

The Use of Cognitive Interviewing to Evaluate Translated Survey Questions: Lessons Learned¹

Gordon Willis, Deirdre Lawrence, and Fran Thompson

National Cancer Institute, National Institutes of Health

Martha Kudela and Kerry Levin

Westat

Kristen Miller

National Center for Health Statistics, Centers for Disease Control and Prevention

Contact author: Gordon Willis (Willis@mail.nih.gov)

Abstract

Methods for the development and pretesting of survey questions in languages other than English are undergoing significant evolution. In particular, researchers are increasingly interested in adapting pretesting methods such as cognitive interviewing and behavior coding. It is not a simple matter, however, to apply these methods in the cross-cultural domain. In particular, extension to multiple languages poses particular challenges in terms of staffing, interpretation of results, and data analysis. We review three projects that applied cognitive interviewing to assess cross-cultural equivalence of health questions across English, Spanish, and Asian languages. Based on this variety of experiences we present selected results indicating how cognitive interviewing can be used to make conclusions about the functioning of survey instruments across languages, and make recommendations concerning future practice in this area.

Introduction

Increasingly, survey researchers are concerned not only with the day-to-day process of development, translation, and pretesting of survey questions across culture and language – but also with the development of a methodology that provides a set of “best practices” for doing so (Behling & Law, 2000; de la Puente & Pan, 2003; Gerber, 1999; Harkness, Van de Vijver, & Mohler, 2003; Ji, Schwarz, & Nisbett, 2000; Johnson, 1998; McKay, Breslow, Sangster, Gabbard, Reynolds, Nakamoto, & Tarnai, 1996; Rose & de la Puente, 2003; Stewart & Napoles-Singer, 2000; Warnecke, Johnson, Chavez, Sudman, O’Rourke, Lacey, & Horm, 1997). Rather than reinventing wheels in this regard, researchers are increasingly engaged in adapting existing pretesting methods such as cognitive interviewing (Willis, 2005), behavior coding (Cannell, Fowler, & Marquis, 1968; Fowler & Cannell, 1996), and respondent debriefing (DeMaio & Rothgeb, 1996). Applying these methods is not a straightforward matter, however. Extensions to languages other than English, and especially to multiple languages, pose particular challenges in terms of staffing, analysis, and interpretation of results. The current paper will summarize three projects that involve the application of cognitive interviewing and make recommendations related to future practices.

Study 1. The Current Population Survey Tobacco Use Supplement: Evaluation of Asian Language Translations

The project is described in detail by Kudela, Levin, Tseng, Hum, Lee, Wong, McNutt, & Lawrence (2004), and centered on the evaluation of a National Cancer Institute (NCI) tobacco use questionnaire when translated from English to (a) Chinese (Mandarin and Cantonese), (b) Korean, and (c) Vietnamese. We summarize the difficulties that arose and solutions that were enacted.

Staff Selection

Because the Westat and NCI staff responsible for the cognitive evaluation spoke no Asian languages, they created the position of “Survey Language Consultant” (SLC) for each of the three groups (the Chinese SLC was fluent in both Mandarin

¹ Pre-conference draft of paper to be presented at the 2005 Conference of the Federal Committee on Statistical Methodology. The paper has not been officially approved by NIH and is not to be cited.

and Cantonese) to oversee cognitive interviewing activities, contribute cultural insights, and translate recruiting, data collection, and training materials. Although SLCs varied as far as being local or at a distant location, the investigators found that that location had much less impact on project-related communication than did SLC level of prior survey experience. SLCs in turn each hired two cognitive interviewers by spreading word of the openings through their extensive networks in the Asian community. SLCs in general and Korean interviewers in particular were found to be especially difficult to recruit.

Instrument Translation

Translation received considerable attention: SLCs first reviewed the prior work of the initial questionnaire translators, in order to reconcile differing opinions. The SLCs found a much greater need for revisions than had been anticipated; each reported that the initial translations were very literal, resulting in questions that were too wordy or even confusing (Chinese and Korean teams suggested revisions to about 60 percent of the approximately 200 survey items, and translation revision was even more extensive for the Vietnamese version).

