example is displayed above. How many separate cooktops do you have in your home?	The next question asks about **separate cooktops**. An example is displayed above. How many separate cooktops do you have in your home?	An example **separate cooktop** is displayed above. How many separate cooktops do you have in your home? (Count the entire cooktop, not the number of burners. Do not include cooktops that are attached to an oven.)
Frequency (*n*=36)	Frequency (*n*=33)	Frequency (*n*=23)
1: 21	1: 25	1: 26
2: 10	2: 6	2: 7
3: 1	3: 0	3: 0
4: 0	4: 0	4: 0
5: 4	5: 1	5: 0
6: 0	6: 1	6: 0

To illustrate how this iterative multi-method pretesting process led to improved survey instruments, Figure 3 focuses on the development of an item about compact fluorescent light bulbs (CFLs). As previously mentioned, we had decided to discard the lighting items used in the previous RECS CAPI instrument and completely redesign this series. Many lighting questions were included in all phases of the pretesting process, and the findings and recommendations were incorporated into the final CAPI, web, and mail instruments. While data collection is ongoing for the CAPI and second round pilot, the team is optimistic that the revised items will have higher quality and better meet the needs of internal and external data users.

Figure 3. Development of Question about CFLs through the Pretesting Process

Pretesting phase	Item wording	Recommendation
QAS	How many of these lights use energy-efficient bulbs, such as compact fluorescent (CFL) bulbs?	Some people may have these types of bulbs but may not know what they are called. Add a show card with images of the different types of energy efficient bulbs.
Cognitive interviews, round 1	Do you have any compact fluorescent or CFL bulbs installed in your house?	Recommend including images and/or descriptions of the different kinds of light bulbs, including standard incandescent lights. If images cannot be used, describe the CFL as the "curled" or "spiral" light bulb. Including the retail price of bulbs on a showcard may also help respondents remember which they have. Crowdsourcing may be a good option for gauging respondent recognition of different bulb types.

Online pretesting, round 1	*Please look at the following pictures.* **Example CFL Bulb** **Example LED Bulb**  What portion of the light bulbs inside your home are **[CFL/LED] bulbs**? Response options: All, Most, About half, Some, None, Don't Know	We recommend including these images, as well as an image of an incandescent bulb.
Final web/mail instrument *Note: CAPI instrument wording is the same, but pictures for all three questions are included on one show card.*		

Discussion

In this paper, we discuss two traditional methods of survey pretesting and one newer option. The traditional method of expert review has the advantage of applying a systematic and organized process to evaluating draft questionnaire items to identify key potential sources of error. By involving methodological and substantive experts, these reviews do not require involvement from research subjects and therefore involve no burden and relatively little resources and schedule time to complete. On the down side, the perspective of those in the target population is not reflected in the questionnaire design as a result of expert review. To the extent that experts and the target population differ in their reactions to draft items, issues can be missed and measurement error can occur.

To involve those who represent the target population and can provide insights on the thought processes of real respondents who will be completing the survey, researchers often conduct one or more rounds of cognitive testing. This method can identify problematic questions or response options and suggest improvements to mitigate these issues. However, cognitive interviewing can be cost prohibitive and time intensive. It usually requires recruiting participants who travel to complete the interview. Interviewers are needed to administer the protocol and can introduce other errors into the process, just as with a survey. Cognitive interviews are typically limited to the extent that they can be conducted with a geographically dispersed population and often rely on small numbers of participants who may or may not provide a broad perspective on potential questionnaire issues.

With the advent of the internet age, crowdsourcing has facilitated online pretesting, a new option for surveys. Online pretests can be conducted with a panel of individuals ready and willing to participate, in geographically diverse locations, at a relatively low cost, without the presence of an interviewer, and with a larger number of individuals than other pretesting methods like cognitive interviews. However, the remote nature of online pretests has some drawbacks, as it can be difficult to investigate particular issues in great depth without the aid of in-person administration by an interviewer. Nevertheless, platforms like Amazon Mechanical Turk offer new options to supplement traditional pretesting methods in the evaluation of a draft questionnaire.

For the 2015 Residential Energy Consumption Survey and Pilots, we used each of the above methods in a staged fashion with a single round of expert review, followed by two rounds of cognitive interviewing and online pretesting. We found that these multiple methods allowed for a well-rounded view of potential sources of measurement error in the questionnaire and pointed to solutions to problems that could be implemented and tested in later stages of pretesting. We highlight examples of such items, including those asking about compact fluorescent (CFL) lightbulbs. The pretesting methods revealed potential and actual participant issues with the question wording and appearance. They allowed us to update the question and test the effectiveness of the improvement. As a result, we feel confident that the measure included in the RECS is measuring the intended concepts.

We feel that this staged approach worked very well for pretesting the RECS. In hindsight, we would have planned for many more online pretests to supplement the design. These pretests can be conducted so quickly and inexpensively that researchers can efficiently conduct hundreds of pretests in a matter of days. For future studies, we would recommend online pretesting to evaluate changes from earlier phases of pretesting with substantially larger sample sizes than those used in cognitive testing.

Finally, we find that multiple pretesting methods were valuable for the RECS which has traditionally been conducted as a CAPI survey but is also evaluating self-administered mail and web versions. By varying the mode used for conducting pretests, we were able to gain additional insight into some issues facing respondents in different survey environments. We encourage other researchers to build off our findings to design their own multi-method pretest approaches and share the results with the survey community to provide further empirical evidence of effective pretesting methods.

References

Dean, E., Caspar, R., McAvinchey, G., Reed, L., & Quiroz, R. (2007). Developing a low-cost technique for parallel cross-cultural instrument development: The question appraisal system (QAS-04). *International Journal of Social Research Methodology*, 10 (3):227-241.

Edgar, J. (2012) Cognitive Interviews without the Cognitive Interviewer? Presented at the 67th Annual Conference of the American Association for Public Opinion Research, Orlando, FL

Edgar, J. (2013). Self-Administered Cognitive Interviewing. Presented at the 68th Annual Conference of the American Association for Public Opinion Research, Boston, MA.

Lee, Y. O., Richards, A. K., Majeed, B., & Smith, A. C. (2015, June). *Patterns in use behaviors across electronic cigarette device types*. Presented at FDA CTP public workshop "Electronic Cigarettes and the Public Health." Hyattsville, MD.

Murphy, J. J., Keating, M. D., & Edgar, J. (2014). Crowdsourcing in the cognitive interviewing process. In *Proceedings of the 2013 Federal Committee on Statistical Methodology Research Conference*, pp. H-1:1–11.

Murphy, J.J., Sha, M., Flanigan, T.F., Dean, E.D., Morton, J.E., Snodgrass, J.A., & Ruppenkamp, J.W. (2007). Using Craigslist to Recruit Cognitive Interview Respondents. Paper presented at the Annual Conference of the Midwest Association for Public Opinion Research, Chicago, IL.

Willis G. B., & Lessler J. T. (1999) Question Appraisal System: QAS 99. National Cancer Institute.

Willis, G. B. (2015). *Analysis of the Cognitive Interview in Questionnaire Design*. Oxford University Press.