

RESEARCH REPORT SERIES
(*Survey Methodology #2014-01*)

**Multilingual Survey Design and Fielding:
Research Perspectives from the U.S Census Bureau**

Yuling Pan
Jennifer Leeman
Marissa Fond
Patricia Goerman

Center for Survey Measurement
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: July 30, 2014

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Multilingual Survey Design and Fielding: Research Perspectives from the U.S. Census Bureau

Yuling Pan, Jennifer Leeman, Marissa Fond, and Patricia Goerman
Center for Survey Measurement, U.S. Census Bureau

1 Introduction

Demographic data show increased international migration in recent decades, with this trend projected to continue at least through mid-century (IOM 2010). Many migrants have only limited proficiency in the national languages of the receiving countries, meaning that linguistic diversity and “super diversity” are becoming the new norm (Blommaert and Rampton 2012). In many places there has also been greater formal recognition of the language rights of minority language speakers, including both long-standing linguistic minorities as well as more recent immigrant groups. As a result, survey organizations around the world have shown increased interest in creating measurement materials and procedures in multiple languages to ensure that minority language respondents are accurately represented in statistical data. In the United States, increases in the percentage of the population that speaks a language other than English at home have engendered new federal, state, and local policies governing language rights. At the federal level, Executive Order 13166, signed in by President Clinton 2000, requires agencies to develop and implement systems to provide meaningful language access to federal programs for Limited English Proficient (LEP) individuals. However, the Census Bureau was already engaged in multilingual data collection prior to Executive Order 13166; translation of Census Bureau questionnaires into Spanish was mandated by Congress in 1976, as part of a broader directive to improve social, health, and economic statistics regarding the Hispanic population (Public Law 94-311). Further, in addition to the commitment to language access and equal representation, there is also now greater recognition that failure to collect data in minority languages may lead to nonresponse bias, since some individuals who do not speak English report hesitancy or unwillingness to respond (Bates and Pan 2009). For these reasons, in the past few decades, the U.S. Census Bureau has devoted increased resources to its data collection program in languages other than English. The 2010 decennial census questionnaire was available in five languages, and language assistance guides designed to serve as response aids for the English-language questionnaire were produced in 59 languages (Kim and Zapata 2012).

Surveying a linguistically diverse population is a complex process that involves numerous linguistic and logistical considerations and affects every step of survey production and operations. In addition to creating questionnaires -- including questions and response options, instructions, and help text -- in multiple languages, survey organizations also need to prepare in-language supplementary materials such as advance letters that inform respondents about the upcoming survey and encourage participation, brochures that provide detailed background information about the survey and its purpose, and reminder postcards designed to increase survey cooperation. While translation is obviously a key component in the development of instruments

and supplemental materials, there is also a need to develop appropriate processes and procedures for fielding surveys with multilingual populations.¹

An often overlooked aspect of carrying out surveys in multiple languages is the need for research and testing. It is now widely recognized that survey development is an iterative process involving multiple rounds of pretesting and revision. However, in the case of surveying multilingual populations, some survey organizations initially conceive of this undertaking as simply a matter of hiring translators to render a survey in additional languages after pretesting in the original language is complete. The assumption is that, if the source version has been thoroughly tested, other language versions will perform similarly and data quality will be comparable. However, as we discuss in this report, there is a need for testing and revising translated surveys and multilingual field procedures, and these practices should be integrated into the survey development process.

In this report, we identify challenges and crucial considerations for multilingual survey development, illustrating our discussion with research conducted by the U.S. Census Bureau pertaining to three major stages of multilingual survey development and deployment: translation of survey materials, qualitative and quantitative pretesting, and fielding. The Census Bureau confronts linguistic issues on a scale much greater than other organizations because of its status as the flagship federal statistical agency and its mission to count everyone living in the United States. The kinds of studies conducted at the Bureau can provide insights for other agencies as well as commercial and non-profit survey organizations.

In our examination of key issues and findings regarding the impact of cross-linguistic differences in multilingual surveys, we analyze two issues related to cross-linguistic and cross-cultural standardization: (1) the standardization of survey materials and procedures (i.e., standardization of the stimulus) vs. data quality outcomes (i.e., standardization of the effect); and (2) the appropriateness of utilizing existing research methods with populations other than those for whom they were originally developed. We conclude with a review of the main issues and next steps for research and evaluation of multilingual surveys, with the goal of informing best practices.

2 Translation of Questionnaires and Survey Materials

In the U.S., public and private survey organizations undertaking a survey in multiple languages typically start with an English-language “source questionnaire” and then translate it, as well as supplemental materials, into other languages.² However, rather than finalizing the source questionnaire and then translating it, survey methodologists advocate incorporating cross-linguistic and cross-cultural input from the inception of the survey design process (e.g., Dorer 2011; Fitzgerald, Widdop, Gray, and Collins 2011). Translatability assessments (Conway, Patrick, Gauchon, and Acquadro 2010) conducted early in the design process can identify potential issues and constraints so that they can be addressed before the source questions and terminology are finalized. This parallel development of multilingual surveys is only possible for the U.S. Census Bureau on the rare occasions when a new survey is created or new questions are added to an existing survey. For example, when the Current Population Survey’s series of

¹ Other issues that survey organizations should address are the mechanisms for capturing, processing and disseminating data in multiple languages. The linguistic and cultural appropriateness of outreach, advertising and promotional campaigns must also be considered.

² In the U.S. federal system this is the case because surveys are often commissioned and designed before decisions about how many – and which – additional language versions will be created are made. In addition, in some cases, a survey has been in circulation for years before the decision is made to offer translated versions.

questions on respondents' source of health insurance was modified to include the new health exchanges, English and Spanish versions were developed and tested simultaneously, allowing feedback from tests in both languages to inform the process (Pascale, Leeman, Cosenza, and Schoua-Glusberg forthcoming). However, the most common scenario at the Census Bureau is that there is a pre-existing English-language survey for which multilingual versions are a relatively recent add-on. Therefore, how best to address issues in cross-linguistic and cross-cultural translations is an important consideration.