Materials Preparation and Interviewer Training

The investigators spent a considerable amount of time educating SLCs about survey development work, especially the purpose and uses of cognitive interviewing, in-depth discussion of how cultural issues may impact translation and data collection, recruiting techniques, supervising interviewers, and writing up results. The SLCs' main task related to the training was to translate the interview protocols into their individual languages. During this process, computer compatibility problems created a variety of unforeseen problems, especially related to cutting and pasting across language versions between multiple computers.

Cognitive interviewer training was conducted primarily in English, although trainee interviewers and SLCs practiced role plays in their individual languages. This approach was effective, but it was somewhat difficult for those with no survey research background to grasp cognitive interviewing purposes, concepts, and techniques. Most noticeably, SLCs and interviewers struggled to distinguish between survey items and cognitive interview probes. In retrospect, Westat staff recommended increasing training sessions from six to eight or ten hours, to equip interviewers with an in-depth understanding of why each survey item is being asked, and the purpose of each cognitive probe. Overall, investigators felt that the best alternative would be to have the SLCs themselves conduct the interviews, given that SLCs had much more exposure to survey methods, and to the evaluated survey itself, over the course of several weeks as they prepared for testing. This observation is consistent with the feeling within U.S. cognitive laboratories that interviews are best conducted by a cadre of experienced, long-term employees, as opposed to lower-level interviewers who are trained as needed to conduct cognitive testing.

Subject Recruitment

SLCs initially expressed a variety of concerns related to recruiting and interviewing respondents, and suggested that monetary incentive be supplemented with small gifts such as fruit or cookies, particularly if the interview were held in the respondent's home. To locate eligible subjects, the SLCs used a variety of methods, including flyers, newspaper advertisements, community events, word-of-mouth, and personal contacts. These efforts were generally effective, though more demanding than those required for usual (general, English-language) cognitive interviews. In particular, some potential subjects suggested that newspaper ads are more trustworthy than flyers. In the end, personal contacts were by far the most effective tool.

Conducting Cognitive Interviews

Each of the SLCs oversaw completion of nine interviews, and five additional Vietnamese interviews were conducted in a second testing round. Interview protocols included "Anticipated" probes (Willis, 2005) that were fashioned prior to the interview and were administered using concurrent verbal probing (that is, during the interview, as opposed to retrospectively as a debriefing activity). Given the inherent language barriers, Westat team members were not able to monitor interviews. As an alternative quality control measure, SLCs participated in a telephone or in-person debriefing meeting after every second or third interview. During a meeting held after the first few interviews, SLCs reported that subjects were impatient with the redundancy of the interviewers' probe questions. The investigators determined that interviewers were administering the probes word-for-word, regardless of whether subjects had already provided the information the probes were designed to elicit. Further, interviewers were hesitant to deviate from probes, either by using their own words or by following up unanticipated problems with unscripted ("Emergent") probes.

From the start of the project, the researchers had concerns about the consistency of cognitive interviewer behavior across languages, and to some extent these concerns were found to be justified. Based on debriefing meetings with the Vietnamese

SLC, the Westat team experienced doubts about the quality of Vietnamese interviews. In particular, the SLC suggested that Vietnamese speakers cannot understand true/false items because they are not posed as questions, and to remedy the problem, she recommended converting each true/false item into a question. However, a Vietnamese-speaking Westat reviewer who was independently consulted felt that the true/false statements were clear and understandable. In further reviewing the tapes, that reviewer discovered a variety of errors: The probes and interviewer instructions had been mis-translated, and the survey itself was administered improperly. For example, the series of true/false items were read with no pauses, so the respondent was not given an opportunity to answer each individual statement before the interviewer presented the following question. Further, interviewer instructions and skip patterns were sometimes read aloud to the respondents; survey items and probes were read using a monotone delivery; and probes were administered as if they were survey items rather than as-needed. Because of these problems, the Vietnamese interviews were considered unusable, and five more were conducted. This finding speaks to the difficulty that English-only researchers may have in retaining appropriate control and monitoring of non-English cognitive interviews.

