

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

Kerry Levin  
Westat

Gordon B. Willis  
National Cancer Institute

Barbara H. Forsyth  
University of Maryland

Alicia Norberg  
Westat

Martha Stapleton Kudela  
Westat

Debra Stark  
Westat

Frances E. Thompson  
National Cancer Institute

We present results from a qualitative evaluation of the Spanish-language version of a dietary intake questionnaire and characterize the types of findings which emerged from several rounds of cognitive testing. Cognitive interviews were used to test the Spanish translation of the National Health Interview Survey (NHIS) Cancer Control Supplement dietary questions, with 36 Spanish-speaking and 9 English-speaking participants. Analyses of the results identified (a) translation issues, (b) culture-specific issues, and (c) general design issues that affected both English and Spanish speakers. Results indicated that general design-oriented difficulties were particularly frequent. Our findings suggest that when appropriately structured, cognitive interviews that feature flexible probing can be useful for identifying a range of problems in survey translations, even after translations have been developed using currently accepted methods. We make several recommendations concerning practices that may be optimal in the conduct of empirical cross-cultural questionnaire evaluations.

**Keywords:** Cognitive interviewing, dietary questionnaire, Spanish translation, questionnaire adapting

## Introduction

Increasingly, researchers are urged to focus on issues of cross-cultural comparability of survey measures (Hambleton, Merenda and Spielberger 2005; Johnson 1998), and to employ a variety of qualitative and quantitative techniques to develop and evaluate questionnaires (Krause 2006; Weech-Maldonado et al. 2001). In particular, producing a survey instrument that is culturally appropriate across populations or sub-groups requires subjecting the instrument to rigorous testing across language and cultural groups (Willis and Zahnd 2007). For this study, we examined the contributions of cognitive interviewing an increasingly popular qualitative evaluation method in improving a Spanish-language translation of survey questions on food intake. Our focus is on identifying and illustrating the kinds of translation and survey response issues that cognitive interviews uncover and the revisions that are indicated based on these results. A major purpose of this study is to augment existing research about using cognitive interviews to identify problems with survey translations and to develop improved versions. We examined diet items previously included in the 2005 fielding of

the National Health Interview Survey (NHIS), an ongoing, nationally representative survey conducted in the U.S. by the National Center for Health Statistics (NCHS) and administered by specially trained Census Bureau interviewers. Questions about diet are included in the Cancer Control Supplement (NHIS-CCS), administered as part of the NHIS in 1987, 1992, 2000, and 2005.

## Cognitive Interviews in Dietary and Translation Research

Cognitive interviewing has been used extensively for nearly 20 years to pretest a wide range of survey questionnaires (Conrad and Blair 2004; Willis 2005). Practitioners use cognitive interview findings to identify problems with questions and offer recommendations for improvements (U.S. Census Bureau 1998, 2004; Presser et al. 2004; Willis 2004, 2005).

Cognitive testing consists of instruction to small sets of participants (generally 8 to 12 per testing round) to respond to both open-ended and targeted verbal probe questions posed by the cognitive interviewer (e.g., "Tell me what that question was asking you;" "Can you tell me why you said your health is 'fair'?" ). In some situations, respondents are asked to "think aloud" as they answer survey questions and probes are administered throughout the interview. In other situations, verbal probing techniques take place when the respondent has completed the questionnaire instrument

---

Contact information: Kerry Levin, Westat, 1650 Research Boulevard, Rockville, Maryland 20850, e-mail: kerrylevin@westat.com

(i.e., retrospective probing). Cognitive interviewing exists as part of a series of quality-control and pretesting steps, including expert review, behavior coding, and field pretesting (Willis 2005).

Although questionnaires are sometimes evaluated through the use of psychometric techniques appropriate to scale development, such as assessment of Cronbach's alpha, or other classical psychometric approaches such as Item Response Theory (Hambleton et al. 2005), qualitative techniques such as cognitive testing are often the only feasible method for empirical pretest evaluation. In particular, for questions concerning behaviors (e.g., tobacco use, diet, and physical activity) that do not give rise to multi-item scales, quantitatively oriented psychometric techniques, such as computation of Cronbach's alpha, cannot normally be applied (DeVellis 2003 refers to these as *atheoretical* as opposed to *theoretical* constructs).<sup>1</sup> For such questionnaires, qualitative cognitive testing appears to have become the major evaluation method of choice (Willis 2005). Research on diet-related measurement suggests that cognitive testing can lead to improved approaches for estimating dietary intake (e.g. Millen, Midthune, Thompson, Kipnis and Subar 2006; Subar et al. 2001; Subar et al. 1995; Subar et al. in press; Thompson et al. 2002a; Thompson et al. 2002b).

### Cross-cultural Cognitive Interviewing

As an extension to the cross-cultural arena, researchers in the area of educational and psychological test development have advocated qualitative evaluation steps, similar to cognitive interviewing, whenever self-reported items are translated (Merenda 2005). Questionnaire designers have made several initial forays in this area (Goerman 2006; 2006a; Schoua-Glusberg 2006a; Pan 2003; Pasick, Steward, Bird and D'onofrio 2001; Napoles-Springer and Stewart 2006; Willis and Zahnd 2007). For example, Agans, Deeb-Sossa, and Kalsbeek (2006) used cognitive interviews and focus groups as part of a multistep process to develop, test, and refine culturally appropriate Spanish-language survey items to ascertain the date of last menstrual period. The authors advocated cognitive testing as a vital step in ensuring conceptual and linguistic equivalence across languages and cultures. Further, two of the studies described in Willis et al. (2005b) included items measuring dietary intake and food preparation. Those studies found that cognitive testing with translated diet items revealed linguistic distinctions that had important effects on question understanding and functional equivalence across languages, nationalities, and cultures. Based on such investigations, it is evident that cognitive interviews can detect at least some translation- or language-related issues in cross-cultural questionnaire design, and that these cover a wide conceptual range.

However, the impact of cognitive testing on survey measurement accuracy has not been demonstrated as clearly in the area of translation of questionnaires as they have been in the area of monolingual research. For example, translation researchers have reported that some respondents, and partic-

ularly those with low educational backgrounds, have difficulty understanding the cognitive probes designed to elicit information about thought processes (e.g. Pasick 2001; Goerman 2006). Further, cultural differences in communication styles may influence the effectiveness of cognitive interviews. Pan (2003) suggested that cultural differences in direct and indirect communication styles and cultural differences regarding openness in expressing opinions may affect the utility of cognitive interviews for testing questionnaire translations in some languages. In addition, cognitive interviewer experience may impact the quality and utility of cognitive interview test results (e.g., Forsyth, Kudela, Lawrence, Levin and Willis 2007; Willis, Lawrence, Kudela and Levin 2005a, Willis et al. 2005b). If multilingual cognitive interviewers have relatively little experience and/or training, the method may be ineffective because of flaws in implementation. Based in part on concerns like these, Goerman (2006, 2006a) conducted preliminary research on approaches for conducting cognitive interviews across cultures. For Spanish-language interviews, Goerman found that carefully crafted introductions that explain cognitive interview purposes can help circumvent difficulties in understanding and answering cognitive probes, but that interviewers must be allowed some flexibility in application of these.