Critical to questionnaire translation is data comparability and conceptual equivalency across languages, a goal that can prove deceptively difficult to achieve. The first challenge is how to ensure that questions and answer categories convey the same meaning in the source and target languages, in particular, with respect to concepts that demonstrate significant cross-cultural variation in their social construction and use (such as racial categories and classifications). Another challenge is how to ensure the "perceived meaning" of a question corresponds to the "questioner's intended meaning" (Braun and Harkness 2005). This is critical in multilingual surveys because the interpretation of meaning is governed by cultural norms of communication (Saville-Troike 1989). Since communicative norms differ across languages, when a question is translated into another language, the respondent may interpret it differently from how the survey designer intended.³ Finally, in cross-linguistic and cross-cultural survey contexts, there may be a heightened risk of "ulterior signaling" where a particular terminology, structure, or context of the question inadvertently raises respondent suspicions that other forms of information are being sought (see also Pan and Lubkemann forthcoming). Therefore it is crucial to securing "question-focused responses" – i.e., responses that reference the (intended) terms or focus of the questions, and not other topics or concerns.

A large and growing body of research regarding survey translation has generated theoretical guidelines for optimal survey translation. The two approaches to translation that have received most attention and discussion in survey research are *adoption* and *adaptation* (Harkness, Van de Vijver, and Johnson 2003). *Adoption* consists of a direct word-for-word translation of the questionnaire from the source language into the target language, and emphasizes the standardization of question wording and formats across languages. This approach is often the default approach adopted by those with little linguistic or translation experience. In contrast, *adaptation* seeks to produce translations that measure the same constructs as the original, even when this requires modifying the original structure in the target language rendition. Because *adapted* translations appear to differ from the source language originals, they can raise concerns regarding standardization of survey questions and response options. This is a particularly salient concern for large-scale surveys with questions that have been asked for many years, such as those conducted by the U.S. Census Bureau.

The *adoption* approach presumes that every word in the source language has a corresponding word in the target language, and thus it characterizes translation as a matter of simply finding the correct terminological equivalent in the target language. However, such one-to-one correspondences do not consistently or reliably exist across languages. There are sometimes subtle differences in the information conveyed by similar words in different languages. In addition, some languages require multi-word phrases to express what other languages can convey with a single word. For example, the English-language term "nursing home," which appears in numerous surveys, can mean a residence for senior citizens in addition

³ While it is true that an English-speaking respondent may interpret a question differently from how the survey designer intended, a translated version contains an additional level of challenges in conveying meaning.

to a convalescence or rehabilitation facility. In some non-English languages (e.g., Chinese, Korean, and Vietnamese), a single translated term of “nursing home” does not convey this additional meaning, and thus an additional phrase is required (Pan, Sha, Schoua-Glusberg, and Park 2009).

However, even words with similar meanings at the semantic or referential level may have different sociocultural meanings or implications. Questions which reference social practices unique to a specific country, or for which there are differences in cross-cultural norms, are particularly problematic for the *adoption* approach. Consider, for example, the challenges of translating the educational attainment questions on the U.S. Census Bureau’s American Community Survey (ACS) into Spanish (Goerman, Fernandez, and Quiroz 2013). The original question makes reference to “home school,” an educational practice that is closely linked to U.S. educational regulations and policies, and which has no direct equivalent in the countries from which many monolingual Spanish-speaking immigrants have arrived. In an attempt to standardize across languages by using the *adoption* approach, a translation based on the terms “home” and “school” was tested: *enseñanza en el hogar*, or “teaching/schooling in the home.” Although respondents were familiar with these terms individually, they interpreted the compound term differently from what was intended. Specifically, they understood it as referring to private tutoring, online classes, or informal lessons taught by parents to their children. These concepts are relevant in educational systems in their countries of origin, and do not relate to the notion of “home school” common to many American English speakers. Improving the translation by finding a “better” or “more accurate” terminological equivalent, as would be suggested by the *adoption* approach, is simply not possible because no such term exists.

The same study also highlights another shortcoming of the assumption of terminological equivalence: its failure to consider linguistic variation. All speech communities exhibit variation in their language use, including differences in pronunciation, vocabulary, sentence structure and pragmatic norms.⁴ One key parameter of linguistic variation is geography; a single language tends to vary from place to place, and from country to country. As a result, one language spoken by immigrants from a range of different countries may present significant terminological variation. Although there may be a single English-language term used in the U.S., the corresponding Spanish term may vary from country to country, leading to confusion when translations strive for strict word-for-word equivalence and standardization of question formats. For example, the terms typically used to refer to “high school,” “college” and “bachelor’s degree,” which are central to the ACS educational attainment question (see Figure 1), vary across Spanish-speaking countries.

⁴ This is an issue that we believe merits more attention and research in the field of survey methodology, in English and other languages, however a discussion is beyond the scope of this report.

11 What is the highest degree or level of school this person has **COMPLETED**? Mark (X) **ONE** box. If currently enrolled, mark the previous grade or highest degree received.

