



Using Cognitive Interviews to Evaluate the Spanish-Language Translation of a Dietary Questionnaire

Kerry Levin (1), Gordon Willis (2), Barbara Forsyth (3), Alicia Norberg (1), Debra Stark (1), Martha Kudela (1), and Frances Thompson (2)

1. Westat; 2. Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute; 3. Center for Advanced Study of Language, University of Maryland

This paper presents results from a qualitative evaluation of a Spanish-language version of a dietary questionnaire and characterizes the types of findings which emerged. The focus will be what was learned from cognitive interview pretesting that helped improve the questionnaire translations. Cognitive interview pretesting helped identify various types of problems respondents had understanding and answering the translated questionnaire items. The results from the cognitive interviews showed some translation errors, a few culture-related issues requiring tailored wording and several general design problems common across languages and cultures.

1. Background – Using cognitive interviews to test translated questionnaires

Cognitive interviews are commonly used in the U.S. and western Europe to pretest draft questionnaires – to identify problems respondents have understanding and answering draft questionnaire items and then to revised the items to improves understanding and response accuracy. Cognitive interview pretests can involve a range of interview techniques, but the general interview structure is pretty standard. A cognitive interviewer administers draft questionnaire items to a cognitive interview respondent who answers them. Sometimes the interviews are conducted in-person and sometimes they are conducted over the phone. Regardless of mode, the interviewer uses a cognitive interview script to administer the questions and probes to gather additional information from the respondent about how they interpret the question, how they go about remembering the information requested and how they select a response. Typically, the types of probes used focus on difficulties respondents have understanding, remembering or answering the draft questionnaire items.

Because cognitive interviews have been very useful for pretesting questionnaires, researchers have extended the methods for the purposes of pretesting questionnaire translations and pretesting wordings for questionnaires designed to be administered in cross-cultural settings.

Some have suggested using cognitive interview pretest methods cautiously for these broader, cross-language and cross-cultural purposes. For example, Pan (2003) observed that the indirect communication styles prevalent in some Asian cultures may make it difficult for cognitive interview respondents to answer traditionally direct cognitive interview probe questions. In addition, Pasick and colleagues (2001) and Goerman (2006) both reported culture- or language-related differences in how respondents interpreted some cognitive interview probe questions.

At the same time, several researchers have used cognitive interview results to identify problems with translated questionnaires and other cross-culture survey questionnaires (e.g., Willis et al, 2006; Carrasco, 2003; Johnson et al., 1995; Schoua-Glusberg, 2006; Napoles-Springer et al., 2006). The authors advocate cognitive testing as a vital step in ensuring conceptual and linguistic equivalence across languages and cultures.

The purpose of our research was to use cognitive interview methods to pretest a Spanish-language translation of items on diet administered in the U.S. as part of the National Health Interview Survey (NHIS). We hope that by documenting the kinds of problems identified through cognitive interviewing, this research contributes to the continuing conversation on the value of using cognitive interview methods to pretest and improve questionnaire translations.

2. Study design and research methods

Overview. We pretested the Spanish translation of diet items included in the 2005 National Health Interview Survey – the NHIS. The NHIS is a telephone survey conducted in the U.S., and the Spanish-language translation is used to interview U.S. respondents who prefer to complete the interview in Spanish. The U.S. Census Bureau conducts the NHIS data collection for the U.S. National Center for Health Statistics (NCHS) and so the Census Bureau directed the survey translation activities. Professional translators developed an initial translation that was reviewed by language experts at the Census Bureau and NCHS. Then, the Census Bureau convened a Translation Review Conference that included bilingual field representatives, field operations staff, and substantive and translation experts. Participants in the Committee

Translation Review worked as a team to discuss translation issues, develop options for addressing them, and make decisions about final translation wording. These NHIS translation activities followed the translation, review and adjudication processes that are recommended by translation researchers (e.g., Harkness et al., 2003) and are standard practice at the U.S. Census Bureau (2004).