For future cognitive testing in foreign languages, Westat staff recommended dividing the first round of interviewing, such that the first sub-round consists of three interviews, and after completion of these, a bi-lingual expert reviewer listens to the taped interviews and summarize any problems or issues interviewers encountered in administering either the survey instrument or the probes. If needed, a remedial training would take place before proceeding.

Problems Identified Through Cognitive Interviews

To analyze interview results, SLCs wrote a detailed summary of each interview, using a note-taking template to guide them. For some items, detailed questions were presented to ensure that the issues of interest were addressed consistently across interviews and among the SLCs. Overall, the following categories of problems were found (with illustrative examples):

A. Translation Errors:

(1) When hearing the question that asked whether they had smoked 100 cigarettes, most Chinese respondents answered with some version of, “In one day?” It was later determined that the phrase “in your entire life” had been left out of the translation.

(2) A Korean translation reversed the meaning of the response choices for the item: “Does anybody smoke cigarettes, cigars, or pipes inside your home?” Translated back to English, the Korean version read, in essence “Is there *not* anyone smoking cigarettes, cigars, or pipes anywhere inside of your home?” The correct answer, if no one smokes, is “yes” (i.e., it is true that there is not anyone smoking cigarettes”).

(3) In Vietnamese, most respondents interpreted the word “community” (“In your opinion, how easy is it for minors to buy cigarettes and other tobacco products in your community?”) as the Vietnamese people in general, presumably because the translation of “community” carries a political connotation. The Vietnamese word for “neighborhood” was also problematic as it implies that the respondent has a relationship with his/her neighbors. The authors recommended translating the question to emphasize physical location: “In your opinion, how easy is it for a minor to buy cigarettes or cigarette products in the area where you are now living?”

B. Culturally Specific Issues:

Problems attributed to differences in cultural viewpoint, or to differences between structural characteristics across societies, were surprisingly infrequent. As an exception, an item which asked whether respondents have ever switched from a stronger to a lighter cigarette posed difficulties for respondents who started smoking a Korean brand of cigarettes and then switched to an American brand, as cigarette manufacturers in Korea are not required to include information about tar and nicotine on each package

C. General Problems:

(1) When asked whether they would go to the store in a bad rainstorm for cigarettes, some subjects who said “no” indicated, upon probing, that they would never have to do this, because they always buy enough cigarettes to ensure they do not run out, or would simply borrow from friends until the weather cleared up. This did not appear to necessarily be associated with Asian culture, however (and may relate more to the culture of smoking).

(2) Similar to previous (English-language) findings, some Chinese respondents did not think of themselves when answering whether anyone smokes inside their homes; the sponsor was encouraged to add “including yourself” or “including those who live here” to the question.

- (3) As has generally been found to be the case, older respondents tended to have more difficulty answering questions, and required many to be repeated several times before they felt able to answer.
- (4) Koreans, as members of other previously tested groups (Spanish, English), had little trouble understanding the question “What is the total number of years you have smoked every day?” until confused by a follow-up instruction to exclude any time they stayed off cigarettes for six months or longer.
- (5) In keeping with another common finding, the Korean translation asked whether respondents have been to a health professional, but does not distinguish whether it was for their own health or someone else’s. Several respondents answered “yes” because they had accompanied others for a doctor visit (e.g., taking children to the pediatrician or one’s wife to the OB/GYN). Project staff recommended adding to the question the phrase “for your own health...” (a similar recommendation has previously been made for an English version of that question).