NO SCHOOLING COMPLETED

No schooling completed

NURSERY OR PRESCHOOL THROUGH GRADE 12

Nursery school

Kindergarten

Grade 1 through 11 – *Specify grade 1 – 11*

12th grade – **NO DIPLOMA**

HIGH SCHOOL GRADUATE

Regular high school diploma

GED or alternative credential

COLLEGE OR SOME COLLEGE

Some college credit, but less than 1 year of college credit

1 or more years of college credit, no degree

Associate’s degree (*for example: AA, AS*)

Bachelor’s degree (*for example: BA, BS*)

AFTER BACHELOR’S DEGREE

Master’s degree (*for example: MA, MS, MEng, MEd, MSW, MBA*)

Professional degree beyond a bachelor’s degree (*for example: MD, DDS, DVM, LLB, JD*)

Doctorate degree (*for example: PhD, EdD*)

Figure 1. 2013 ACS educational attainment question

In some places, “colegio” is the most common term for high school, while in others “escuela secundaria” is the most accurate equivalent. Goerman et al., (2013a) found that employing a single direct translation in the questionnaire led to inconsistent, variable responses that were dependent upon a respondent’s country of origin. As a result, respondents from some countries appeared to over-report their educational levels.

In addition to concepts such as educational attainment, which vary by language, culture, and country, more abstract concepts (such as ratings of “satisfaction” and “happiness”) are difficult to translate directly because even though terminological equivalents may be apparent, the interpretations of these concepts are not comparable across multilingual and multicultural respondents (see Smith 2004 for a review).

Another problem with word-for-word translations is that they generally do not take cross-linguistic comparability of pragmatic effect into account, as can be seen in a recent project undertaken by the Census Bureau to develop translations of existing English-language survey advance letters for the ACS (Chan and Pan 2011, Pan, Hinsdale, Park, and Schoua-Glusberg 2006, 2008; Pan and Landreth 2009). See Figure 2 below for an example of a Census Bureau

advance letter incorporating multiple languages.⁵ In the first phase of the study, four versions of the advance letters were translated word-for-word into Spanish, Chinese, Korean, Russian, and Vietnamese and then pretested with monolingual native speakers in each of the target languages.

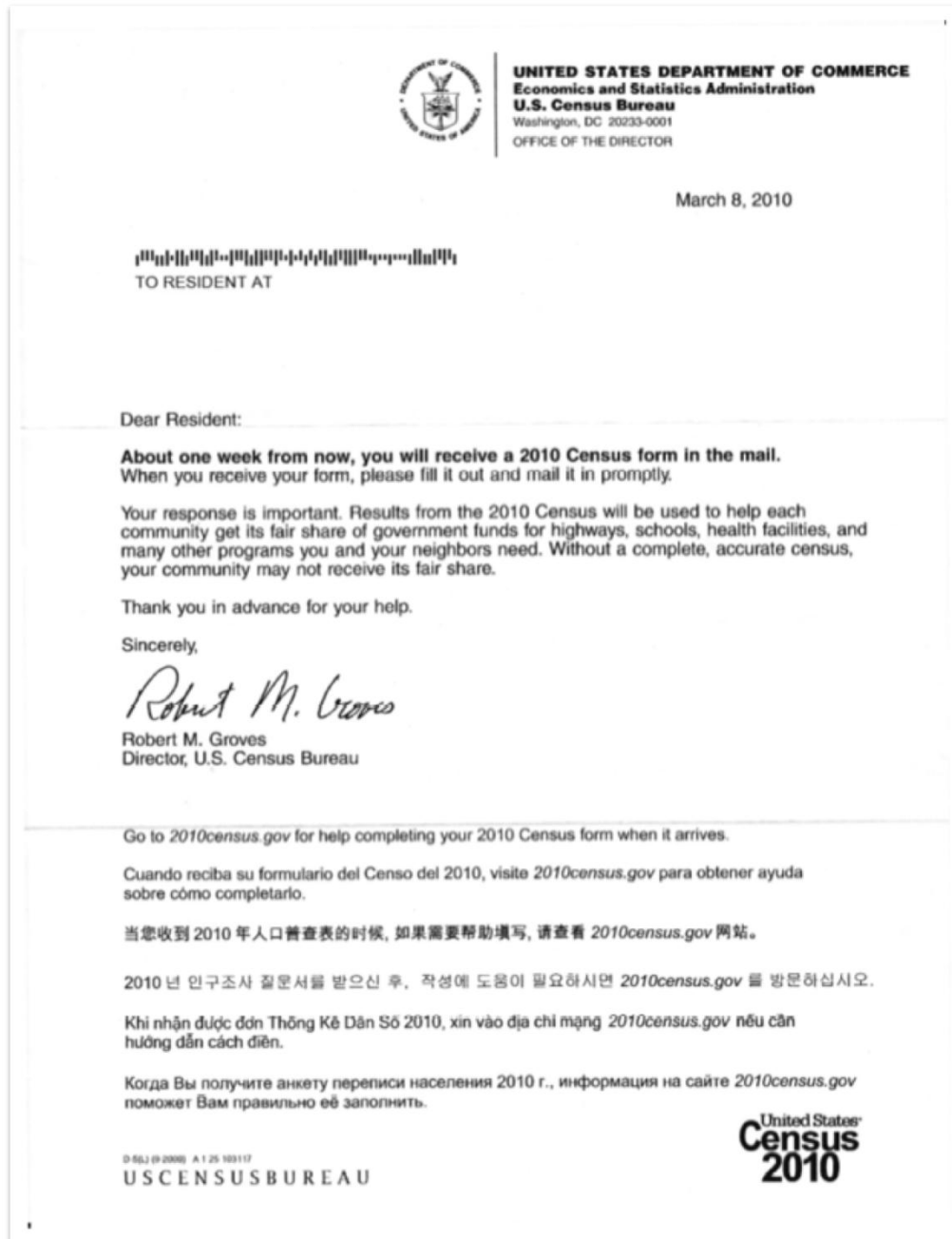


Figure 2. 2010 Census English-language advance letter.

⁵ Survey organizations use advance materials as part of a strategy to reduce household nonresponse rates (Dillman 1991; Fox, Crask, and Kim 1988; Yammarino, Skinner, and Childers 1991) and to establish the legitimacy of a survey (Dillman 1978). Among limited English proficiency (LEP) respondents, knowledge of the decennial census and other U.S. national surveys was minimal, and the belief that such surveys are mandatory and responses are confidential was not commonly held (Bates and Pan 2009).