Pretest Design. We conducted three rounds of cognitive interviews. The next slide illustrates the composition of the three interview rounds. Spanish-speakers in the U.S. come from a variety of regional and national backgrounds. We anticipated regional differences in dietary habits and the vocabulary used to talk about food. To ensure we included pretest respondents from a range of Spanish-language backgrounds, we conducted Spanish-language interviews in 3 different U.S. locations. In round 1, we conducted 9 interviews with Spanish-speaking respondents in Washington, DC and 9 interviews with Spanish-speaking respondents in San Jose, CA. We used results from the Round 1 interviews to test the draft Spanish-language translation, identify problems and suggest revisions. In round 2, we conducted 9 interviews with Spanish-speaking respondents in Miami, FL to test the revised Spanish-language translation. In round 3, we conducted 9 interviews with English-speaking respondents in Washington, DC. The round 3 interviews gave us a chance to assess whether we found different types of problems with the Spanish and English-language questionnaires.

The next slide illustrates the three pretest locations. In addition to including a mix of national backgrounds we also recruited respondents who represented a mix of genders, ages, and education levels.

3. Analysis and results

We used a two-step analytic process to analyze the results from the three rounds of cognitive interviews. In step 1, we reviewed narrative summaries of the interviews in each round and developed documents that summarized results for each item, across the interviews in each interview round. In step 2, we reviewed the item-level summaries for each interview round, identified problems observed, and classified the observed problems into four general categories. (1) **Translation** problems occurred when translated item wording altered the intent of the original question. **Culturally-related** problems occurred when an item's intended meaning was difficult to convey using Spanish-language constructs or constructs from a specific Spanish-

language nationality or subculture. **General design** problems occurred when respondents had difficulties with comprehension, recall, or response selection that seemed independent of culture or language. **Mixed** problems occurred when a single problem seemed related to a combination of translation, cultural-related and/or general cognitive issues.

The next slide shows counts for the four general categories of problems, by cognitive testing round. Four results are noteworthy. First, the Round 1 Spanish-language interviews revealed many problems. With 21 items tested, we found an average of approximately three problems per item. Second, many of the problems in the Round 1 interviews were general design problems that were not specific to one culture or language. Third, revisions made after the first round of interviews dramatically reduced the number of problems identified in the second round of Spanish-language interviews. This reduction was especially pronounced for general design problems. Also, revisions between Rounds 1 and 2 nearly eliminated translation problems. Fourth, in the second round of testing, the English- and Spanish-language questionnaire versions were roughly comparable in terms of the numbers of problems identified. In other words, it seems that revisions based on the first set of cognitive interviews enhanced comparability between the English- and Spanish-language questionnaires.

Let's turn to a few examples to give you a more concrete idea about the kinds of problems cognitive interviewing helped us find. First, let's look at problems identified as translation issues. Most of these problems were relatively easy to resolve by selecting alternative translation wordings or by refining or restructuring translation wordings.

Example 1. One set of translation problems involved translation wordings that did not adequately convey the intended construct. This item on salsa is an example. "During the past month, how often did you have salsa?" (You see the tested Spanish translation here, below the English-language version.) The item intends to ask about a spicy tomato-based sauce that Americans often eat with Mexican food. Cognitive interview respondents thought of the Spanish term, "salsa" more generally as "sauce" of no specific type. Furthermore, the Spanish translation literally asks about "sauce containing fruits or vegetables." Spanish-speaking cognitive interview respondents did not focus the intended spicy (or "picante") sauce. Instead, respondents (particularly in Miami and California) reported including sauces like marmalade, apple sauce, and fruit sauce for topping ice cream.

You can see that the revisions we recommended to help respondents focus on the intended spicy tomato sauce. We suggested moving the “salsa” item ahead of additional items on tomato sauces (such as spaghetti sauce); removing the Spanish-language reference to “fruits or vegetables”, and adding descriptions to focus on the intended, “spicy” type of salsa – “including spicy (hot), pico de gallo or Mexican style salsa.”