Study 2: Cognitive Testing of Fruit and Vegetable Questions in Spanish

Study 2 is described in detail elsewhere (Hunt, Willis, Levin, Thompson, Norberg, & Franklin, 2005). Dietary behavior is potentially interesting in a cross-cultural project because diet is culturally specific and its cross-cultural measurement may present unique problems. The project involved the testing of a Food and Drink Consumption section from the (interviewer-administered) National Household Interview Survey (NHIS), sponsored by NCI, and conducted by the National Center for Health Statistics and the U.S. Census Bureau. Westat staff and their consultants conducted independent rounds of interviews in English and in Spanish (9 of each). As for the previous study, the testing procedure involved concurrent verbal probing. Probe questions were Anticipated, and therefore written directly into each language version of the cognitive testing protocol. English-language interviews were conducted by two experienced Westat staff members, and Spanish language interviews by two consultants specifically trained to conduct cognitive interviewing by lead Westat Staff, and by one of the experienced (bilingual) Westat staff interviewers.

Results Concerning the Cognitive Interviewing Process

Novice bilingual interviewers who conducted Spanish interviews relied heavily on the fully scripted probe questions; they rarely used interviewer latitude to follow up with “Reactive” probes that follow up unanticipated findings. Further, for Spanish cognitive interviewers, it was helpful to have an experienced researcher present (serving a mentor role). However, the project staff observed that:

- (1) Spanish-language interviewers tended to follow the probes robotically. Therefore, as in the first study, such non-selective probing caused problems for subjects;
- (2) General comprehension probes (e.g. “*What does this question mean?*”) were problematic, as subjects felt they were being tested;
- (3) Specifically targeted probes (e.g. “*Did you include X when you were thinking about the fruits that you ate?*”) were more effective.

Results Concerning Demographic Characteristics

Between Spanish and English groups, tested subjects differed in ways other than language or ethnicity: Spanish-speaking subjects were much more likely to be less than 35 years old (89% of the former but only 11% of the latter); Spanish-speakers had lower income (89% had <25K/year, as opposed to 11% of English speakers); and approximately half of Spanish-speakers had at most a High School education, whereas 11% of English did. Language, ethnicity, and other demographic features were therefore confounded and observed differences cannot be uniquely attributed to any one factor.

A. Translation Errors:

- (1) For a question on cookies, cake, pie, brownie, “...*comio GALLETAS, PASTELES, TORTAS o PASTELITOS DE CHOCOLATE BROWNIES*”- In Mexico the term “torta” means “sandwich,” so at least one subject misunderstood the question being asked. One subject wondered whether the question included salty cookies as well as sweet ones, because the term “galletas” can be used to refer to either in Spanish;
- (2) The term “salsa,” was problematic because it was comprehended differently by English and Spanish subjects. In English “salsa” refers to a Mexican-style sauce; whereas in Spanish it simply comes across as “sauce,” a more general interpretation

than intended. As such, in Spanish the more specific phrase “salsa picante, pico de gallo o salsa tipo mexicana” was selected. Ironically, even a Spanish term that is understood by English-only speakers may produce translation problems.

B. Culturally Specific Issues:

As for Study 1, these were overall found to be rare. However -

(1) One exception was a case in which a problem was found in English, but not in Spanish: “*Not counting what you just told me about (lettuce salads, white potatoes, cooked dried beans), and not counting rice, how many times per ____ did you eat OTHER VEGETABLES?*” - Foods in mixtures were included by Hispanics. However, 5 of 9 in English neglected to include vegetables in soups/stew mixtures;

(2) Recent arrivals in US were not familiar with “bran” or with names of breakfast cereals (e.g., “Cheerios”; “Fiber One”);

(3) Some allowances had been made for differences in questions to capture cultural differences that were known to exist based on the cultural background of translators and previous field work: (a) As an example of food containing cheese, English used “lasagna;” Spanish used “quesadilla”; (b) Within a question about beans, “green beans” was included in English but not in Spanish, as Hispanics seemed not to spontaneously include green beans in this category. These variations seemed to function well. There was some degree of non-correspondence of items as literal or word for word translation appeared to be less useful for study purposes.