The problems experienced by respondents differed somewhat across languages, but some common complaints were verbosity, difficulty following the sequencing of information, perception of a stiff or unfriendly tone, and intimidating and incomprehensible legal language. Researchers found that while respondents understood the words in the direct translations of the letters, they did not grasp the intended messages or “point” of the letters. For example, the phrase “response to this survey is required by law” was intended by the Census Bureau to convey the mandatory nature of the forthcoming survey with the goal of making the respondent understand that he or she must answer the survey questions. However, many respondents reading the translations in their native language did not interpret the mandate as applying to them. Some reasons they gave for this conclusion implied assumptions about U.S. law (e.g., “The U.S. is a free country and thus its residents cannot be forced to respond.”), a lack of experience with survey procedures, and misunderstanding concerning the intended survey respondent (e.g., the survey is meant for U.S. citizens only, thus the letter was not relevant to the respondent). Direct word-for-word translation of such letters from English into target languages may not result in letters that encourage the desired outcomes even when they are faithful to the original in terms of grammatical structures and lexical items.

Although *adoption* is often seen as the default approach to survey material translation, many survey researchers advocate the alternative *adaptation* approach to survey translation (e.g., Behling and Law 2000; Harkness 2003; Harkness and Schoua-Glusberg 1998; McKay, Breslow, Sangster, Gabbard, Reynolds, Nakamoto, and Tarnai 1996; Harkness, Villar, and Edwards 2010). Recent research has also illuminated the benefits of this approach (e.g., Levin, Willis, Forsyth, Norberg, Kudela, Stark, and Thompson 2009; Fitzgerald et al. 2011; Willis and Zahnd 2007; Pan and Fond 2010, forthcoming). Proponents of *adaptation* recommend using a source questionnaire as a template, but then allowing survey questions to be significantly modified to make them more easily comprehensible in a different target language. The modifications this approach allows can take several forms, including providing supplemental explanations of terms, adjustments to grammatical rules that are language-specific, and some accommodation of language-specific conventions and sensitivities (such as indicators of politeness). In contrast with structural or *formal equivalence* stressed in the adoption approach, the adaptation approach aims to ensure that translated survey questions measure the same or similar constructs as the source questions, thus achieving *functional equivalence* even if this requires reformulation of question wording, structure, and sequencing of information in the questions, instructions, and response options. Proponents of the *adaptation* approach do not reject standardization but they prioritize cross-linguistic standardization of semantic and pragmatic meaning, rather than standardization of the stimuli. For example, in the case of the ACS educational attainment question discussed earlier, *adaptation* might consist of adding information about what home schooling consists of, or reformulating the question to inquire about the number of years of schooling completed (instead of highest level completed). However, there is some institutional resistance to such an approach, as a result of concerns that reformulating the question in non-English languages might introduce cross-linguistic discrepancies, and revising the English-language question would have an impact of historical data comparisons.

In comparison with institutional resistance to *adaptation* in the translation of survey questionnaires, there is generally a greater willingness to adapt supplementary materials such as advance letters or informational brochures, as these are designed to encourage compliance from respondents, and are not as closely linked to data collection instruments. The main

communicative intent of a typical survey advance letter is to notify respondents of an upcoming event (i.e., the survey) that will require their attention, and to request future action from them (i.e., response). It is often also necessary to include information about legal protections and obligations of respondents, as well as how their information will or could be used by the survey organization. Given that respondents may not interpret direct translations of the messages contained in translated advance letters as intended, as was the case in the study described above, adaptation can be useful for making these communicative intentions clearer and more effective. The Census Bureau's cognitive research on multilingual ACS advance letters (Pan and Fond 2012) found that in order to achieve the letters' intended outcome, the main messages, rather than the lexical items, grammatical structure, or information-sequencing conventions of the original should take precedence. The goal of encouraging participation in surveys across language groups can only be realized by providing convincing, inspiring, and culturally appropriate advance materials tailored as well as possible to the linguistic, cultural, and social backgrounds of the respondents. For this reason, the final version of the Chinese version of ACS survey letters included additional details clarifying the identity of the sender (the director of a U.S. Census Bureau regional office) and addressee. Messages were re-sequenced to lead with background about the survey and the respondent's role in the survey process rather than the message about mandatory participation that had come first in the original English letter. Additionally, the tone of the Korean version of the ACS letters was adjusted to contain more politeness signals and friendliness to mitigate the legal terminology that struck respondents as harsh and threatening to non-English speaking respondents. Similar modifications were made to Russian and Vietnamese versions of the advance letters.

In this section we argued that word-for-word adoption of English source language questions and materials into other languages cannot adequately address the key challenges facing survey translators. We discussed various issues involved with cross-linguistic and cross-cultural standardization and demonstrated that standardizing the stimuli across language groups does not lead to standardization of (comparable) responses and communicative effect. Therefore, in order to ensure translation quality and cultural appropriateness, and ultimately achieve the goal of collecting quality data from all respondents, it is necessary to allow for flexibility in expression. To determine whether a translation is successful, regardless of whether it is conducted via the adoption of the adaptation approach, it must be thoroughly reviewed and pretested. In the next section we address this type of research.

