The American Community Survey (ACS) en Español: Using Cognitive Interviews to Test the Functional Equivalence of Questionnaire Translations

Lorena Carrasco
U.S. Census Bureau, Decennial Management Division, 4700 Silver Hill Road, Stop 7100, Washington, DC 20233

INTRODUCTION

This report documents the results of a pioneering study to test the functional equivalence of Spanish questionnaire translations. Cognitive interviews were used to determine how respondents interpreted the translation of key questions or terms, and to determine whether respondents' interpretations were consistent with the source language (English) questionnaire. This research represents a milestone in the development of the Census Bureau's blueprint for obtaining high quality data in foreign languages as well as data collected from persons with limited English proficiency. The blueprint consists of four interrelated components: translation guidelines, pretesting standards, quantitative and qualitative research, and sociolinguistic research on multiple language use. The blueprint seeks to apply known scientific methodologies to the development and implementation of bilingual or multilingual questionnaires to ensure data collected in foreign languages are of high quality. The study discussed herein also represents the Census Bureau's first attempt to apply cognitive interview techniques to existing translations of demographic surveys. The findings of this research are in some ways consistent with wording and questionnaire design issues that affect English instruments. In other ways, the findings are important because they confirm and document potential sources of error that have been suspected to affect target language questionnaires, but for which no empirical data were previously available. The findings reveal key linguistic and questionnaire design issues that are relevant to the development and assessment of questionnaire translations.

Translation of Spanish language questionnaires at the Census Bureau typically begins after the English questionnaire has been designed, tested, and finalized. Financial constraints, lack of time, and lack of qualified bilingual personnel are among the most common reasons why questionnaire translations are not routinely pre-tested. Perhaps the greatest challenge to managing bilingual survey instruments is the absence of bilingual survey specialists who can review questionnaire translations from a survey methodology perspective, so the issue of functional equivalence has remained at large. What is functional equivalence and why does it matter? Questions in multiple languages are said to be functionally equivalent if they are measuring the same construct, and in doing so take into account fundamental differences between language and cultural groups (Smith 2002). Because the goal of multilingual questionnaires is to collect valid and reliable data in more than one language, it becomes imperative to have questionnaire translations that are functionally equivalent to the original (or source) language.

THEORETICAL FRAMEWORK

Literature dealing with cognitive interviews and questionnaire design is vast; however, research dealing with these issues in bilingual settings appears to be limited at best. Authors writing about this topic are only beginning to explore the complex relationship between data collection and bilingualism. Much work is needed to explore data collection methodologies using bilingual questionnaires, bilingual enumerators, and bilingual respondents. There is also a need for empirical knowledge about how use of interpreters or use of monolingual questionnaires in bilingual settings affects data quality. Four important assumptions from monolingual research have influenced the character of this research. Response error is one of the most common sources of survey error (Groves 1991). Cognitive discrepancies between how the respondent interprets a question and the actual intent of the question may lead to response error (Tourangeau, Rips & Rasinski 2000). Verbal report methods, such as cognitive interviews, have been successfully used as a tool for questionnaire development, including discovering questions that are flawed, improving questions, as well as content and construct validation (Willis, DeMaio & Harris-Kojetin 1999; Crutcher 1994; Payne 1994; Sudman, Bradburn, & Schwarz 1996). The models have thus shown the respondent may be a source of response error because he or she may not interpret the intent of survey questions, and that cognitive interviewing is a useful tool to identify and remedy these types of shortcomings. In a bilingual or multilingual setting, response error may be compounded by linguistic or cultural interference. Specifically, respondents whose native language is not the same as the source language may encounter difficulty accessing and conceptualizing information of the source culture they may have not yet encoded, and may have difficulty responding to questions that ask about concepts or experiences that may be foreign to them. To complicate matters, the target language may not have equivalent terms to account for some of