Example 2. Another type of translation error involved selecting Spanish wordings that had different meanings across regions or nationalities. This item on cookies is an example. “During the past month how often did you eat cookies, cake, pie, or brownies?” Based on cognitive interview responses, we found that “galletas” can mean either cookies or salty crackers. Also, for respondents with Mexican backgrounds, “torta” can mean a sandwich. At the same time, respondents generally indicated they recognized the intended “cookie”-related meanings. You can see the revision we recommended to accentuate the intended meanings. We suggested moving the two potentially confusing terms (“galletas” and “torta”) toward the end of the question, using the question context to clarify the intended meaning.

(Side note: we classified this as a translation problem rather than a culture-related problem because it could be circumvented by avoiding the use of “galleta” and “torta”.)

Next, let’s focus on culture-related problems. We found relatively few and resolved most of them by removing unnecessary words and adding alternative wordings to help items function effectively across regions and nationalities.

I have two examples of culture-related problems caused because the dietary concepts of interest differ across cultures or nationalities.

Example 3. Here’s an item about the consumption of “white potatoes.” “During the past month how often did you eat other white potatoes?” You see the item has additional instructions on which potato-based foods to include and exclude as “white potatoes.” Spanish-language cognitive interview results reminded us that Spanish-speakers in the U.S. eat several types of potatoes that are not white but that are nutritionally similar to white potatoes. Responses to this “potato” item should include these other types of non-white potatoes.

You can see the revision we recommended to encourage respondents to include the full range of potatoes the item intends to cover. We recommended removing “blancas” (for “white potatoes”) from the Spanish translation, and retaining the instruction to include potatoes like red-skinned and Yukon Gold potatoes.

Example 4. This item on cereals is another example of a culture-related problem caused because concepts differ across cultures or nationalities. “During the past month when you ate cereal, which kinds did you usually eat?” The response options list types of cereals that differ in terms of their fiber content. Cognitive interview responses from Spanish-speaking respondents suggested that they didn’t think about cereals in terms of “fiber content,” and so we had difficulty finding Spanish wording that communicated the question goals. Our difficulty is reflected in the revisions you see here. The general question goal seemed culture-specific. Because fiber content was not a salient characteristic of cereals for Spanish-speaking respondents, we think that addressing the issue might require reconsidering the measurement goals and modifying the English-language questionnaire accordingly. This was not an option for these NHIS items.

Example 5. This cereal item is an example of another type of culture-related problem – due to knowledge or habits not shared across cultures, regions or nationalities. On this slide, I’ve given you a little more detail about some of the cereal response options to make the problem clearer. You can see that the item uses the brand names of cereals popular in the U.S. to help respondents distinguish the categories of fiber content. This strategy works pretty well for acculturated, English-speaking respondents. But for less acculturated Spanish-speaking cognitive interview respondents, the brand names were completely unfamiliar. We think that unfamiliarity with the listed brand names amplified the general difficulties Spanish-speaking respondents seemed to have answering this item because “fiber content” was a dietary factor they were not used to thinking about.

As indicated in the recommended revisions discussed, we think that developing a Spanish-language question that addresses new immigrants’ would probably require reconsidering measurement goals and modifying the English-language questionnaire accordingly.

Example 6. The fruit juice item illustrates an example of a culture-related issue where reproducing the question intent in Spanish requires some culture-specific wording. The English

language item asks “During the past month, how often did you drink 100% FRUIT JUICE or 100% fruit juice blends, such as orange, mango, apple, and grape juices?” Several cognitive interview respondents incorrectly reported fruit juice consumption. We think this item had several problems and I’ll talk about it again in a minute. One thing cognitive interview respondents told us was that the examples used to help define “fruit juice” were juices they had little experience with. Several items that used examples to define the foods of interest had similar issues due to food differences across cultures. We recommended revising the Spanish-language items to include examples more familiar to Spanish-speaking respondents in the U.S. *Papaya* for the fruit juice item, *yucca* for the “other vegetable” item and *avena* for the cereals item are some examples.

Next, let’s look at the general design problems that cognitive interviewing helped us identify. These problems were most common and we used a variety of approaches for addressing them.