3 Pretesting of Multilingual Questionnaires and Materials

In general, pretesting of survey items aims to ensure that the questions are asking what researchers intend, that they are understood consistently across respondents, that respondents are willing and able to answer the questions, and that the questions are not overly burdensome (Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau 2009). However, research in this area has evolved and there is now also a focus on the definition and empirical evaluation of the methods themselves (Presser, Couper, Lessler, Martin, Martin, Rothgeb, and Singer 2004). Common pretesting techniques include expert review, focus groups, cognitive interviewing, behavior coding, usability testing and field testing. While various pretesting methods have been applied to English-language, U.S. surveys for years, the techniques have also been increasingly applied to survey translations and multilingual surveys (Harkness, Mohler, van de Vijver 2003; Harkness, Vilar, and Edwards 2010). In addition to linguistic concerns, two other issues regarding the pretesting multilingual surveys must be considered. First, it is crucial to take the

social experiences and expectations of LEP respondents into account. Second, survey researchers should be aware that it may not be possible to conduct pretesting in the same way, and to obtain directly comparable data, across language groups. We begin by discussing key pretesting methods and illustrating the insights and challenges associated with them.

One of the most basic quality control procedures applied to survey translations is expert review, in which a panel of experts is assembled and asked to assess a preliminary translation and to identify potential problems or difficulties for respondents (U.S. Census Bureau 2003). Ideally, such panels include survey methodologists with knowledge of the theoretical and/or practical aspects of questionnaire design as well linguists with expertise in the target language and experts with specific knowledge of the target language population as well as the subject matter of the survey. In addition to reviewing materials independently, the entire panel may meet to discuss problems, propose solutions and identify areas for additional testing. Expert reviews can eliminate major design errors that may create problems or add undue respondent burden.

One limitation of expert review is that, because it does not include respondent input, it may not accurately anticipate and capture respondents' backgrounds and life circumstances (U.S. Census Bureau 2003). It can be difficult for experts to understand all of these variables, particularly when the target population includes immigrants from a variety of national origins or cultural groups. Further, the fact that expert reviewers are bilingual in the source and target languages as well as highly knowledgeable about surveys means that they may not anticipate the difficulties experienced by monolingual respondents unfamiliar with survey constructs and procedures.

Thus, in order to incorporate the insights of potential respondents, survey organizations have increasingly turned to cognitive interviewing to pretest multilingual surveys (e.g., Blair and Piccinino 2005; Dean, Caspar, McAvinchey, Reed, and Quiroz 2007; Forsyth, Kudela, Levin, Lawrence, and Willis 2007; Levin et al. 2009; Willis, Kudela, Levin, Norberg, Stark, Forsyth, Brick, Berrigan, Thompson, and Lawrence 2010). Cognitive testing is a method in which survey respondents are interviewed one-on-one in order to examine how they interpret, comprehend and respond to survey questions and other materials (Willis 2005). The goal of the process is to identify aspects of a survey instrument (e.g., question wording, response options, instructions, etc.) that are not perceived by respondents in the way that the survey designers intended. Results from cognitive interviews are then used to recommend alternative wording or other modifications, which are then tested in a second round of cognitive interviews. This can be an iterative process involving multiple rounds of modification and testing until success is achieved.

The cognitive testing method has been used to pretest translated survey questions and materials. Typically, translated materials are tested with monolingual speakers of the target language in order to make sure that they are equivalent to the original in content, tone, meaning, and effect (Pan, Landreth, Hinsdale, Park, and Schoua-Glusberg 2007). For example, in a Census Bureau study of pretesting the 2010 Census form in multiple languages (Pan, Sha, Park, and Schoua-Glusberg 2009), a multilingual research team was assembled and researchers utilized cognitive interviews conducted in each of the target languages to gauge potential respondents' comprehension of and response to the translated census questions. Based on the results of the first round of interviews, researchers proposed alternative translations, and then conducted a second round of interviews to test them, in order to identify the best translation.

Cognitive interviewing can also help detect LEP respondents' noticing of and reaction to multilingual materials. In a cognitive interview study of reactions to a multilingual brochure containing information in five different languages, Chan and Pan (2011) found that when

respondents opened envelopes containing survey advance materials primarily in English, they often overlooked multilingual brochures contained in the same envelope. However, when their attention was directed to the brochures, respondents were able to extract the important messages of the advance material, e.g., a survey will be arriving in the mail, response is mandatory and confidential, and language assistance is available. Importantly, respondents reported that the existence of the brochure in their language made them feel respected and appreciated by the survey organization and a concomitant greater likelihood that they would consider it their duty to respond when the survey arrived (Chan and Pan 2011: 355).

Cognitive interviewing has also been extensively employed to probe respondents' comprehension of social constructs or practices that either do not exist in respondents' native languages or countries, or are conceived differently from how they are understood in the U.S. For example, constructions of race are markedly different in the U.S. and Latin America, with these differences reflected in the numerous U.S. Census Bureau studies that have documented the challenges faced by monolingual Spanish-speaking respondents when asked to identify their race using the categories mandated by the federal government's Office of Management and Budget (OMB) guidelines (e.g., Childs, Terry, and Jurgenson 2011; Goerman, Caspar, Sha, McAvinchey, and Quiroz 2007a; Goerman, Caspar, Sha, McAvinchey, and Quiroz 2007b; Goerman, Quiroz, McAvinchey, Reed, and Rodriguez 2013). In addition to different constructions of "race," Latin American societies generally do not distinguish between "race" and "ethnicity," and this is reflected in cognitive interview research in which many Spanish-speaking respondents considered race and ethnicity to be equivalent (Gerber and Crowley 2005). Along the same lines, many Spanish-speaking participants in another cognitive interview study preferred a combined question that includes both ethnicity and race, rather than two distinct questions (Terry and Fond 2013).