these experiences or concepts (Mackey 1970). It is also likely that interpretations will vary depending on the respondent's native language, cultural frame of reference, and degree of acculturation or assimilation into the dominant society. A certain degree of linguistic interference (e.g., cultural, semantic, lexical, or phonological interference) should also be expected. Results show cognitive interviews are a feasible tool for pretesting questionnaire translations in order to ensure functional equivalence of questionnaire translations. Results from these cognitive interviews identified a series of conceptual and interpretive problems with key terms and questions, and helped document salient issues that may be attributed either to linguistic or questionnaire design issues. The findings are not reflective of inadequate translations, or poor translation techniques, but rather point to issues that may become apparent only through pre-testing.

METHODOLOGY

Researchers used Spanish questions as they appear in the American Community Survey (ACS) computer-assisted personal interview (CAPI) instrument. The ACS provides demographic and socio-economic data as well as housing profiles. If funded, the ACS would replace the decennial supplementary survey and collect this information on an annual basis from a sample of 3 million households in every county, including samples in American Indian and Alaska areas, Hawaii, and Puerto Rico. The ACS has three methods of data collection. First, households are asked to complete a paper-and-pencil (PAPI) instrument using mail-out/mail-back methodology. Computer-assisted telephone interviewing (CATI) methodology is used for households that do not return their PAPI questionnaire. A final attempt is made to collect data for missing households using computer-assisted personal interviewing (CAPI) methodology. Like most other demographic surveys, the ACS CAPI questionnaire was first developed in English (source language), and then translated by private contractors into Spanish (target language).

Three bilingual researchers from the Census Bureau, the National Center for Health Statistics, and a private contractor completed a total of 35 face-to-face interviews between January and February 2002 in California, Illinois, Maryland, and Texas. Researchers used the CAPI version of the Spanish ACS and conducted interviews exclusively in Spanish using probes about the meaning of questions or specific terms. The interviews were conducted in community centers and lasted between 30 to 50 minutes. Interviews were audio taped, and respondents received a $30 incentive for their participation. All respondents were over the age of 16. Most were monolingual Spanish speakers with no knowledge of English; a few had limited English proficiency. The distribution of Hispanic heritage is the following: 20 were from Mexico, 6 from Guatemala, 3 from Ecuador, 3 from El Salvador, 2 from Colombia, and 1 from Puerto Rico. A total of 19 women and 16 men participated. Respondents were schooled both in the USA (42%) and abroad (58%). Seventeen percent had 5 years or less of formal schooling, 25% had completed grades 6-8, 25% had completed grades 9-12 but did not graduate from high school, 17% were high school graduates, and 17% had completed some college (no degree).

GENERAL CONTRIBUTIONS

Most questions were found to be functionally equivalent to the source document. That is, the Spanish speaker's interpretation of the question or concept reflected the meaning of the English question about three-fourths of the time. For example, if the English question asked "What is your name" and the person provided a name when he heard the Spanish question, the intent of the target question was considered to be functionally equivalent to the source question. If the English question asked "Are there any boarders who stay here" and the person interpreted the translation for "boarders" as students, then the question was not considered to be functionally equivalent to the source language, since the source language does not ask about students. The findings below are discussed in the context of general patterns that may affect the functional equivalence of questionnaire translations. The results have been classified into four linguistic and three design categories. One survey question is used to exemplify in detail the specific process under discussion.