One particular challenge within cross-cultural cognitive interviewing studies is the conceptualization of the broad range of problems identified, which tend to exceed the scope of those detected within monolingual pretesting. Several conceptualizations of these problem types have been proposed. Across multiple studies using cognitive interviews to test questionnaire translations, Willis and his colleagues identified three categories of questionnaire problems: *translation* problems where source-language questions are rendered incorrectly; *culture-specific* problems where differences in social structures, norms, viewpoints, or access to information and materials require variations in measurement approach or in question wording across languages and cultures (i.e., cultural adaptation); and *general* problems that appear to be universal, such as difficulties in comprehension of vague and poorly defined concepts (e.g., "vigorous physical activity") or in long-term recall of mundane behaviors (e.g., frequency of intake of particular foods) (Forsyth, Kudela, Lawrence et al. 2007; Kudela et al. 2006; Willis, Lawrence, Kudela and Levin 2005a; Willis et al. 2005b).

Carrasco (2003) identified similar types of issues using cognitive interviews to test a questionnaire translation, classifying problems into two general categories: *linguistic* problems reflecting culture-specific language uses and *questionnaire design* problems reflecting culture-specific questionnaire formatting (structural) issues. Schoua-Glusberg (2006b) used cognitive interviews to study culture-related response errors in translated items on educational background. Her analyses identified three sub-types of issues in mea-

<sup>1</sup> Certain psychometric measures, such as test-retest reliability, are in principle useful for the evaluation of some behavioral measures. However, these are often not feasible as a part of production survey pretesting.

suring educational background for immigrants to the U.S.: *translation* issues that arise because educational systems for the nationalities included in her study used similar words in different ways, *questionnaire design* issues related to composing standardized questions on education in multicultural studies, and *response coding* issues related to helping interviewers correctly interpret respondents' answers, particularly when interviewers and respondents have different nationalities and educational experiences.

Further, in characterizing the types of problems found through testing of English and Spanish instrument versions, Goerman and Caspar (2007) distinguish between (a) *Translation* issues; (b) *Crosscutting* findings that are common across language versions; (c) *English-only* problems, in which the functioning of the translated version is superior to that of the source English version; (d) *Combination problems*, where neither source or target versions function well, but for different reasons; and (e) *Contextual and Navigational* issues, where (for self-administered questionnaires) respondents to English and Spanish versions are differentially influenced by the visual and organizational layout of the forms. Taking a broad perspective on the concept of adaptation, Harkness, Mohler and van de Vijver (2003) distinguish between (a) features that are *non-problematic* in that they are assumed consistent across source and target language; (b) problems that are *linguistic* in nature; (c) problems having a *cultural* foundation, and (d) those that are *psychometric*. Finally, Malda, van de Vijver, Srinivasan, Transler, Sukumar and Rao (in press) have proposed a more fine-grained taxonomic approach to identifying problems in the cultural adaptation of instruments, and include adaptations that are (1) *construct-driven*, (2) *language-driven*, (3) *theory-driven*, (4) *familiarity-/recognizability-driven*, and (5) *culture-driven*. Given the wide range of potential problems, and in means for classifying these, the current investigation in part sought to empirically determine which of these varieties of difficulties surfaced as a product of cross-cultural cognitive interviewing.

## Goals

Based on previous research, we hypothesized that cognitive interviews would be an effective method for pretesting and refining the Spanish-language version of the 2005 NHIS diet questions for future administrations of the questionnaire, particularly when study protocols follow the guidance in Goerman (2006, 2006a).<sup>2</sup> In addition, we attempted to answer more specific questions concerning the evaluated dietary instrument:

1. Is the Spanish-language translation of diet questions accurate and effective?
2. Are item wordings consistently understood across Hispanic subcultures and nationalities (e.g., Caribbean, Central American, Mexican)?
3. Does the Spanish-language translation include appropriate foods as examples?
4. Are translated food terms commonly understood?

We also aimed to assess several process-related and interpretative issues:

1. What type of experience and training is necessary for bilingual cognitive interviewers?
2. Are special adaptations in techniques necessary for cognitive interviews of Spanish-speakers?
3. How well does a cognitive protocol operate for Spanish-speakers?
4. Should findings from Spanish-language interviews be characterized as uniquely related to Spanish language or Hispanic culture, or are they similar to problems identified in English-language, non-Hispanic interviews?
5. Most generally, how should the identified problems be characterized, in terms of the types of adaptations that are necessary to achieve between-version comparability?

## Method

### *Evaluated Questionnaire*

The evaluated questionnaire contained 21 questions asking frequency of consumption of a number of food items or categories of foods. A representative segment of the tested NHIS dietary questionnaire (in English) is presented in Table 1 (and is available in full from [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Survey\\_Questionnaires/NHIS/2005/English/QCA\\_NCHS.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2005/English/QCA_NCHS.pdf)). The Spanish version used in 2005 is also downloadable from the NCHS site).

### *Translating the 2005 NHIS Diet Questions*

The U.S. Census Bureau enlisted the Library of Congress (LOC) to translate the 2005 NHIS diet questions into Spanish. In accordance with practices that focus on team-based, forward translation that are increasingly advocated in the survey methods field (Census Bureau 2004; Harkness and Schoua-Glusberg 1998; Harkness, Pennell and Schoua-Glusberg 2004; McKay et al. 1996), the survey sponsors held a Spanish Translation Review Conference to review the initial translation, and to make decisions about translation wordings.<sup>3</sup> Participants in the Review Conference included bilingual field interviewers from the Census Bureau, methodologists, field operations staff, substantive experts and translation specialists from NCHS and the Census Bureau, and a researcher with translation adjudication experience from the National Institutes of Health (NIH). Participants worked as a team to discuss identified translation

<sup>2</sup> Although the 2005 NHIS Cancer Supplement was translated and fielded in Spanish, no cognitive or other evaluative testing was done of the Spanish version, and no information was available concerning operation of the questionnaire within the field environment. As such, the current investigation endeavored to more intensively investigate the functioning of that instrument.

<sup>3</sup> Although the practice of back-translation of instruments has been used extensively, it appears to be falling out of favor, especially as the sole process used to evaluate survey questionnaires.