In addition to being used to pretest question wording and translations, cognitive interviews are also one method employed in usability testing. Usability testing investigates how well respondents are able to utilize an instrument, with the goal of assessing not only how well they process the wording, but also whether they are able to locate key informational elements and successfully navigate through the survey. Usability testing is a key component of pretesting computer-based and online materials. For example, a usability study of the online version of the Spanish-language Puerto Rico Community Survey (PRCS; similar to the ACS) found evidence of respondents having difficulty entering their names into the fillable fields provided. Because the use of two last names (maternal and paternal) is common in Puerto Rico, as well as other Spanish-speaking countries, many respondents were unsure whether to report just one or both, a problem that was confounded by a limited number of characters that could be entered (Leeman, Fond, and Ashenfelter 2012; see also Goerman et al. 2007a and 2007b). Also, and more specific to the PRCS, cognitive testing revealed that a question about home heating fuels did not take into account that home heating is not generally used in Puerto Rico; so although respondents understood the translation, they felt that it was not possible to answer the question accurately because the question presumes that a respondent's home must have heat (or, be in need of heat). These are the types of insights that can be gained about survey instruments that might not be identified in expert review because of the respondents' different daily experiences.

Clearly, cognitive interviewing is a useful method for gaining a better understanding of how survey questions, and their translations, are perceived by respondents. However, one caveat of conducting cognitive interviews to inform multilingual survey and materials development is that this methodology was developed primarily with respondents from North America and

Europe, whereas the vast majority of LEP individuals in the U.S. are immigrants from Latin America or Asia. There is some evidence from research utilizing cognitive interviews that LEP respondents either did not understand the process or experienced difficulty completing the tasks in line with the researchers' expectations (Carrasco 2003; Coronado and Earle 2002; Goerman, 2006a, 2006b; Kissam, Herrera, and Nakamoto 1993; Pan 2004; Pan, Landreth, Hinsdale, Park, and Schoua-Glusberg 2010; Potaka and Cochrane 2004). For example, Pan (2004, 2008) has found that the common cognitive interview technique of soliciting a paraphrase of a question or response was not effective with Chinese-speaking respondents because the exercise was unfamiliar; they would typically repeat the text as-is, and so the probe was not very useful for eliciting information. In addition, issues that are significant factors in survey responses such as social desirability, acquiescence bias, nonattitudes and neutral opinions, etc. (Smith 2004) may show cross-linguistic differences.

As a result of findings such as these, researchers have stressed the need for further studies of multilingual pretesting (Smith 2004). Specifically, cross-cultural and cross-linguistic awareness should inform the entire cognitive interview process for the development or translation of testing protocols, the tailoring interview probes, the organization of language teams, the selection and training bilingual interviewers, and the reporting of research results (Pan et al. 2007, 2010; Goerman and Casper 2010a). A more systematic approach to multilingual pretesting is needed to ensure that the impact of cultural variation is reduced to minimum and that cognitive testing results are valid and reliable.

In addition to cognitive interviewing, another common method of survey pretesting is to conduct a field test, or a small-scale practice run of data collection involving all phases of survey production, from initial contact to non-response follow-up (Groves et al., 2009). Field tests sometimes contain split-panel designs in which different segments of the survey sample receive different versions of the questions. In an effort to ensure high-quality surveys and avoid unexpected problems, the U.S. Census Bureau generally requires that new questions or changes to existing questions be field-tested prior to implementation. However, there are many logistical challenges to conducting field tests with minority languages, such as the difficulty and cost of identifying and drawing a large enough sample to have a geographically dispersed non-English speaking population. Thus, it has not been common to include non-English languages in U.S. Census Bureau field testing operations, although in recent years, there has been a greater effort to include Spanish, such as in Census non-response follow-up and field tests of the ACS (see Childs, Landreth, Goerman, Norris, and Dajani 2007, Pascale, Goerman, and Drom 2012).

It should be noted that even the rare split-panel field tests conducted by the Census Bureau in languages other than English have used a single, direct translation of each English-language question. Thus, although there are two target language versions of each question, they each correspond to a different source language question. In other words, split-panel field testing has not been used to compare between two possible translations or to assess the adequacy of a translation. As was discussed earlier, comparisons between different translations, including comparisons between *adoption*s and *adaptation*s, have generally been made based on cognitive testing results. In an institution which places high value on the quantitative data obtained in field tests, one implication of the lack of such data regarding different translations is that adaptations are sometimes seen as 'untested,' because only qualitative data are available. Given that direct word-for-word translations are perceived as maintaining cross-linguistic standardization, *adoption* is seen as the default in the absence of strong evidence for *adaptation*.

We have attempted to provide a snapshot of some of the research and practice relating to the pretesting of multilingual surveys. Expert review, cognitive interviewing, and field testing are all used to varying extents at different phases in the survey development and deployment process, and we want to emphasize the necessity for more research into how to make these methods most effective. Currently, pretesting is not always performed consistently across different surveys and completely at each phase. With that said, however, the current state of pretesting helps to get a better version to field, where the survey and its administration encounter additional challenges and concerns.

4 Fielding the Survey: Research and Practice

Even after the survey is translated and pretested, conducting a survey in a multilingual context presents several challenges. In addition to concerns such as interviewer-respondent interaction, interviewer errors, and contact strategies, there is the critical issue of the effect of language and culture on interview interaction. In particular, survey interview protocols reflect assumptions that are grounded in a set of experiences and cultural knowledge specific to the U.S. social context and these may not be shared by LEP respondents. Possible differences include whether there is similar social practice of survey research in the country of origin, the experience speakers of different languages have participating in a survey interview, and the cultural norms regarding interacting with a stranger and answering questions. In addition, carrying out the interview in a language other than English can have an impact on the meaning of the questions and the way that they are interpreted.