Linguistic Issue #1: Pre-Existing Bias
Cognitive interviews revealed pre-existing meaning for certain terms in the respondents' cultural frame of reference led respondents to misinterpret the intent of the question. The question *"Did you receive any Social Security or Railroad Retirement benefits during the past 12 months?"* elicits a quick response. Most respondents do not show signs of hesitancy, either verbal or non-verbal, that would suggest they have misunderstood the question. The question about Social Security uses the term *Seguro Social*, which is a proper noun used in many Spanish-speaking countries, albeit not in the same context it is used in the United States. Questions that use terms with pre-existing meaning in different cultural contexts, as in the case of Social Security, will not necessarily show high allocation rates. Thus, these types of questions require special attention from bilingual personnel who are familiar with linguistic and cultural nuances of the target population. Probing for meaning

revealed only a few of the respondents actually interpreted "Social Security" as intended. The results showed 83% of the 35 respondents either reported not knowing what the term *Seguro Social* is, or interpreted the term in a manner that was not consistent with the intent of the question. Several techniques may be explored to overcome the biases associated with pre-existing cultural frames of reference. In this particular case, one option might be to keep (borrow) the English proper noun "Social Security," which may help respondents realize the question asks about benefits in the United States. Another option may be to add a short explanation about what Social Security payments include. This might steer respondents away from interpreting the concept in general welfare or healthcare terms. A third alternative may be a combination of both suggestions. A comparative approach should help determine which wording is best understood as intended.

Linguistic Issue #2: No Equivalent Term or Marker
Perhaps the greatest challenge to functionally equivalent questionnaires is lack of equivalent terms or markers for words or concepts that do not exist in the target language or culture. Thus, the target language will not have a term to account for the concept being measured. In the United States, the foster system trains and pays adults to care for children on a temporary basis. Collecting information about foster children in a bilingual context is quite challenging for linguistic as well as cultural and semantic reasons. The Spanish ACS questionnaire uses *hijos de crianza* to refer to foster children. Conveying the concept of foster child using these words, however, is vague. The word *crianza* stems from the verb *criar* (to raise). This is the most likely explanation for why 75% of the 35 respondents interpreted the term as "any" child the respondent is raising. Those who interpreted the term as intended were familiar with the foster system and used the English term "foster" when referring to the program. For example, some respondents used "el niño foster," a phenomenon which linguists refer to as borrowing (Mackey 1970). The presence of borrowing in this case suggests trying to find a name in Spanish for a U.S. institution such as the foster program might create more methodological harm than remedy.

Linguistic Issue #3: Frequently Occurring Lexicon
A term or concept may be translated in any number of ways. Some of those ways may include terms that are used with greater frequency than others. For example, in the discourse of some areas "car" may occur more frequently than "automobile." In terms of functional equivalence, the challenge to questionnaire translations is finding among all acceptable translations, the specific wording which most consistently elicits the information desired. The question *"At any time during the past 12 months, did anyone in this household receive food stamps?"* illustrates a frequent lexicon used for "food stamps" may pose measurement challenges. In the survey context, c*upones para la compra de alimentos* is widely misinterpreted as a more general concept encompassing public assistance or welfare. After probing for specific names of programs, 42% of the 35 respondents offered the following examples: WIC, cash assistance, food stamps, TANF, and AFDC. Respondents also mentioned assistance from charities, coupons for food that come in the mail (e.g., Val Pak), coupons clipped from the Sunday newspaper, gift certificates or vouchers. These interpretations may result from using the generic term *cupones para la compra de alimentos*, which literally means "coupons to buy food," and which could feasibly include the different types of coupons mentioned by respondents.

Linguistic Issue #4: Literal Translations
Using the same words in two or more languages will not always convey the same idea. For example, if a language has a word for "room" and a word for "mate," does putting those two words together result in a conceptual equivalent of "roommate"? This technique is reflected in *compañero de cuarto*, which was intended to convey the concept of "roommate." Results showed only 33% of the 35 respondents interpreted c*ompañero de cuarto* as a functional equivalent of "roommate." Most thought this concept referred to people who share the same bed; people who sleep together; or sexual partners. The question about roommates, as intended by the Census Bureau, is meant to identify persons who are unrelated to the householder and who share living quarters primarily to share expenses. The word "roommate" poses both conceptual and linguistic challenges for the researchers. Carrasco and Músquiz's (2003) research findings suggest this phenomenon is not very prevalent in the Hispanic culture and in many Latin American countries, where the roommate situation may be limited to college students or migrant workers. There also appears to be some degree of social desirability associated with this term; mainly the disapproval of unmarried women living with unrelated adults. Extended family dynamics also affect clear differentiation between household members and relationship to the head of household. Census Bureau interviewers have reported respondents use pidgin words (mixture of English and Spanish) for this concept, such as "el ruma" or "rumi."