Table 1: Selected dietary items from NHIS 2005

Item wording
1. During the past month, how often did you eat HOT OR COLD CEREALS?
2. How often did you have MILK, either to drink or on cereal? Do NOT include small amounts of milk in coffee or tea.
3. How often did you eat FRUIT? COUNT fresh, frozen, or canned fruit. Do NOT count juices.
4. How often did you have TOMATO SAUCES such as spaghetti sauce or pizza with tomato sauce?
Response Format
[ ] Never; [ ] 1-3 times last month; [ ] 1-2 times per week; [ ] 3-4 times per week; [ ] 5-6 times per week; [ ] 1 time per day; [ ] 2 times per day; [ ] 3 times per day; [ ] 4 times per day; [ ] 5 or more times per day; [ ] Refused; [ ] Don't know

issues, articulate options for addressing them, and make decisions about final translation wording.

Prior to the cognitive testing step, a further review of the existing translation was conducted under the supervision of our project team. Following the guidelines used by the Census Bureau (2004), (see also Forsyth et al. 2007), two independent reviewers examined the translation of the NHIS dietary questions which had resulted from the earlier Review Conference, to review and document potential issues that had been overlooked. Both reviewers were fluent in English and Spanish, were experienced survey researchers, and had different Spanish-language backgrounds. The reviewers identified a small number of concerns with the translation, (e.g., the observation that in Spanish, 'abase' should be corrected to 'a base') and then submitted this documentation to an experienced, independent adjudicator who made final decisions about revisions prior to cognitive testing.<sup>4</sup> The adjudicator made relatively minor additional changes prior to cognitive testing. Overall, the problems that were identified and presumably addressed within the initial translation, review, and adjudication steps varied. Issues included the selection of improper terms as Spanish equivalents to English, grammatical, syntactic and spelling errors, and omissions of words in the stem of the question or the answer categories. Cognitive testing was conducted following these steps in an attempt to ensure that these problems had been resolved, and to identify more subtle problems that might emerge through the process of intensive interviewing of members of Spanish-speaking subcultures.

### *Cognitive Testing Design and Methodology*

Three iterations of testing were conducted:

1. A procedural pretest
2. Round 1 cognitive interviews
3. Round 2 cognitive interviews

For the procedural pretest, nine cognitive interviews, conducted in Spanish in the Washington DC area, served as a training vehicle for the interviewers and as a source of data for refining the cognitive interview protocol. After minor modifications to the cognitive test protocol following the procedural pretest, we conducted two iterative rounds

of cognitive interviews. Round 1 consisted of interviews with 18 Spanish-speaking participants, in San Jose, California, and Miami, Florida. Round 2 consisted of 18 more interviews, following review and question revision based on Round 1 results: 9 with Spanish-speaking participants, and 9 with English speakers. Round 2 interviews were conducted in the Washington, DC area. We included English speakers in Round 2 primarily to test the English-language version of the NHIS diet questions. As described below, it seemed important to include a test of the English-language questions because the procedural pretest and Round 1 interviews revealed general issues with the translated questions that appeared to transcend language or culture, and English-language interviews therefore provided an opportunity to determine whether similar issues emerged with the English-language questions.

### *Interview Participants*

For Spanish-speaking interviews, recruiting activities were conducted in Washington DC, Miami, and San Jose. In each city, firms providing research support services screened, recruited, and scheduled interviews with Spanish-speaking volunteers. The study design explicitly included participants from three cities in different regions of the U.S. to ensure that we included individuals with a range of Hispanic/Latino national backgrounds. Each firm recruited Spanish speaking volunteer participants who were born somewhere other than the U.S. and who represented a mix of ages, gender, education levels, and language use (i.e., monolingual versus bilingual). Although some researchers have proposed that translation equivalence be established by administering items in two languages to bilinguals, we decided this approach would be problematic because bilinguals may understand and process language differently than do monolinguals do (Blais and Gidengil 1993; Ellis et al. 1989). We were unable to assess language skills in advance of the interview, beyond participants' self-report of whether they spoke Spanish exclusively, or at least, spoke more Spanish than English. Because

<sup>4</sup> Note that this procedure differed slightly from that advocated by the Census Bureau (2004), which recommends a combined review and adjudication step.

*Table 2:* Demographic characteristics of cognitive interview participants

	Language spoken by participants	
	Spanish	English
<i>Gender</i>		
Male	17	5
Female	19	4
<i>Age</i>		
18-29	15	1
30-39	10	1
40-49	6	2
50+	5	5
<i>Education completed</i>		
Less than high school	12	1
High school	15	-
Some college	6	1
Bachelor's degree or greater	3	7
<i>National background</i>		
Mexico	16	-
Cuba	10	-
El Salvador	4	-
Guatemala	2	-
Honduras	2	-
Dominican Republic	1	-
Puerto Rico	1	-
Hispanic (background not reported)	-	1
Non-Hispanic	-	8
<i>Total years residing in United States</i>		
1 year and less	10	-
Between 1 and 5 years	8	-
6-10 years	3	-
11-20 years	3	-
21-30 years	5	-
More than 30 years	7	-
n	36	9

some researchers assert that using a measure of acculturation to U.S. society is important when examining differences across culturally-diverse groups (Bethell et al. 2003; Willis and Zahnd 2007), we measured acculturation level, in rough terms, as the reported length of time living in the United States.

We recruited English-speakers from a proprietary database of study volunteers. All English-speaking volunteer participants were born in the U.S., though one of the English-speakers reported an Hispanic ancestry. Table 2 summarizes participant demographic characteristics.

### *Cognitive Interview Procedures*

Senior members of the project team with backgrounds in survey methods and qualitative research trained three bilingual cognitive interviewers to administer the questionnaire and the cognitive interview protocol. The selected cognitive interviewers were skilled qualitative researchers with survey design experience; all were fluently bilingual in Spanish and English and reflected Mexican and Central American backgrounds. All three had previous experience in conducting questionnaire evaluation independently, in off-site locations. One of the interviewers, who had significant questionnaire translation and cognitive interview analysis skills, served as the interviewing team leader. The team leader coordinated all interviewing and reporting activities, conducted roughly half of the cognitive interviews, and assisted in monitoring interviews conducted by the other cognitive interviewers.