Until recently, research on fielding a survey with LEP respondents by the U.S. Census Bureau focused on issues associated with undercounting, such as immigrant mobility (de la Puente, Hunter, and Salo 2003); the greater frequency of complex household structures (e.g., de la Puente 1993; Schwede 2003; and Schwede, Blumberg, and Chan 2005); and the difficulty of enumerating the population of *colonias*, which have high concentrations of Spanish-speaking LEP respondents (de la Puente and Stemper 2003). During the 2010 Census, two ethnographic studies were conducted to observe live face-to-face census interviews with LEP respondents (Pan and Lubkemann, 2013) and with minority ethnic groups (Schwede and Terry 2013) for the first time. The goal of the research was to identify linguistic and cultural barriers to quality data collection. A common theme emerging in the two studies was that it is not always possible to disentangle linguistic issues from cultural issues in fielding a survey. For example, the LEP respondents from seven language groups in Pan and Lubkemann's study had no prior cultural experience with an interview event and had little knowledge of the U.S. Census. Thus, many of the interactional guidelines prescribed by the Census interview protocol not only sometimes failed to secure and sustain the interview itself, but also sometimes inhibited it. Bilingual interviewers had to depart from the interview protocol in order to obtain the needed information from respondents. This raises a key methodological question of how to design an interview protocol that can achieve the intended communicative effect. Research is needed on how to tailor interview interactions to specific cultural norms while maintaining the standardization of survey interviews.

The intertwining of linguistic and cultural issues can also be seen in a recent behavior coding study from a 2010 Census follow-up operation conducted by telephone in English and Spanish (Childs, Leeman and Smirnova 2012). Analysis of cross-linguistic differences revealed that while the English-language version of the yes/no Hispanic Origin question (see Figure 3

below) did not appear to cause problems, in Spanish, respondents sometimes interpreted it as a three-way choice, between “hispano,” “latino” and “origen español.”

The screenshot shows a software window titled "U.S. Census Bureau Coverage Followup". At the top left, there are buttons for "Help F1", "Roster F3", and "Exit F10". Below these, it says "SECTION G: Demographics - GHO_A". On the right side, there are buttons for "Interview F11" and "Jump Back F12". The main content area contains the question: "Are you of Hispanic, Latino, or Spanish origin?" with two radio button options: "1. Yes" and "9. No". At the bottom left, there is a link "Back to Previously Completed Loop" and a "Case Id: 13483856". In the center, there is a "Previous F5" button with a left arrow. On the right, there is a "Next F8" button with a right arrow. The "U.S. CENSUS BUREAU" logo is centered at the bottom.

Spanish Text:

¿Es {fill usted/FULL NAME} de origen hispano, latino o español?

Figure 3. Hispanic Origin question from 2010 Census Follow-Up telephone instrument.

The study identified several factors that likely contributed to the difference between Spanish- and English-language response patterns, some of them related to the differences in Spanish and English word order, and others related to the different meanings of “Spanish” and “Spanish origin” in English and Spanish. Of particular interest here is that the language in which the interview is conducted has an impact on what questions might be considered reasonable. Because conversational interaction is governed by the assumption that questions do not request obvious or redundant information (Grice 1977), and because in the U.S., Spanish is widely associated with Hispanic identity, it is unusual for someone speaking Spanish to ask another person speaking Spanish whether or not they are Hispanic. Thus, the interpretation of the Spanish-language question as requesting a “yes/no” response, rather than a choice, is less likely. In contrast, such a question would be less unusual in an English-language conversation or interview, where there would be no default presupposition of Hispanic/Latino identity, especially over the telephone. In addition to underscoring the importance of testing questions with various demographic groups, the results of this study also suggest that it is insufficient to examine the comparability of survey

questions in isolation. Instead, the broader discursive and interactional context must be taken into account.

Another issue facing survey researchers is the use of interpreters in field interviews. As it is not practical to translate a survey into every language, field interviewers must often rely on a bilingual person to interpret the interview. Field interviewers draw on a wide range of bilinguals with a wide variety of language competency and interpreting skills, ranging from professional interpreters to neighbors to household children. The use of interpreters, and the great variety among interpreters used, raises the question of how to ensure data quality. Additionally, although survey interviewers normally are trained to read each question as worded to lessen measurement error, in practice standardization of survey interviews – even interviews with no interpreter involved – is challenging (Houtkoop-Steenstra, 2000). Standardizing interpreted interviews presents an even greater challenge; because there is no single correct translation of a question, there will inevitably be some variation when survey interviews are conducted through interpreters.

Research has revealed that the interpreter's role has a great impact on the interview situation and its outcomes. Interpreting for survey interviews requires unique skills: language, culture, and subject matter (Freed, 1988). Unskilled or untrained interpreters sometimes deviate widely from the original question meaning, possibly compromising data quality (Doerr, 2005, Kapborg and Bertero 2002; Martir and Willis 2004). Untrained interpreters also sometimes dominate, or even take over survey interviews, as was documented in ethnographic Census Bureau research in the field (Pan, 2007). To minimize these negative effects, the Census Bureau has developed guidelines for survey interviewing (Pan, Leeman, and Fond 2013). In addition to an overview of best practices for the interpreters, the guidelines also provide a checklist of questions that interviewers can use to choose among various potential interpreters (e.g., 'Do you have any formal education in the target language?'), as well as a list of "Dos and Don'ts" regarding how to conduct the interview through an interpreter (e.g., 'Do maintain eye contact with the respondent; Don't direct your questions to the interpreter').

5 Next Steps

As we have shown, multilingual survey development is significantly more complex than simply adopting source-language data collection instruments and procedures into target languages. In order to ensure functional equivalency across versions in different languages, multilingual surveys require instruments that are linguistically, culturally, and socially appropriate in each language, and the development of such instruments involves extensive review and in-language pretesting. By synthesizing research on multilingual survey research conducted at the US Census Bureau, we examined some of the methodological and institutional challenges for a large statistical agency working within the constraints of not being able to carry out simultaneous parallel development of surveys in multiple languages 'from scratch.' In addition to illustrating the shortcomings of the *adoption* approach and the impact of cross-linguistic differences, we also presented the kinds of studies needed to identify and address such differences.