C. General Problems:

(1) Problems in reporting behavior frequency were noted for both groups: When asked “*How many times in the past day, week, or month did you eat/drink (food)?*” subjects in both groups tended to give responses but omitted the denominator (per week, month, or year);

(2) A decision to remove an earlier reference to “*fruit juice blends*” (“mezclas de jugos puros de fruta”) was found to be beneficial for both languages;

(3) For a question asking about “fruit drinks”: Despite an instruction to NOT include sugar-free drinks, all Spanish respondents (and 5 of 9 in English) said they would include sugar-free Kool-Aid;

(4) For “*During the past 12 months, how many times... did you eat FRUIT?*” 3 of 9 in Spanish, and 3 of 9 in English, neglected to include fruit they eat in food mixtures (e.g., with cereal or yogurt);

(5) “*During the past month, how many times... did you have TOMATO SAUCES such as spaghetti sauce or pizza with tomato sauce?*” - In response to a probe about whether ketchup should be included, two in Spanish and two in English, said it should (in departure from researcher intent).

(6) “*How many times did you eat red meat*” - Use of a show card worked well in both languages;

(7) “*... How many times did you eat sugar-baked goods such as cookies, cake, pie, doughnuts or pastries (Include low-fat kinds)*” - In both groups, subjects failed to recognize that the answer was intended to include information reported for prior questions, and said “none,” thinking the question was asking something additional.

Study 3: Systematic Analysis of Cognitive Interviewing Results Across English and Spanish

The project is described more fully by Miller, Willis, Eason, Moses, and Canfield (under review), and involved a very different approach to cross-cultural cognitive interviewing than that used in Studies 1 and 2 above. Rather than systematically documenting the *procedures* appropriate for the conduct of cross-cultural (and multi-lingual) cognitive interviewing, the investigators instead emphasized the manner in which cognitive testing results are processed, evaluated, and compared. The investigators first noted that there is little agreement among practitioners regarding the standards or criteria appropriate for cognitive interviews (Snijders 2003), especially with respect to the nature of writing up results in a way that makes clear to investigators if problems exist. Therefore, they advocated supplementing the normally open-ended written

cognitive interviews with quantifiable outcome codes, where reliance on purely qualitative and sometime impressionistic interviewing approach might lead to erroneous conclusions about cross-cultural discrepancies.

Procedure

The investigators conducted sixty-seven cognitive interviews, divided equally between (a) (self-reported) Hispanic and (b) Non-Hispanic (both White and Black) participants, in both urban (Washington D.C) and rural/suburban locations (two areas of Northwest Ohio). The D.C. area interviews were conducted in the Questionnaire Design Research Laboratory at the National Center for Health Statistics. Ohio interviews were conducted either in the participant's home or in a private room of a community facility. Interviews of Hispanics were conducted by two bilingual consultants, one of whom had translation experience and had previously been trained in both questionnaire design and in cognitive interviewing techniques. Non-Hispanic interviews were conducted by several NCHS staff members, ranging in cognitive interviewing experience (from moderately to very experienced).

The interviews were based on an interviewer-administered health-survey questionnaire containing items covering chronic conditions, cancer screening, diet, physical activity and demographics. All but one of the cognitive interviews of Hispanics were conducted in Spanish, and all Non-Hispanics interviews were in English. The instrument and cognitive probe questions were translated from English to Spanish by one of the Spanish-speaking cognitive interviewers, with no further reconciliation or review. Spanish-language interviews were conducted by two interviewers who were trained for the activity; one of whom was an investigator, the other a college professor. Hence, both represented the approach advocated by Westat researchers, in which cognitive interviewers are fairly high-level project investigators who have prior experience in questionnaire design and survey methods.

The cognitive interviews were semi-structured; along with the survey questions, the interview guide (protocol) consisted of several pre-scripted follow-up questions pertaining to participants' interpretations of key terms and overall comprehension of questions. These fixed (Anticipated) probes ensured that this particular information was collected in every interview and could then be compared across all interviews. To allow a more flexible approach, interviewers were also instructed to exhibit latitude and to inquire as to the ways in which participants constructed their answers to the survey questions, which further provides insight into potential sources of response error. These Emergent, non-scripted probes were designed to help interviewers make sense of gaps or contradictions in participants' explanations, and to provide information needed to interpret question problems.

