

Building Experiments into Cognitive Testing: An Example

Carol Cosenza & Jack Fowler

Center for Survey Research, University of Massachusetts, Boston

1. Background

CAHPS® is a family of survey instruments designed to capture people's experiences with obtaining medical care. Teams of researchers from many organizations around the United States have taken part in this collaborative effort. The instruments have been used for people with both public (Medicare and Medicaid) and private insurance. Over the years, the survey has been modified and tailored to meet the needs of specific populations, such as those with End Stage Renal Disease (focusing on care in dialysis centers) and children (to be answered by a guardian). In general, the CAHPS instruments are primarily geared to people who are ambulatory. Yet, many people with public insurance are living in nursing homes. That fact raised the question of whether or not a CAHPS instrument could be adapted to capture the experiences of those in nursing homes. A pilot effort was launched to develop and evaluate a survey instrument that might be appropriate for use in nursing homes.

Focus groups were conducted with nursing home residents and with family members of nursing home residents. Starting with what we learned in the focus groups and building on instruments previously used by others in nursing home settings, as well as the core CAHPS instrument itself, a new survey instrument was designed. The next phase of the project was doing cognitive interviews with nursing home residents using this instrument.

In total, during the first round of cognitive testing, 52 interviews were completed in 5 different nursing homes by 3 different survey organizations. All three organizations found similar results. Most obviously, almost half of the respondents had trouble answering the questions.

The instrument we tested was fairly traditional (see Example 1) and asked for reports about how often certain things happened and for ratings about how they felt about things. However, interviewers felt that many respondents were not able to provide meaningful answers. This was not surprising, since many nursing home residents have a difficult time even with straightforward cognitive tasks that others can handle more easily.

We found that there were two major problems respondents had with the instrument :

- 1) summarizing across people and across time, and
- 2) focusing on a particular time period.

Example 1: Questions from instrument used in cognitive testing

<p>In the last week, how often were you given enough time to finish your meals?</p> <p><input type="checkbox"/> Never <input type="checkbox"/> Sometimes <input type="checkbox"/> Usually <input type="checkbox"/> Always</p> <p>How often do you <u>have trouble understanding</u> the nurses' aides and other staff when they talk to you because of the way they speak your language?</p> <p><input type="checkbox"/> Never <input type="checkbox"/> Sometimes <input type="checkbox"/> Usually <input type="checkbox"/> Always</p> <p>During the <u>day</u> last week, was your call light answered promptly?</p> <p><input type="checkbox"/> YES <input type="checkbox"/> NO <input type="checkbox"/> DID NOT USE CALL LIGHT</p> <p>In the last four weeks, did you get the special therapy you needed?</p> <p><input type="checkbox"/> Yes, all that you needed <input type="checkbox"/> Yes, some of what you needed <input type="checkbox"/> No, none of what you needed</p> <p>In the last six months, we want to know how you rate all the staff, nurses and nurses' aides who took care of you. How would you rate your care?</p> <p><input type="checkbox"/> Excellent <input type="checkbox"/> Very Good <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Poor</p>

2. The task

As we continued to analyze the results of the cognitive interviews, we realized that until we could figure out what type and form of question most nursing home residents could answer, there was little need to worry about the content. At that point, we decided to combine a complex experiment with the cognitive testing. We determined that there were 3 key features that could vary in questions to measure nursing home experiences:

1. **Type of question**
 - Report (asking whether or how often something happened)
 - Rating (asking how someone felt about something)
2. **Time period** that is asked about
 - One day (such as yesterday or today)
 - Multiple day period (such as "last week," "last 7 days," or "past month")
 - Non-specific present (no specific time period defined)

3. Type of **response task**

Reports

- Yes/No
- Frequency report with adverbs (such as “always-never”)
- Number of days
- Days-based frequency task (such as “every days, some days, no days”)

Ratings

- Ordered adjectives (such as “excellent-poor”)
- Comparative evaluation (such as “ok, could be a little better, could be a lot better”)
- Numbered ratings (such as a 0-10 scale)

3. The experiment

The experiment involved taking a few concepts (such as food, getting help, and noise) and developing alternatives that varied all the question characteristics listed. Then, these variations were administered to a sample of nursing home residents, who were questioned about how they understood the questions and how they arrived at their answers. The goal was a *systematic test* of how we can get information from people who were cognitively challenged. By creating a taxonomy of possible options, we were able to test many different ways these kinds of questions can be asked. Table 1 shows an example of the different questions that could be asked about one concept.