Building on Goerman's (2006, 2006a) guidelines, cognitive interviewer training stressed the importance of administering the tested survey questions as worded, while remaining flexible when administering cognitive interview probe questions. That is, participants were asked to answer (orally) the NHIS diet questions as read to them, and the cognitive probes administered were either pre-scripted, devised spontaneously at the time of the interview, or "emergent" in that they were created to react to clear indications of problems (see Willis 2005, for a full description of this probing taxonomy). Training included several role-play exercises to practice cognitive interviewing skills. In addition, throughout all testing rounds, the senior bilingual team leader monitored ongoing interviews and provided feedback to interviewers. Interviews took place in cognitive laboratory facilities at various recruitment facilities across the country. Participants received \$60 incentive payments. Interviews lasted 60 to 90 minutes and followed the cognitive interview protocol (which was approved by Westat and recruitment firms' Internal Review Boards (IRB) and determined Exempt from IRB review at NIH). The protocol consisted of the following elements:

1. *Interview introduction.* Interviewers explained the purpose of the project and the nature of the cognitive interviewing procedures to be conducted (e.g., that the interview would consist of think-aloud and interviewer-based probing, that the overall purpose was to identify defects in the tested instrument, and that the participant should feel uninhibited in expressing difficulties). Interviewers assured participants that all information would be treated as confidential. Also, interviewers requested permission to audiotape interviews and, when applicable, informed participants that NCI staff would remotely observe the interview. Interviewers reminded participants that they could refuse to answer any questions and that they could end the interview at any time. Participants provided written informed consent.
2. *Administration of the draft questionnaire.* Participants answered the NHIS diet questions.
3. *Cognitive probing.* In order to identify potential "silent misinterpretations" (DeMaio and Rothgeb 1996) that

were not otherwise in evidence, interviewers administered probes in the cognitive interview protocol that assessed how participants interpreted key questions or phrases. In addition to proactive probing of such pre-identified concrete concerns (Cosenza 2002), cognitive interviewers also followed up by probing participants who had any observed confusion or difficulties answering the questionnaire items. Interviewers administered most probes concurrently with the diet questions (as described by Forsyth and Lessler 1991; Willis 2005). Between testing rounds, scripted probes were revised when the translations into Spanish were found confusing for participants. For example, probes that asked “What do you think this question is about?” and “What time frame were you thinking about?” were unclear, and were reworded for the second round of cognitive interviews with participants. The reworded probes that were much better understood included statements such as, “Tell me more,” and “Why did you answer like that?”

After each interview, interviewers reviewed the audiotapes of their sessions and prepared detailed summaries, using a standard summary template designed by senior team members to ensure complete reporting and to facilitate subsequent analyses.

## Results

### *Analytic Approach*

Because cognitive interviews primarily produce qualitative data, interview summaries drafted by cognitive interviewers or transcripts serve as important data sources for analysis. Interview summaries from this study were used as the primary source for the analysis. Findings from cognitive interviews may result in recommendations to (a) modify the English version of the survey instruments (sometimes referred to as ‘decentering’; see Hambleton 2005), by either adding domains to capture the experiences of Latino participants or modifying the construction of items in English to make them more translatable into Spanish; (b) modify the wording of items within the Spanish version to accommodate ethnic and regional variations in Spanish language use; or (c) simplify the translation so the complexity level of the document is appropriate for the target population (Weech-Maldonado 2001). For the translation of the NHIS 2005, ultimate decisions for changes in both English and Spanish resulting from the cognitive testing were client-driven.

Analyses focused on item-specific results, combined across interviews within each testing round of interviewing. Given that survey items tested in the Procedural Pretest and Round 1 interviews were identical, and because very few changes were made to protocol procedures following the Procedural Pretest, results from these first two steps were combined. Analyses focused on Round 2 interviews separately, and for those interviews, analyses focused separately on Spanish- and English-language results. Table 3 summarizes our two-step analysis process. Analysis Step 1 ap-

plied qualitative data reduction methods (Miles and Huberman 1994) to review interview summaries and develop item-level summaries. The Step 1 item-level summaries constituted the primary data for Analysis Step 2, which relied on content analysis (i.e., organizing issues and problems into meaningful groupings) to identify categories of item problems within interview rounds and to develop item revisions (e.g., Silverman 1993; Patton 2002; Miles and Huberman 1994).

In Step 2, analysts identified item issues or problems, and considered how these clustered into particular categories of questionnaire defects. Based on this process, identified problems were classified into three unique categories (and one combination): *Translation* problems, *Culture-specific* problems, *General design* problems; and *Mixed* problems that appeared to be caused by a combination of the other subtypes (Table 4). Following Willis and his colleagues (Forsyth et al. 2007; Kudela, Forsyth, Levin, Lawrence and Willis 2006; Willis, Lawrence, Kudela and Levin 2005a; Willis et al. 2005b), analysts identified a *Translation* problem whenever translated item wording altered the intent of the original question, a *Culture specific* problem whenever an item’s intended meaning was difficult to convey to subjects from Hispanic sub-cultures, and a *General design* problem when questions presented difficulties that appeared independent of culture or language.

Table 4 presents frequencies of each category of problems, by cognitive testing round. Four results are noteworthy. First, the Pretest and Round 1 Spanish-language interviews revealed numerous problems<sup>5</sup>: For the 21 items tested, we identified an average of approximately three problems per item. Second, roughly three-quarters of the problems identified in early interviews were General design problems that we judged not to be specific to culture or language.<sup>6</sup> Third, revisions made to both the English and Spanish instruments after the first round of cognitive interviews reduced the number of problems identified in the second round of Spanish-language interviews. This reduction was especially pronounced for General problems. Further, revisions between the two rounds of cognitive interviews nearly eliminated translation problems. Fourth, in the second round of testing, the English- and Spanish-language versions of the dietary questionnaire were roughly comparable in terms of the total number of problems identified. This latter result suggests that revisions based on the first set of Spanish cognitive

<sup>5</sup> We tested in Spanish but not in English in Round 1, based on the expectation that observed problems would be specific to Spanish-speakers. Given the preponderance of General design problems within that round, we determined this to have been an incorrect assumption.

<sup>6</sup> A reviewer has pointed out that, based on Spanish-only testing within Round 1, we could not confidently make the assessment that these problems were General in nature and would also afflict non-Hispanics. It was in part for this reason that we decided to conduct interviews, within Round 2, of English-speakers as well. Based on the total set of results obtained over both rounds, we felt that we could then, in retrospect, label some Round 1 results as General in nature.

Table 3: Overview of two-step qualitative analytic process

Step	Analytic methods	Analytic products	Goal
Step 1	Qualitative data reduction	Item-level summaries by interview round	To facilitate subsequent content analysis
Step 2	Content analysis	Pass 1: Identification of items with problems and/or wording issues	To determine general magnitude of problems and issues evident in each round of interviews
		Pass 2: Classification of problems into three general categories: Translation problems, Culture-specific problems, and General design problems	To characterize general types of issues found and assess prevalence of each type in each round of interviews
		Pass 3: Characterization of problems more specifically, producing more detailed description of problem	To distinguish problems more precisely for purposes of item revision and communication

Table 4: Number (and percent) of problems identified in cognitive interviews over all 21 tested questions, by type of problem, testing round, and language

Testing round	Type of problem identified				
	Translation	Culture specific	General design	Mixed	Total
Pretest + Round 1: Spanish	9 (14.8%)	3 (4.9%)	46 (75.4%)	3 (4.9%)	61 (100%)
Round 2: Spanish	2 (13.3%)	2 (13.3%)	9 (60.0%)	2 (13.3%)	15 (100%)
Round 2: English	0 (0.0%)	0 (0.0%)	15 (93.8%)	1 (6.3%)	16 (100%)

interviews enhanced comparability between the English- and Spanish-language questionnaires.