Results

Despite the development and application of multiple coding categories that represented the key presumed stages of the survey response process (comprehension, retrieval, decision/judgment, and response), the investigators ultimately determined that the most useful overall measure was one indicating if *any* of these problems were found in a particular interview, and that the (text-based) qualitative results were useful for further diagnosing the specific nature of the problem. To assess whether observed problems were systematically related to language, ethnicity or other measured subject characteristics, cross-tabulations and logistic regression analyses were conducted for 18 evaluated questions. Regression analysis included age, gender, income, education, and ethnicity as independent variables; and the presence of a cognitive problem, as represented by the overall problem code indicator, as the dependent variable. These analyses determined that for these items, ethnic group membership (or language, which was almost perfectly correlated with ethnicity) was the strongest overall predictor of problem code occurrence (as computed for each individual subject, within an individual interview), with Spanish-speakers generally experiencing more difficulties than English (for 5 items), but with Spanish-speakers seemingly having fewer problems for two other questions. Spanish produced significantly more problems with a question on ever having cancer, and significantly less for one on combined household income. Five other questions for which language predicted problem frequency involved foods and meals (see Table 1). Somewhat surprisingly, the analysis revealed that other measured demographic characteristics (gender, age, educational level, and income) had, at most, weak effects.

Qualitative analysis was used to facilitate interpretation of these results, and to pinpoint the character of the observed problem. Key observations were as follows.

A. Translation Errors:

It was clear from the qualitative and quantitative data that some translated survey questions caused interpretation difficulties for Spanish-speaking subjects. That is, particular words translated literally from English simply did not convey the same meaning. For example, the phrase *frijoles con chile* was intended to mean chili beans, but was interpreted by most Spanish participants as "beans with hot sauce." Additionally, some terms vary by region (e.g., Puerto Rican Spanish uses *nami* for

yam, while Mexican Spanish uses *camote*) or were inappropriately formal forms of Spanish (e.g., the word *fiambre* for lunchmeat). Consequently, these terms were not universally understood in the same manner across Spanish-speaking subjects. This type of translation problem seemed to account for the high percentage of Spanish-speakers experiencing problems with the red meat question.

Similarly, some words in Spanish consisted of more than one meaning and could easily be misinterpreted, depending on the context in which they occurred. The word *comida* can mean either “meal,” “food,” or the name of a particular meal (e.g., the English word for dinner). Consequently, the question “*Did you eat a morning meal?*” was translated as “*¿Ayer comió Usted la comida de la mañana?*” but misunderstood by some Hispanic participants as “*Did you eat your dinner in the morning?*” This interpretive issue accounted for a significant degree of the ethnicity-based problems regarding the meal questions.

B. Culturally-Specific Issues:

A qualitative examination of the interviews revealed why English speakers subjects, in comparison to Spanish, were likely to experience problems with a question on cooking oil. Many Spanish-speakers reported using only butter or lard, and were easily able to provide a report concerning the one they used most often. Non-Hispanic, English participants were much more likely to cook with a variety of oils, and experienced trouble in reporting the most frequently used type.

C. General Problems:

Reiterating the results of Study 1, some findings were unrelated to language or culture. During the course of conducting the interviews, it became obvious that for the question “*Yesterday did you eat any beans such as kidney beans, refried beans, chili beans, bean soup, bean salad or lentils?*,” participants adopted differing interpretations of the word *bean*. Some participants viewed the question as asking about legumes only, while others included green (string) beans. Yet, at that point, interviewers could only speculate whether there was a particular group of participants that relied on a specific interpretive pattern. For example, the researchers hypothesized that older participants and perhaps less educated participants would be less inclined to view the question as asking about legumes only. As it turned out, based on statistical (regression) analysis, the patterns of interpretation were not related to specific demographic group membership, as “green bean error” was found to be essentially random.