Design Issue #1: Mapping Responses
The question *"What is the highest degree or level of school you have completed?"* elicits a quick response from respondents. For the most part, respondents know what their highest level of education or highest degree attained is. The problem is only about one third of initial responses (n=35) correctly mapped onto the pre-coded response categories. Most initial responses were in the form of some ordinal number, such as "fourth." Detailed probing allowed researchers to accurately map most of

the initially misleading responses onto the pre-coded answer categories. In order to find a response that would more accurately reflect the respondent's highest level of education, it was necessary to ask for a description of the education system in the country where this education was completed, and then compare the foreign school years to the U.S. system to determine what the corresponding level might be. Putting the initial answer in a comparative context also verified the level was accurately recorded. For example, a respondent whose initial answer was "fourth," had actually completed what is considered tenth grade in the United States:

Int:	¿Cuál es el título o nivel de escuela mas alto que usted ha terminado?	Resp:	De primero a sexto.
Int:	**What is the highest degree or level of school you have completed?**	**Resp:**	**From first through sixth.**
		Int:	¿Y después qué sigue?
Resp:	Cuarto.	**Int:**	**And what comes after that?**
Resp:	**Fourth.**		
		Resp:	La secundaria
Int:	¿Cuarto de qué? ¿De primaria?	**Resp:**	**"Secundaria."**
Int:	**Fourth of what? Elementary?**		
		Int:	¿Y eso abarca de qué año a qué año?
Resp:	No, de secundaria.	**Int:**	**And that's from what grade to what grade?**
Resp:	**No, "secundaria."[1]**		
		Resp:	De primero a sexto.
Int:	A ver, dígame como se estructura el sistema de educación en Ecuador, empezando con la priamaria. ¿La primaria abarca de qué año a qué año?	**Resp:**	**From first through sixth.**
		Int:	¿Y despues qué sigue?
Int:	**Tell me what the structure of the education system in Ecuador is like, beginning with elementary school. Elementary school goes from what grade to what grade?**	**Int:**	**And what comes after that?**
		Resp:	Un año de pre-universitario para los que quieren ir a la universidad.
		Resp:	**One pre-university year for those who want to go to the university.**

By asking these probes, it was possible to determine that in fact, fourth grade of *secundaria* in Ecuador corresponded to the sophomore year in high school, which is the tenth year of school in the United States. This was also apparent because after two more years, the respondent could have chosen to go onto a university-preparation course for one year. Although the initial response was [fourth], probing showed it was not the fourth grade that corresponds to elementary school in the United States, since fourth graders do not have the option of going to college after the sixth year of schooling. The few responses that were not mapped onto the pre-coded response categories were mostly for vocational education completed in a country outside the United States. For example, one of the respondents had studied sewing for three years after elementary school. She had a certificate in *corte y confección* [sewing], and worked as a seamstress. It is likely that in terms of formal schooling, she finished the sixth grade and perhaps that would be the most accurate response in her case.

False cognates and homonyms. False cognates and homonyms presented additional challenges to data collection. For example, c*olegio* in Spanish is a synonym for "school," but because it looks and sounds like the English word "college," it is a false cognate (i.e., it may be easily but erroneously confused with college). Thus, the interviewer may be inclined to select the response category for "college," which is not necessarily the appropriate level of education. Relying on the respondent's initial ordinal response (e.g., "third") may not be necessarily appropriate either. The same word may be used in different countries, but not necessarily to refer to the same grades or levels (i.e., they are homonyms). For example, a Mexican respondent might use "third" to refer to 3rd year of elementary school, 3rd year of middle school, or 3rd year of high school. Results from these cognitive interviews suggest levels of education will not necessarily be consistent across or within countries. For this reason, false cognates and homonyms might be particularly challenging to education or literacy surveys.