Example 7. The fruit juice item gives an example of a generic problem where the fruit juice food category definition was insufficient – for both Spanish- and English-speaking respondents. Spanish-speaking respondents interpreted “100% fruit juice” as any drink made at home using real fruit even when the drink included added ingredients such as water, milk or sugar. English-speaking respondents interpreted “100% fruit juice” as excluding juices made from concentrate. Also, both Spanish- and English-speaking respondents indicated they did not understand the intended distinction between “fruit juice” in this item and “fruit-flavored drinks” in a later item.

Among the revisions listed here, we recommended the Spanish item ask about “pure” fruit juice rather than “100%” fruit juice. Also, we added a description, “without added sugar”, and removed the confusing reference to “fruit-flavored drinks.”

Example 8. Several items included instructions with examples that interviewers read as necessary to help define food categories. Cognitive interview results indicated that many of these optional instructions were actually necessary to help respondents interpret items correctly. For example, the item on “milk” used the instruction with examples show here. “Include skim, no-fat, whole milk, buttermilk, and lactose-free milk”. Also include “chocolate or other flavored milks.” When interviewers did not read these instructions, respondents routinely left some foods out of their reports that they should have included. Additionally, respondents said the examples

in the instructions were helpful when interviewers read them. You can see that we recommended revising interviewer instructions so they always read defining examples.

Example 9. This slide shows an example of a similar general design problem – due to optional instructions about the reporting period. When interviewers omitted these optional instructions, respondents reported for a host of time frames. For example: “the past week,” “the past 2 years (since the respondent arrived in the U.S.), “since becoming pregnant,” and “yesterday.” Again, we recommended revising interviewer instructions so they always mention the time frame, “during the past month.”

Example 10. This last example represents a general problem because the question wording elicited uncodeable responses. This item on green leafy salads illustrates the wording that was problematic across the full set of dietary items. The questions asked “During the past month how often did you eat lettuce or green leafy SALAD?” The item requires a numerical frequency response. Instead, respondents routinely provided verbal responses such as “not too often”; “hardly ever”; or “all of the time.” We recommended revising all questions to ask, “During the past month, how many times per day, per week or per month did you drink/eat....” – in this case green leafy salad.

4. Summary and conclusions

I'll finish up by reviewing some of the things we learned from the cognitive interview pretest. Most importantly, cognitive interview pretesting helped us identify problems in translated questionnaire items that would likely interfere with measurement accuracy. Notably, the process of cognitive interviewing of Hispanics, in Spanish, presented no persistent obstacles that were not also present in the English-language interviews. This result is important because other researchers have reported mixed results using cognitive interviews to test survey translations. We hypothesize that at least two factors contributed to our successes using cognitive interviews to identify design problems. First, we used protocols that followed the guidelines developed by Goerman (2006) to help respondents understand cognitive interview probes and cognitive interview purposes. Second, Spanish-speaking cultures may be relatively close to the U.S. along the continuum of directness, a factor Pan identified as likely to be important in predicting the effectiveness of the cognitive interview pretest method. NCI is currently conducting similar research testing Asian-language translations of diet questions. We hope that future comparisons between Spanish-language and Asian-language results using cognitive interviews may shed additional light on how these two factors affect cognitive interview results. We hope the result will be clearer understanding about when cognitive interviewing is likely to be useful and when other pretest methods are likely to be more useful.

In addition, it is interesting that the cognitive interview results helped us to identify a range of question design issues – including translation, culture-related and general design issues. The translation issues identified were apparently overlooked by the questionnaire translation process, even though the translation process followed well-accepted practices. This leads us to wonder whether there are some types of translation problems that are difficult for language professionals to detect and easier to notice based on input from more naïve language users. This is one topic that we are interested in exploring in future research.