Two additional rounds of cognitive testing yielded clear results about how best to get meaningful answers from a population that has severe cognitive difficulties. The first round of testing focused on figuring out the best *question type* and *time period*.

We found that asking about “yesterday” did not work well because it provided a very limited basis for respondents to report about. Some things that might happen fairly often might not have happened on the one day in question. There were also respondents who answered about the times something did happen (even if it wasn’t on the day in question). The “last week” was also problematic, since we again found that respondents had difficulty summarizing over time and focusing on a specific reference period. Often, the events that constitute the denominator for calculating the answers occur frequently, they do not stand out as events very much, and therefore they are hard to collect cognitively for anyone. Think about the denominator of all the times in the last week that eating or going to the bathroom occurred. That is a large, complicated, diverse denominator which is very hard to put together. Respondents often simplified the cognitive task by either focusing on a single event or totally ignoring the time frame. The non-specific present (asking about how are things going now) provided the most reliable responses. As we found in the initial testing, this experimental round proved that summarizing was very cognitively complex for respondents and in general was not a task they did well. We also found that asking for a rating was easier than asking for a report of the same thing (since ratings do not rely on respondents having to summarize their experiences).

Knowing which format seemed to work better, we refined our testing further for the next round of cognitive interviews. Our focus was now on the various types of *response tasks*. We tested rating with adjectives, numbered rating scales, and comparative evaluations. We found that adjective scales (including “excellent - poor”) were hard for respondents to remember and use, even when the responses were listed on a show card. Respondents also had trouble with the comparative

evaluations. We are not certain whether it was the scales themselves that were difficult, or the concept of comparative evaluation that was cognitively complex for people. What our testing did show was that the best form for questions for this population (nursing home residents) was to ask ratings in the non-specific present using a 0-10 rating scale (Example 2). Most respondents seemed comfortable with the numbers and the 11-point scale provided a much wider spread of responses than would have been possible using words.

Example 2: Example of final format

Use any number from 0 - 10 where 0 is the worst possible and 10 is the best possible. What number would you use to rate how comfortable the temperature in the nursing home is?

4. Conclusions

- The taxonomy we developed for testing encompasses most of the possible variations in how to get people to describe their experiences. Considering all the ways in which a concept could be asked about, this is a tool that could be used in almost any setting asking about almost any type of experience.

- The idea of testing multiple variations of one or two questions might be a useful exercise in deciding on the best form in which to ask questions to a given population. If researchers are worried that the question form itself might influence how the question is answered, it may be worth spending the time to test various question formats before testing content. When working with special populations, tailoring the task to what the population can do best is as important as asking about the correct concept and using the correct vocabulary.

Table 1: Possible variations for “staff comes quickly when called” concept.

				TIME PERIOD	
		One Day (Yesterday/Today)	Week (Last 7 days/Last week)	Non-Specific Present	
REPORTS					
Yes/No		Yesterday, did someone come quickly when you called for help?	In the last week, did someone come quickly when you called for help?	Does someone come quickly when you call for help?	
frequency report with adverbs such as: Always to Never		Yesterday, how often did someone come quickly when you called for help?	In the last 7 days, how often did someone come quickly when you called for help?	How often does someone come quickly when you call for help?	
days-based frequency task such as: 1)How many days? or 2) Every days, most days, some days, no days)		--NOT POSSIBLE --	In the last 7 days, on how many days did someone come quickly when you called for help?	How often does someone come quickly when you call for help?	
RATINGS					
ratings with adjectives such as: Excellent to Poor		Yesterday, how would you rate how quickly someone came when you called for help?	In the last 7 days, how would you rate how quickly someone came when you called for help?	How would you rate how quickly someone comes when you call for help?	
comparative evaluation such as: 1) needs no improvement, needs a little improvement, needs a lot of improvement or 2) ok; could be a little better, could be a lot better		Yesterday, how would you rate how quickly someone came when you called for help?	In the last week, how would you rate how quickly someone came when you called for help?	How would you rate how quickly someone comes when you call for help?	
numbered ratings such as: 0 to 10 (Use any number from 0 to 10 where 0 is the worst possible and 10 is the best possible.)		What number would you use to rate how quickly someone came when you called for help today?	Thinking about the last week, what number would you use to rate how quickly someone came when you called for help?	What number would you use to rate how quickly someone comes when you call for help?	