The following sections present more detailed examples of each unique problem type, as well as the types of recommendations for modification that were made as a result, across both testing rounds.

### Translation Issues

Analysis revealed two general sub-types of translation problems:

1. Words did not convey intended constructs when translated from English to Spanish.
2. A Spanish term was not equally familiar, or had different meaning, across nationalities, regions, or cultures.

Table 5 contains examples illustrating these translation problems, and the revisions made to eliminate them. Most of these problems were relatively simple to resolve by selecting alternative translation wordings or by refining or restructuring translation wordings.

### Culture-Specific Issues

Analysts identified three general types of culturally-specific problems; that is, those that appeared to afflict one or more Hispanic sub-cultures, but not non-Hispanics, and represent failures of adaptation. More specifically, these problems occurred when the question involved concepts, knowledge, or behavior that differ between cultures or nationalities (see Table 6). Overall, we detected relatively few culture-specific issues. Even for these, however, we were seriously restricted in making adaptations to resolve these problems, as

this would involve nontrivial changes to the English version, and the client was reluctant to significantly alter the English-language questionnaire. That is, decentering was severely restricted, especially given the use of sequential rather than simultaneous questionnaire development (see Tanzer 2005 for a further description of these approaches).

### General Design Issues

Finally, analysts found a variety of general problems that affected both Spanish- and English-speaking participants. Table 7 contains examples of key problems, and the (limited) revisions applied to address them. Most, but not all, of these problems were identified in Round 1, and led to revisions that were then further evaluated in the second round. Several approaches helped to address these issues, including rewording items, revising interviewer instructions, and explicitly stating the item reporting period as part of each item. Paradoxically, the survey client recognized the severity of many of these problems, and the fact that they likely affect not only Spanish but English-speaking respondents, but was again reticent to enact changes to either version, based on the desire to maintain comparability with previous survey administrations using the instrument. A compromise was reached in which minor alterations were made to both English and Spanish questionnaires to improve the questions generally, while maintaining cross-cultural comparability.

### Discussion

Results from cognitive testing identified translation, culture-specific, and general design problems that seem

Table 5: Examples of Translation issues identified in cognitive interviews

Translation issue		Illustrative examples
Spanish words do not convey intended construct	<i>Problematic Translation:</i> Salsa for a “picante”-type tomato sauce	<i>Description:</i> “Salsa” is a general term equivalent to “sauce” in English. The original translation literally asked about “sauce containing fruits or vegetables” and did not convey the intended meaning of a picante-like sauce. As a result, subjects in Miami and California reported thinking about marmalade, apple sauce, and fruit sauce for topping ice cream. <i>Revision:</i> In Spanish version, place “salsa” item ahead of additional items on tomato sauces; Remove reference to “fruits or vegetables”; Include descriptions to define salsa, including “salsa picante o pico de gallo, o tipo Mexicana.” (That is: “spicy (hot), pico de gallo or Mexican style.”)
	<i>Problematic Translation:</i> Carne de res for “red meat”	<i>Description:</i> “Carne de res” translates literally as “beef”, a narrower construct than the intended one (“red meat”). Participants were therefore confused by question instructions to include veal, pork, lamb, and cold-cuts. <i>Revision:</i> Remove reference to “res” (“beef”) and rephrase as “carne roja” (“red meat”).
Spanish words are unfamiliar or have different meanings in some regions or to some nationalities	<i>Problematic Translation:</i> Refrigerios for “snack”	<i>Description:</i> “Refrigerios” (“snack”) was familiar to participants in the Washington, DC area (mainly Central American and Mexican) but generally unfamiliar to Cuban and Caribbean participants in Florida, and to Mexicans in California. “Bocados” (“snack”) was more familiar to participants across regions and nationalities. <i>Revision:</i> Use “bocados” in place of “refrigerios.”
	<i>Problematic Translation:</i> Galletas and torta for “cookies” and “cakes”	<i>Description:</i> “Galletas” can mean either cookies or salty crackers. Also, for Mexican participants, “torta” can mean a sandwich. <i>Revision:</i> Move “galletas” and “torta” toward end of question on cakes and cookies, and use context to further clarify the intended meaning.

likely to interfere with intended survey measurement goals. Notably, the process of cognitive interviewing of Hispanics, in Spanish, presented no persistent obstacles that were not also exhibited in English-language interviews. The evident success using cognitive interviewing to identify problems with translated survey items is important, because researchers have previously reported mixed results (Goerman 2006; Pan 2003; Pasick et al. 2001). At least two factors probably contributed to successes reported here. First, the cognitive interview protocols followed preliminary guidelines recommended by Goerman (2006, 2006a) to convey the overall purpose, and procedures to be used, in the cognitive interview. In particular, interviewers were trained to emphasize that the purpose of testing is the evaluation of questions, as opposed to collection of individual dietary data.

Second, cognitive interviews in Spanish may avoid some problems associated with studies of Asians noted by Pan (2003), especially because Spanish-speaking cultures may be relatively close to the U.S. along the continuum of social directness, a factor Pan identified as likely to be important in pretest method effectiveness. On the other hand, Willis and Zahnd (2007) reported that cognitive interviewing tech-

niques appeared to function well for Korean speakers, including those who are monolingual and relatively unacculturated to U.S. society. Future comparisons between Spanish-language and Asian-language results using cognitive interviews may shed additional light on how cultural group membership and acculturation affect cognitive interview results.

### *Recommendations for Future Practice*

Based on our results, we make several general recommendations for using cognitive interviews to test survey translations. First, based on the impression that our selection of very experienced Spanish-language interviewers proved beneficial, we advise selection of seasoned cognitive interviewers who are fluent in both the target and source languages, and who have demonstrated ability to build rapport with participants. In this context, rapport, as described by Schaffer and Riordan (2003), refers to the participants’ confidence in the researcher, their overall comfort level with the researcher, and/or their willingness to cooperate with the procedures associated with the survey instrument. We realize that the proposal to utilize experienced interviewers who are

Table 6: Examples of Culture-specific issues identified in cognitive interviews

Cultural issue	Item content	Illustrative examples
Definition of concepts or knowledge differ across cultures or nationalities	<i>Problem:</i> Knowledge of cereal fiber content	<i>Description:</i> Spanish-speaking participants reported that they had never thought before about fiber content of cereals. <i>Revision:</i> None: Fully addressing issue would require significant modification to the English-language version, which was not allowable.
	<i>Problem:</i> Reference to “white potatoes”	<i>Description:</i> Spanish-speaking cultures commonly eat a variety of potatoes that are not white in color but that have nutrition profiles similar to white potatoes. <i>Revision:</i> For the Spanish language version, remove “blancas” (“white”); retain instruction in all versions to include potatoes like red-skinned and Yukon Gold potatoes.
	<i>Problem:</i> Knowledge of cereal brand names	<i>Description:</i> Participants who were recent immigrants were unfamiliar with many of the brand names used to help define the response categories in the cereal item. <i>Revision:</i> None, as revising the Spanish questions to address new immigrants’ needs would require revising item goals and English-language measurement approach.
	<i>Problem:</i> Familiarity with presented examples	<i>Description:</i> For a few items, Hispanic participants found examples to be unfamiliar. <i>Revision:</i> Add examples to the Spanish version that are likely to be familiar for items on fruit juice (papaya), other vegetables (yucca) and cereals (avena).