Table 1: Miller et al. study: Percentage of participants having problems with the question, by language group.

Tested question:	Spanish	English
1) How many times did you eat red meat, including beef, pork, lamb, or lunchmeat, hot dogs or sausages made from beef, pork or lamb yesterday?	77.1% (27/35)	35.5% (11/31)
2) Did you eat a morning meal yesterday?	66.7% (24/36)	6.5% (2/31)
3) Did you eat a midday meal yesterday?	54.2% (13/24)	13.8% (4/29)
4) Did you eat an evening meal yesterday?	57.7% (15/26)	6.7% (2/30)
5) When you use butter or oils for cooking or preparing your food, which of the following types do you use most often? a) Butter, Margarine, Lard or Shortening, b) Olive oil or Canola oil, c) Corn oil, Vegetable oil, Peanut oil, Soy oil, d) non-stick spray, e) Don't use fat	8.3% (3/36)	38.7% (12/31)

Conclusions and Caveats

Study 3 suggested that non-trivial differences may exist between English and Spanish language groups in responding to common health survey questions. As found in the other studies, these were due to either translation errors, cultural influences (though minimally), and general difficulties with survey questions. Further, some outcomes “favored” Spanish speakers, others English speakers. Concerning the basic research question – whether systematic coding of results enhances cross-cultural interpretation – this looked promising, but the investigators recognized several limitations and caveats. From an operational point of view, systematic coding of cognitive interviewing results may not always be feasible. In many pretesting studies (e.g., Studies 1 and 2), much smaller rounds of interviews are generally conducted than were included in the current study; restrictions by the U.S. Office of Management and Budget often limit the size of an interviewing round to no more than nine subjects. In such cases strict quantification of results will not supply information sufficient to identify cross-cultural differences. Further, the authors admit that the amount of time required to ensure that all codes were applied

consistently across all interviews, and to develop an analyzable data set, was considerable. Therefore, just as the project staff for Study 1, the investigators on Study 3 cite the need for significant resources in cross-cultural cognitive interviewing studies, but for different reasons (i.e., those related to analysis as opposed to the conduct of interviews).

Finally, and most importantly, it remains the case that even though cognitive interviewing results were systematically coded and therefore produce quantitative, statistically-analyzable data, these data are only as good as the information they derive from. To the extent that the Spanish interviewers were behaving differently from the English-speaking ones (unknown in this case), it is unclear whether significant differences in coding frequency between cultural groups reveal variation that is due to ethnicity/language, or whether this simply represents interviewer variation. On the other hand, the fact that the qualitative results appear coherent, in the context of the qualitative results, does tend to obviate (though not to eliminate) this concern.

Overall Conclusions and Recommendations Concerning Cross-Cultural Cognitive Interviews

Based on experience from Studies 1-3 above, we suggest the following:

- 1) In selecting cognitive interviewers to test translated instruments, bilingualism is a necessary but not sufficient condition. Interviewers are optimally well-versed in questionnaire design, and highly trained in cognitive interviewing. It would be best if the field could develop a cadre of professionals who specialize in a particular language, and who could serve as expert translators, cognitive interviewers, and high-level design consultants.
- 2) Translation difficulty should not be underestimated as a challenge to cross-cultural equivalence, even if “best practices” of translation have been used. The recoding of language for use in standardized surveys is a complex and vexing enterprise, often requiring empirical testing to affect properly. As such, cognitive testing may be integral to the translation process. We recommend that translation receive adequate resources, and always be viewed as incomplete prior to the cognitive testing process.
- 3) Many problems with survey questions are general in nature, and represent core conceptual, retrieval, decision, and response processes that supersede language or culture. To some extent, cognitive interviewing reliably reveals the same problems, across groups. Though the broad presence of problems is generally not welcome to the data collector, this fact is somewhat reassuring from the vantage point of the cognitive interviewer, as it lends credence to the notion that our findings have substance and staying power, rather than representing ephemeral idiosyncrasies. As a practical matter, one might conclude simply that the more cognitive interviews we do, across languages, the more likely we are to find common problems that afflict respondents in any group or language. As such, investigators must be prepared to consider “decentering” of the English version, to allow the incorporation of changes that seem advisable based on testing in languages other than English.