Cognitive interviewing identified relatively few culture-related problems. We anticipated that culture-related problems might be more frequent because dietary habits are often closely related to culture. Two hypotheses suggest themselves. First, perhaps the cognitive interview pretest found relatively few culture-related problems because the team-based translation and review process used to develop the Spanish translation identified and eliminated most culture-related problems. Second, perhaps the cognitive interview pretest identified relatively few culture-

related problems because cognitive interview methods are not effective for identifying culture-related problems. We are exploring these two hypotheses in a new study that documents and compares the types of problems found at different steps in the translation, review, adjudication and pretesting process. This research will also compare results across survey content areas to examine the effect of survey topic on types of problems found.

Cognitive interview pretesting identified relatively many general design problems. The prevalence of these problems was surprising because English-language versions of the NHIS dietary items have been fielded several times.

We suggest three hypotheses that might explain why we observed so many general problems in translations of these previously fielded questions. First, perhaps some of the generic problems found through cognitive interviewing are not obvious when the questions are administered in field settings. For example, interviewers may be unaware when respondents overlook some foods that should be included in their survey responses. Second, well-trained interviewers may intervene in field settings to fix some of the problems found through cognitive interviewing. For example, when respondents provide uncodeable responses to items that ask “how often,” interviewers may assist respondents by asking for the number of times per day, per week, or per month. Furthermore, if interviewers intervene and respondents learn the response format quickly, then the “how often” stem may pose a real problem for relatively few items. Third, problems identified using cognitive interview methods, may not be problems at all in standard field settings.

This last point is related to a general caveat for results reported here. We have interpreted issues identified by cognitive testing as indicators of “problems” with the Spanish-language translation. It is clear that these issues were problems for the cognitive interview respondents. We do not have survey data to verify that issues we identified through cognitive interviews are real problems in fielded interviews. Future administrations of the NHIS diet questions will provide opportunities to determine whether issues identified in cognitive interview testing also predict survey responses or survey response errors.

References

- Carrasco, L. (2003). The American Community Survey (ACS) en Español: Using cognitive interviews to test the functional equivalency of questionnaire translations. (Study Series, Survey Methodology # 2003-17. Washington, DC: Statistical Research Division, U.S. Census Bureau.
- Goerman, P. (2006a). Adapting cognitive interviews for use in pretesting Spanish language instruments. (Research Report Series, #2006-3), Washington, DC: Statistical Research Division, U.S. Census Bureau.
- Goerman, P. (2006b). Cognitive interview methodology revisited: Development of best practices for pretesting Spanish survey instruments. Paper presented to the American Association for Public Opinion Research, Montreal, Canada.
- Harkness, J.A., Van de Vijver, F.J.R., & Mohler, P. (2003). *Cross-cultural survey methods*. New York: Wiley.
- Johnson, T.P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial*, 3, 1–39.
- Napoles-Springer, A., & Stewart, A. (2006). *Overview of Qualitative Methods in Research with Diverse Populations. Making Research Reflect the Population*. *Medical Care*, Vol 44, No 11, Suppl 3, November, 2006.
- Pan, Y. (2003). The role of sociolinguistics in the development and conduct of federal surveys. *Proceedings of the Federal Committee on Statistical Methodology Research Conference*. Arlington, VA.
- Pasick, R.J., Stewart, S.L., Bird, J.A., & D'onofrio, C.N. (2001). Quality of data in multiethnic health surveys. *Public Health Reports*, 116, 223 – 243.
- Schoua-Glusberg, A. (2006a). Eliciting education level in Spanish interviews. Paper presented to the American Association of Public Opinion Research, Montreal, Canada.
- Schoua-Glusberg, A. (2006b). Assessing comprehension of translated questionnaires with qualitative methods. Proceedings of the Statistics Canada Symposium 2004: Innovative Methods for Surveying Difficult-to-reach Populations. Downloaded February 2, 2007 from http://www.statcan.ca/english/freepub/11-522-XIE/2004001/Schoua-Glusberg_eng_final.pdf
- U.S. Census Bureau (2004). Census Bureau Guideline: Language translation of data collection instruments and supporting materials. Washington, DC: Author.

Willis, G., Lawrence, D., Kudela, M., & Levin, K. (2005a). The use of cognitive interviewing to study cultural variation in survey response. Paper presented to the Quest Workshop on Questionnaire Design and Testing.