Table 7: Examples of General design issues identified in cognitive interviews

General design problem	Illustrative example
Food category definition was insufficient	<i>Description:</i> (a) Participants 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; (b) Participants interpreted “100% fruit juice” as excluding juices made from concentrate; and (c) “fruit-flavored drinks” confused both Spanish- and English-speaking participants, as this seemed redundant with “fruit juice.” <i>Revision:</i> For both language versions, remove transitional statement, “We’re next going to ask you about three different types of beverages you might drink: soda, fruit juices and fruit drinks.”
Supplementary definition is needed	<i>Description:</i> Several items included examples that interviewers read as necessary to help define food categories (e.g., READ IF NECESSARY: “Incluya leche desnatada, descremada, baja en grasa, leche entera, suero de leche y leche sin lactosa. Incluya también leche de chocolate u otras leches con sabores agregados.”/ “Include skim, no-fat, whole milk, buttermilk, and lactose-free milk. Also include chocolate or other flavored milks.”). When interviewers did not read these examples, Spanish and English speaking subjects routinely omitted foods that they should have included. Participants found the instructions helpful when interviewers read them. <i>Revision:</i> For both versions, revise interviewer instructions so they always read defining examples.
Item reporting period is forgotten	<i>Description:</i> Interviewer instructions asked interviewers to repeat item-reporting periods as necessary. When interviewers did not repeat the reporting period, subject reports were based on idiosyncratic and variable time frames. For example: “the past week,” “the past 2 years (since the participant arrived in the U.S.)”, “since becoming pregnant,” and “yesterday.” <i>Revision:</i> For both versions, include “during the past month” in all question wordings so interviewers will read it each time.
Question wording elicits uncodeable response	<i>Description:</i> The question wording “con qué frecuencia”/ “how often” elicited verbal responses (e.g., “not too often,” “hardly ever,” “all of the time”) rather than the intended numerical frequency responses. <i>Revision:</i> Reword all questions to ask, “Durante el mes pasado ¿cuantas veces por día, semana o mes, bebió/comió” / “During the past month, how many times per day, per week or per month did you drink/eat...”

adaptive, as opposed to novices who are simply trained to 'follow a script' as they probe, can often present a 'needle in a haystack' challenge with respect to interviewer recruitment. However, this challenge may dissipate with the ongoing development of multilingual cognitive interviewing capacity within the survey pretesting field (which increasingly appears to be the case for Spanish, and is perhaps at an incipient point for some Asian languages).

A second recommendation, consistent with Goerman (2006; 2006a), is based on the observation that Hispanic participants are amenable to the cognitive testing task once it is explained clearly. The interview introduction should include descriptions of the purposes of the cognitive interview and the cognitive interview probes, including example exercises to make the descriptions concrete. However, the investigators must be flexible and prepared to modify the translations of scripted probes that are found to be ineffective or confusing to participants. To this end, we found it to be useful to include a procedural pretesting round of cognitive interview interviews, to evaluate both the protocol and interview procedures.

Our third recommendation is based on the finding that some translation issues became evident only because the cognitive interview design included participants with different regional and national backgrounds. For languages like Spanish, where there are likely to be regional variations or dialects, it is important to ensure that cognitive testing research includes participants from a range of these backgrounds even when overall sample sizes are small.

Fourth, although our cognitive interviews revealed relatively few culture-specific problems, we do not want to discount these, as other investigations have found them to be important sources of survey non-equivalence (Miller, Willis, Eason, Moses and Canfield 2005; Warnecke et al. 1997). There are several potential explanations for the relative paucity of strictly cultural factors in our results (i.e., those that differentially affected Hispanics, but that were not primarily due to translation or linguistic factors). Perhaps the team-based translation approach used to develop the initial translations anticipated cultural issues and made translation choices that rectified the bulk of these. It is also possible that the diet questionnaire did not present the opportunity for strong cultural divergence, although we suspect that questions on dietary intake are ripe for such effects, given the wide variation in dietary practices between Hispanics and non-Hispanics. Alternatively, it is possible that our recruitment and interviewing procedures were not effective for identifying inherent culture-specific issues. For example, our Hispanic participants may have been relatively acculturated to U.S. society compared to the general Spanish-speaking household population, and may have therefore experienced fewer problems related to a unique Hispanic dietary perspective.

The predominance of General design problems identified by our interviews was somewhat surprising, given that English and Spanish-language versions of the evaluated NHIS dietary items have been fielded previously, on a large scale. In general, the significant number of General design

problems observed in our interviews reflect a few major issues that each affected many items. To reiterate, nearly all items tested used a "how often" question stem which was found to be generally problematic. Nearly all items also included instructions to read reporting periods only as needed (and we found that the instructions were in fact necessary); and many of the tested items included defining examples to be read at the interviewer's discretion (and again, we found reading these to be very helpful).

We suggest three hypotheses as to why we observed these persistent general problems within previously fielded diet questions. First, perhaps some of these problems may not be evident when the same questions are administered in field settings. For example, interviewers may be unaware when participants focus on inappropriate reference periods to answer survey questions (e.g., they experience silent misinterpretation). Further, we observed instances of misinterpretation of key terms and phrases that interviewers would not be expected to identify simply on the basis of administering the question (e.g., as indicated in Table 5, misinterpretations occurred for "100% juice," "salsa," and "white potato"). Again, these problems only emerged once probing was conducted within the cognitive interviewing context. Second, well-trained field interviewers may intervene to rectify these problems. For example, when respondents provide uncodeable responses to items that ask "how often," interviewers may assist them by probing for the number of times per day, per week, or per month. Or, if interviewers intervene and respondents learn the response format quickly, then the "how often" stem may pose a problem for relatively few items.

Third, it is of course possible that problems identified using cognitive interview methods may not be problems at all in standard field settings. It should be possible to test these alternative hypotheses in an experiment using field interview observation methods such as behavior coding (Cannell, Lawson and Hausser 1975; Cannell, Oksenberg and Kalton 1991). Researchers could administer either original or revised questions to two groups of respondents, and observation of both respondent and interviewer behaviors would indicate whether the original translations are more difficult for interviewers to administer or more difficult for respondents to answer, compared to the revised translations.