References

- Behling, O., & Law, K.S. (2000). Translating questionnaires and other research instruments: Problems and solutions. London: Sage.
- Cannell, C. F., Fowler, F. J., & Marquis, K. (1968). The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. Vital and Health Statistics, Series 2, No. 26. Washington, DC: U.S. Government Printing Office.
- de la Puente, M., & Pan, Y. (2003). An overview of proposed Census Bureau Guidelines for the translation of data collection instruments and supporting materials. Paper presented at the meeting of the Federal Committee on Statistical Methodology, Arlington, VA.
- 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 (pp. 177–195). San Francisco: Jossey-Bass.

- Fowler, F. J., & Cannell, C. F. (1996). Using behavioral coding to identify problems with survey questions. In N. Schwarz & S. Sudman (Eds.), Answering questions: Methodology for determining cognitive and communicative processes in survey Research. San Francisco: Jossey-Bass.
- Gerber, (1999). The view from anthropology: Ethnography and the cognitive interview. In M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau (Eds.), Cognition and Survey Research. New York: Wiley.
- Harkness, J.A., Van de Vijver, F.J.R., & Mohler, P. (2003). Cross-Cultural Survey Methods. Hoboken, N.J.: Wiley.
- Hunt, M., Willis, G., Levin, K., Thompson, F., Norberg, A., & Franklin, M. (2005). Can Cognitive Interviewing Be Done Cross-Culturally? Poster presented at the annual meeting of the American Association for Public Opinion Research, Miami.
- Ji, L.J., Schwarz, N., & Nisbett, R.E. (2000). Culture, autobiographical memory, and behavioral frequency reports: {PRIVATE } Measurement issues in cross-cultural studies. Personality and Social Psychology Bulletin 26, 586-594.
- Johnson, T.P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. ZUMA – Nachrichten Spezial, 3, 1-39.
- Kudela, M.S., Kevin, K., Tseng, M, Hum, M., Lee, S., Wong, C., McNutt, S., & Lawrence, D. (2003). Tobacco Use Cessation Supplement to the Current Population Survey Chinese, Korean, and Vietnamese Translations: Results of Cognitive Testing. Final Report submitted to the National Cancer Institute, Rockville, MD.
- McKay, R.B., Breslow, M.J., Sangster, R.L., Gabbard, S.M., Reynolds, R.W., Nakamoto, J.M., & Tarnai, J. (1996). Translating survey questionnaires: Lessons learned. New Directions for Evaluation, 70, 93-105.
- Miller, Willis, Eason, Moses, and Canfield (under review). Interpreting the Results of Cross-Cultural Cognitive Interviews: A Mixed-Method Approach. National Center for Health Statistics.
- Pan, Y. (2003). The role of sociolinguistics in the development and conduct of Federal surveys. Paper presented at the meeting of the Federal Committee on Statistical Methodology, Arlington, VA.
- Rose, D., & de la Puente, M. (2003). Guideline implementation: Translating a new Canadian/U.S. health survey into Spanish. Paper presented at the meeting of the Federal Committee on Statistical Methodology, Arlington, VA.
- Snijders, G. (2003). Cognitive laboratory experiences: On pretesting computerized questionnaires and data quality. Unpublished Doctoral Dissertation, Utrecht University, Netherlands.
- Stewart, A.L. & Napoles-Springer, A. (2000). Health-related quality-of-life assessments in diverse population groups in the United States. Medical Care, 38, II-102—II-124.
- Warnecke, R.B., Johnson, T.P., Chavez, N., Sudman, S, O'Rourke, D.P., Lacey, L, & Horm, J. (1997). Improving question wording in surveys of culturally diverse populations. Annals of Epidemiology, 7, 334-342.
- Willis, G.B. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Sage: Thousand Oaks, CA.