---

1 *Secundaria* is typically used to refer to a level that comes after elementary school.

Design Issue #2: Question-Order Effects
A challenging issue may rise in target-language questionnaires as a result of the order in which questions are asked. Question order effects may or may not reveal the same patterns for both the target and source language. For example, both the English and Spanish questionnaires ask whether the respondent received money from different types of income sources. The question *"Last week, did you do any work for either pay or profit?"* uses the word *lucro* in Spanish to convey the concept of "profit." Although this is an acceptable and perhaps the most accurate way of translating "profit," Spanish speakers tend to ascribe a negative connotation to the term. Rather than interpreting the term as any other work that resulted in earnings, 42% of the 35 respondents thought the question was asking about money that was earned by cheating, taking advantage of someone, gambling, or committing some type of fraud or illegal act.

Design Issue #3: Automation
Automated instruments display text that is hard coded (i.e., text that never changes) and text that is soft coded (i.e., words that change depending on the context of a question). Words that change are called "fills," and grammar rules define the logic or rules for wording changes (e.g., adding or omitting helping verbs, tense agreements, possessive pronoun agreements, etc.). Applying English logic to Spanish fills results in text that is difficult to read and understand. For example, the Spanish language requires defining fills based on different logic from that of English fills. To illustrate this point, the common English fill <your/his/her> is based on 1) whether the respondent is the subject or proxy, and 2) the subject's sex. In Spanish, the same fill would need to be based on the noun's number. This would require identifying a different set of variables to determine what the screen should display. Otherwise, the English logic would display Spanish fills that are nonsensical, or fills in English. The latter situation causes English words to appear embedded within Spanish text. In cases where multiple fills exist, inadequately defined fills result in long strings of English words embedded within Spanish text, making it extremely difficult to read and understand questions. This contributes to interviewers having to constantly toggle between the Spanish and English instruments, which in turn discourages interviewers from using Spanish instruments and encourages them to translate on the fly. Another linguistic aspect that becomes relevant, in terms of English-Spanish bilingual instruments, is one of these two languages may require fills where the other one does not. For example, the fill for helping verbs (do, does, did) is prevalent in the English instrument. These helping verbs, however, are not needed in the Spanish instrument. The Spanish instrument, on the other hand, will require fills to display gender agreement, whereas the English language does not need this type of fill. Although these problems have been reported, identified and documented, there appears to be no clear sense of who might be ultimately responsible for language issues in an automated questionnaire. Should the translator become familiar with specifying logic? Should the translator work with a programmer or should the programmer be bilingual? Answers to these questions will need to be addressed as use of multilingual questionnaires and automated instruments increase.

DISCUSSION

Results from this study reveal the concept of functional equivalence is important to data collection in a bilingual or multilingual setting. As measurement instruments, questionnaire translations require pre-testing to ensure they yield valid and reliable data. Qualitative pre-testing techniques such as cognitive interviews can show which questions present respondents with trouble interpreting the intent of the question, correctly understanding key survey concepts, misinterpreting key terms, and challenges matching the respondent's answer onto pre-coded response categories. Results from qualitative pre-testing may be effectively used to find wording in the target language that is functionally equivalent to the source language. The systematic approach used throughout this research shows an effective way to improve the quality of data obtained in foreign languages as well as data obtained from respondents with limited English proficiency is to apply known pre-testing methodologies, such as cognitive interviews, to questionnaire translations *before* they are fielded.