### *Caveats*

This last point suggests a general caveat for the results reported here. We considered issues identified by cognitive testing as indicators of response error associated with the Spanish-language translation of the evaluated NHIS dietary questionnaire. Although it is clear that these issues were problems for the cognitive interview subjects, we do not possess data verifying that issues we identified through cognitive interviews produce actual problems in fielded interviews. Again, behavior coding and interaction analysis may provide alternate means to assess conceptual equivalence of measures across diverse racial/ethnic groups, extending the utility of the cognitive interview pretest method along a continuum of qualitative analysis (Napoles-Springer et al. 2006).

Future administrations of the NHIS diet questions will provide opportunities to monitor interviews, and to determine whether issues identified in cognitive interview testing also predict survey responses or survey response errors.

A second limitation to the current study is that we were constrained by the focus on an existing questionnaire, in English, and designed mainly for non-Hispanic respondents. This constraint allowed us to ascertain the function of this instrument on Hispanics, but only in a limited sense. We were unable to begin from the more basic vantage point of considering whether the constructs represented by the questionnaire (and not just the established individual survey questions) are appropriate for Hispanics. Optimally, development of a dietary (or other behavioral) question set would begin at an earlier point, in a way that included ethnographic interviewing (Gerber 1999) or the conduct of focus groups, prior to scripting of the questions. It is possible that a questionnaire that is appropriate for Hispanics, and that takes into account unique cultural views or behaviors with respect to diet, would take a different form than the NHIS dietary supplement.

A further limitation of our study is that we cannot clearly characterize the demographic differences between our nominally Hispanic and non-Hispanic groups. Due to small sample sizes, and resultant confounding of variables that may influence the survey response process (in particular, educational level), it is not clear that variation in response between groups was in fact due only to language or ethnicity. Note, however, that despite the potential for multiple sources of variation between groups, a compelling finding was with respect to similarity between these groups (i.e., preponderance of General design problems), as opposed to dissimilarity.

Even given the inherent limitations of the design of the study and the background constraints imposed by the need to rely on a relatively invariant instrument, we conclude that there is merit in applying cognitive interviewing to improve a questionnaire that might otherwise contain both obvious and subtle errors once converted to Spanish (and presumably when translated to other languages as well). As suggested previously by Schoua-Glusberg (2006a), cognitive testing is likely to locate outright errors, to enhance comprehensibility, and to improve question flow, even when translations have already been developed using a systematic, team-based, multi-step process that reflects currently accepted practices (e.g., Census Bureau 2004; Harkness and Schoua-Glusberg 1998; Harkness et al. 2003). Overall, the cognitive interview has a clear role as a vital supplement to previously conducted formal translation procedures. That is, cognitive testing locates translation errors that are otherwise overlooked, and identifies the need for cultural adaptations that cannot be achieved through minor wording changes. In fact, the process of cognitive interviewing itself may motivate clients and designers to consider a degree of latitude in making appropriate adaptation that extends far beyond the level they are normally comfortable with, within the question translation process.<sup>7</sup> Finally, especially with additional methodological development, cognitive testing promises to be a generally adaptive technique that investigates a range of linguistic, socio-cultural, logical, and cognitive factors, and can efficiently

identify not only defects, but also point the way toward resolution of these problems within cross-cultural investigations.

## References

- Agans, R. P., Deeb-Sossa, N., & Kalsbeek, W. D. (2006). Mexican immigrants and the use of cognitive assessment techniques in questionnaire development. *Hispanic Journal of Behavioral Sciences*, 28, 209-230.
- Bethell, C., Carter, K., Lansky, D., Latzke, B., & Gowen, K. (2003). *Measuring and Interpreting Health Care Quality Across Culturally-Diverse Populations: A Focus on Consumer-Reported Indicators of Health Care Quality*. Portland, OR: FACCT - Foundation for Accountability.
- Blais, A., & Gidengil, E. (1993). Things are not always what they seem: French-English differences and the problem of measurement equivalence. *Canadian Journal of Political Science*, 26, 541-555.
- Cannell, C. F., Lawson, S. A., & Hauser, D. L. (1975). *A technique for evaluating interviewer performance*. Ann Arbor, MI: Survey Research Center, University of Michigan.
- Cannell, C. F., Oksenberg, L., & Kalton, G. (1991). New techniques for pretesting survey questions. *Journal of Official Statistics*, 7, 349-365.
- Carrasco, L. (2003). *The American Community Survey (ACS) en Español: Using cognitive interviews to test the functional equivalency of questionnaire translations*. Washington, DC: Statistical Research Division, U.S. Census Bureau.
- Conrad, F., & Blair, J. (2004). Aspects of data quality in cognitive interviews: The case of verbal reports. In S. Presser et al. (Eds.), *Questionnaire Development, Evaluation and Testing Methods* (p. 67-87). New York: John Wiley and Sons.
- Cosenza, C. (2002). *Not your grandparent's cognitive testing: Exploring innovative methods in cognitive evaluations of questions*. Paper presented at the International Conference on Questionnaire Development, Evaluation and Testing (QDET), Charleston, SC.
- DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (p. 177-195). San Francisco: Jossey-Bass.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Ellis, B., Minsel, B., & Becker, P. (1989). Evaluations of attitude survey translations: An investigation using item response theory. *International Journal of Psychology*, 24, 665-684.
- Forsyth, B. H., Kudela, M. S., Lawrence, D., Levin, K., & Willis, G. (2007). Methods for translating an English-language survey questionnaire on tobacco use into Mandarin, Cantonese, Korean and Vietnamese. *Field Methods*, 19, 264-283.
- Forsyth, B. H., & Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathoiwetz, & S. Sudman (Eds.), *Measurement errors in surveys* (p. 393-418). New York: Wiley.
- Gerber, E. R. (1999). The view from anthropology: Ethnography and the cognitive interview. In M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (p. 217-234). New York: Wiley.
- Goerman, P. (2006a). *Adapting cognitive interviews for use in*

<sup>7</sup> We thank an anonymous reviewer for suggesting this point.

- pretesting Spanish language instruments*. 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 at the meeting of the American Association for Public Opinion Research, Montreal, Canada.
- Goerman, P., & Caspar, R. (2007). *A new methodology for the cognitive testing of translated materials: Testing the source version as a basis for comparison*. <http://www.amstat.org/sections/srms/Proceedings/y2007f.html>.
- Hambleton, R. K. (2005). Issues, designs, and technical guidance for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (p. 3-38). Mahwah, New Jersey: Erlbaum.
- Hambleton, R. K., F. Merenda, P., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, New Jersey: Erlbaum.
- Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. A. Harkness (Ed.), *Cross-cultural survey equivalence* (p. 87-127). Mannheim, Germany: ZUMA Nachrichten Spezial.
- Harkness, J. A., Mohler, P. P., & Van de Vijver, F. J. R. (2003). Comparative research. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (p. 3-16). Hoboken, NJ: John Wiley & Sons.
- Harkness, J. A., Pennell, B.-E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (p. 453-473). New York: Wiley.
- 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* (No. 3). Mannheim, Germany: ZUMA Nachrichten Spezial.
- Krause, N. (2006). The use of qualitative measures to improve quantitative measures of health-related constructs. *Medical Care*, 44(11), S34-8.
- Kudela, M. S., Forsyth, B. H., Levin, K., Lawrence, D., & Willis, G. (2006). *Cognitive interviewing versus behavior coding*. Paper presented to the American Association of Public Opinion Research, Montreal, Canada.
- Malda, M., van de Vijver, F., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (n.d.). *Adapting a cognitive test for a different culture: An illustration of qualitative procedures*. *Psychology Science Quarterly*.
- McKay, R. B., Breslow, M. J., Sangster, R. L., Gabbard, S. M., Reynolds, R. W., Nakamoto, J. M., et al. (1996). Translating survey questionnaires: Lessons learned. *New Directions for Evaluation*, 70, 93-105.
- Merenda, P. F. (2005). Cross-cultural adaptation of educational and psychological testing. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (p. 321-341). Mahwah, New Jersey: Erlbaum.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.
- Millen, A. E., Midthune, D., Thompson, F. E., Kipnis, V., & Subar, A. F. (2006). The National Cancer institute diet history questionnaire: Validation of pyramid food servings. *American Journal of Epidemiology*, 163, 279-288.
- Miller, K., Willis, G., Eason, C., Moses, L., & Canfield, B. (2005). Interpreting the results of cross-cultural cognitive interviews. In J. H. P. Hoffmeyer-Zlotnik & J. Harkness (Eds.), *Methodological aspects in cross-national research* (p. 79-92). Mannheim, Germany: ZUMA Nachrichten Spezial.
- Napoles-Springer, A., Santoyo-Olsson, J., O'Brien, H., & Stewart, A. (2006). Using cognitive interviews to develop surveys in diverse populations. *Medical Care*, 44(11), S3-S4.
- Napoles-Springer, A., & Stewart, A. (2006). Overview of qualitative methods in research with diverse populations. Making research reflect the population. *Medical Care*, 44(11), S3-S4.
- 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. November 2003, 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.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Rothgeb, J. (2004). Methods for testing and evaluating survey questions. In S. Presser et al. (Eds.), *Methods for testing and evaluating survey questionnaires* (p. 1-22). New York: Wiley.
- Schaffer, B., & Riordan, C. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, 6(2), 169-215.
- Schoua-Glusberg, A. (2006a). *Assessing comprehension of translated questionnaires with qualitative methods*. Proceedings of the Statistics Canada Symposium 2004: Innovative Methods for Surveying Difficult-to-Reach Populations. [http://www.statcan.ca/english/freepub/11-522-XIE/2004001/Schoua-Glusberg\\_eng\\_final.pdf](http://www.statcan.ca/english/freepub/11-522-XIE/2004001/Schoua-Glusberg_eng_final.pdf).
- Schoua-Glusberg, A. (2006b). *Eliciting education level in Spanish interviews*. Paper presented to the American Association of Public Opinion Research, Montreal, Canada.
- Silverman, D. (1993). *Interpreting qualitative data*. Thousand Oaks, CA: Sage.
- Subar, A. F., Thompson, F. E., Kipnis, V., Midthune, D., Hurwitz, P., McNutt, S., et al. (2001). Comparative validation of the Block, Willett and National Cancer Institute food frequency questionnaires: The eating at America's Table Study. *American Journal of Epidemiology*, 154(12), 1089-1099.
- Subar, A. F., Thompson, F. E., Potischman, N., Forsyth, B. H., Buday, R., Richards, D., et al. (2007). Formative research of a quick list for an automated self-administered 24-hour dietary recall. *Journal of the American Dietetic Association*, 107(6), 1002-1007.
- Subar, A. F., Thompson, F. E., Smith, A. G., Jobe, J. B., Ziegler, R. G., Potischman, N., et al. (1995). Improving food frequency questionnaires: A qualitative approach using cognitive interviewing. *Journal of the American Dietetic Association*, 95(7), 781-788, quiz 789-790.
- Tanzer, N. K. (2005). Developing tests for use in multiple languages and cultures: A plea for simultaneous development. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (p. 235-263). Mahwah, New Jersey: Erlbaum.
- Thompson, F. E., Subar, A. R., Brown, C. C., Smith, A. F., Shambaugh, C. O., Jobe, J. B., et al. (2002a). Cognitive research enhances accuracy of food frequency questionnaire reports: Results of an experimental validation study. *Journal of the American Dietetic Association*, 102(2), 212-225.
- Thompson, F. E., Subar, A. R., Smith, A. F., Midthune, D., Radimer, K. L., Kahle, L. L., et al. (2002b). Fruit and vegetable

- assessment: Performance of 2 new short instruments and a food frequency questionnaire. *Journal of the American Dietetic Association*, 102(12), 1764-1772.
- U.S. Census Bureau. (1998). *Pretesting policy and options: Demographic surveys at the Census Bureau*. Washington, DC: U.S. Census Bureau.
- U.S. Census Bureau. (2004). *Census Bureau guideline: Language translation of data collection instruments and supporting materials*. Washington, DC: U.S. Census Bureau.
- Warnecke, R. B., Johnson, T. P., Chavez, N., Sudman, S., O'Rourke, D. P., & Lacey, L. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, 7(5), 334-342.
- Weech-Maldonado, R., Weidmer, B., Morales, L., & Hays, R. (2001). *Cross-cultural adaptation of survey instruments: The CAHPS experience*. Seventh Conference on Health Survey Research Methods. Hyattsville, Maryland.
- 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.
- Willis, G., Lawrence, D., Thompson, F. E., Kudela, M., Levin, K., & Miller, K. (2005b). *The use of cognitive interviewing to evaluate translated survey questions: Lessons learned*. Proceedings of the Federal Committee on Statistical Methodology Research Conference. Arlington, VA.
- Willis, G., & Zahnd, E. (2007). Questionnaire design from a cross-cultural perspective: An empirical investigation of Koreans and Non-Koreans. *Journal of Health Care for the Poor and Underserved*, 18, 197-217.
- Willis, G. B. (2004). Cognitive interviewing revisited: A useful technique in theory? In S. Presser et al. (Eds.), *Methods for testing and evaluating survey questionnaires*. New York: Wiley.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for questionnaire design*. Thousand Oaks, CA: Sage.