In terms of questionnaire design, the results discussed herein suggest the need to carefully assess questionnaire translations from a questionnaire design perspective. This is necessary to ensure both that translations are consistent with basic principles of questionnaire design, and consistent with research objectives. After all, the most important test a questionnaire translation needs to pass is "Does this question measure the same thing in the target language?" And it may be difficult to provide an accurate response to this question without systematically and empirically assessing the functional equivalence of questionnaire translations. In terms of linguistic challenges, it is likely the types of issues discussed herein might not be evident to the translator, and may become evident only after testing. This may be due to the fact a question itself may be an appropriate translation, yet may not be acceptable as a construct measure. For this reason, questionnaire translations must be assessed both as translations and measurement instruments. It may not always be possible to have a translator who is trained in survey methods, or a survey methodologist who is bilingual, but the two types of assessments are needed. Regardless of the techniques that are used to translate questionnaires, pre-testing is essential to identifying and remedying linguistic and

design challenges that could result in response error. A final note is offered on the advantage of using cognitive interviews to addresses the intra-language variation of Spanish. After all, the notion of "different Spanishes" frequently clouds understanding of Spanish language issues. Keeping this in mind, a main objective of this type of testing is to find the wording that is less frequently misinterpreted, since misinterpretation would be expected to lead to greater response error. Though different terms may be preferred by different Hispanic subgroups, terms that consistently convey the intended construct should be employed. Selecting terms for use in measurement instruments must therefore be based on systematic and empirical testing, and such objective selection should take precedence over subjective preferences of intra-language variation.

This paper has illustrated some of the linguistic and design challenges that cognitive interviews may help identify and remedy. Without systematic testing, there is no empirical basis for evaluating the efficacy of a questionnaire translation as a measurement instrument. Thus, survey specialists should routinely test questionnaire translations regardless of the process, techniques, and quality assurance steps that translators may have used to produce the final target language questionnaire. No other recommendations are offered. Instead, the author invites the reader, and in particular the bilingual reader, to explore and assess the efficacy of current data collection methodologies in bilingual or multilingual contexts.

References
Carrasco, Lorena and Margarita Moreno Músquiz. 2003. "Using Focus Groups to Improve Spanish Census Forms." Paper prepared for discussion at the meetings of the American Association for Applied Linguistics: March 22-25, 2003.

Crutcher, R. J. 1994. "Telling what we know: The use of verbal report methodologies in psychological research." *Psychological Science* 5: 241-244.

Groves, R. 1991. "Measurement errors across the disciplines." In *Measurement Errors in Surveys,* ed. Paul Biemer, Robert Groves, L. Lyberg, N. Mathiowetz, and S. Sudman, pp. 1-28. New York: John Wiley & Sons, Inc.

Mackey, William F. 1970. "The Description of Bilingualism." In Readings in the Sociology of Language, ed. Joshua Fishman.   The Netherlands: Mouton & Co. N. V. Publishers.

Norman, D., and Rumelhart, D. 1975. "Memory and knowledge." In *Explorations in Cognition*, ed. D. Norman and D. Rumehlart, pp. 3-32. San Francisco: W. H. Freeman and Company.

Payne, J. W. 1994. "Thinking Aloud: Insights into information processing." *Psychological Science*, 5: 245-248.

Smith, Tom. 2002. "Developing Cross-National Survey Instruments." Paper prepared for presentation at the meetings of the International Conference on Questionnaire Development, Evaluation, and Testing Methods: November 14-17, 2002.

Sudman, S., Bradburn, N., and Schwarz N. 1996. "Thinking about Answers: The Application of Cognitive Processes to Survey Methodology." San Francisco: Jossey-Bass.

Tourangeau, R., Rips, L., and Rasinski, K. 2000. "The Psychology of Survey Response." New York: Cambridge University Press.

Willis, D., DeMaio T., & Harris-Kojetin, B. 1999. "Is the bandwagon headed for the methodological promised land? Evaluating the validity of cognitive interviewing techniques." In *Cognition and Survey Research,* ed. M. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau, pp. 133-153. New York: Wiley.