Whereas the multilingual survey development procedures adopted by U.S. Census Bureau in the last decade represent important advances, most multilingual pretesting is limited to cognitive or usability testing with a small sample size of LEP respondents. A key next step in improving the development and deployment of multilingual surveys will be field testing to

further investigate effects on data quality and methods to increase response rates and decrease item non-response rates. Another vital area for continued research and innovation is the development of culturally, as well as linguistically appropriate pretesting methods; as our synthesis showed, just as the *adoption* approach to survey translation can be inadequate, so too the default *adoption* approach to testing protocols and instruments is also problematic.

One theme that runs throughout our discussion, and which represents a particularly pressing issue for multilingual survey development is the tension between standardizing different language versions of a multilingual survey and attending to the cultural and linguistic specifics of each version. Underlying this tension is the issue of whether standardization should focus on stimulus inputs (i.e., question wording and administration) and stimulus outputs (i.e., communicative effect). This question highlights another key area for research going forward: how to ensure that differentiated stimulus inputs actually produce comparable outputs (e.g., data). In addition to the specifics involved in the development of any particular multilingual survey, it is also these broader methodological inquiries that need our attention to ensure the highest quality data collection in a linguistically diverse nation.

References

- Bates, Nancy and Pan, Yuling. 2009. "Motivating Non-English-Speaking Populations for Census and Survey Participation." In *Proceedings for the 2009 Conference for the Federal Committee on Statistical Methodology*. Washington, DC. Federal Committee on Statistical Methodology.
- Behling, Orlando, and Kenneth S. Law, eds. 2000. *Translating questionnaires and other research instruments: Problems and solutions*. SAGE Publications.
- Blair, Johnny, and Linda Piccinino. (2005). "The development and testing of instruments for cross-cultural and multi-cultural surveys." In Hoffmeyer-Zlotnik, J.H.P. and Harkness, J. (eds). *Methodological Aspects in Cross-National Research*. Pp. 13-30. Mannheim, Germany: ZUMA Nachrichten Spezial, Volume 11.
- Blommaert, Jan, and Ben Rampton. 2012. *Language and superdiversity*. MPIMMG.
- Braun , Michael, and Janet A. Harkness. 2005. "Text and context: challenges to comparability in survey questions. In J.P. Hoffmeyer-Zlotnik and J.A.Harkness (Eds.), *Methodological aspects in cross-national research* (pp. 95-107). Mannheim, Germany: ZUMA.
- Carrasco, L. 2003. *The American Community Survey (ACS) en Español: Using cognitive interviews to test the functional equivalency of questionnaire translations*. Statistical Research Division Research Report Series (Survey Methodology #2003-17). Washington, DC: U.S. Census Bureau.
- Chan, Anna and Pan, Yuling. 2011. "The Use of Cognitive Interviewing to Explore the Effectiveness of Advance Supplemental Materials among Five Language Groups." *Field Methods*. 23(4)342-361.
- Childs, Jennifer Hunter, Ashley Landreth, Patricia Goerman, Dawn Norris, and Aref Dajani. 2007. "Behavior coding analysis report: Evaluating the English and the Spanish versions of the non-response follow-up (NRFU) for the 2006 Census Test." Statistical Research Division Study Series (Survey Methodology# 2007-16). Washington, DC: U.S. Census Bureau.
- Childs, Jennifer Hunter, Rodney Terry, and Nathan Jurgenson. 2011. "Measuring Race and Hispanic Origin: Cognitive Test Findings Searching for "Truth"." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 111:26-42.
- Childs, Jennifer Hunter, Jennifer Leeman, and Michelle Smirnova. 2012. Behavior Coding Report of 2010 Census Coverage Followup English and Spanish Interviews. Research and Methodology Directorate, Center for Survey Measurement Study Series (Survey Methodology #2012-08). Washington, DC: U.S. Census Bureau.

- Conway, K., Patrick, D. L., Gauchon, T., and Acquadro, C. 2010. "Enhancing cross-cultural appropriateness for newly developed patient-reported outcome (PRO) instruments: The use of translatability assessment." *PRO Newsletter*, 44.
- Coronado, I., and D. Earle. 2002. "Effectiveness of the American Community Survey of the U.S. Census in a Borderlands Colonia Setting." *Draft report submitted to the U.S. Census Bureau*.
- de la Puente, Manuel. 1993. "Why are People Missed or Erroneously Enumerated in the Census: A Summary of Findings from Ethnographic Research." *Proceedings of the 1993 Research Conference on Undercounted Ethnic Populations*, pp. 29-66.
- de la Puente, Manuel, and D. Stemper. 2003. "The Enumeration of Colonias in Census 2000: Perspectives of Ethnographers and Census Enumerators." *Census 2000 Evaluation J.4: Ethnographic Studies*, September 22, 2003.
<http://www.census.gov/pred/www/rpts/J.4.pdf>
- de la Puente, Manuel, Hunter, J. E., and Salo, M. 2003. "Comparative Ethnographic Research on Mobile Populations." *Census 2000 Evaluation J.3: Ethnographic Studies*. July 17, 2003.
<http://www.census.gov/pred/www/rpts/J.3.pdf>
- Dean, Elizabeth, Rachel Caspar, Georgina McAvinchey, Leticia Reed, and Rosanna Quiroz. 2007. "Developing a Low-Cost Technique for Parallel Cross-Cultural Instrument Development: The Question Appraisal System (QAS-04)." *International Journal of Social Research Methodology* 10:227-241.
- Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*, New York: Wiley.
- Dillman, Don A. 1991. "The Design and Administration of Mail Surveys." *Annual Review of Sociology* 17:225-249.
- Doerr, Lauren. 2005. Lost in translation: Data quality and interpreters in multilingual research: Towards an interpreting methodology. Paper presented at the Annual Conference of the American Association for Public Opinion Research. Miami, FL. May 12-15, 2005.
- Dorer, B. 2011. Advance translation in the 5th round of the European Social Survey (ESS). FORS Working Paper Series, paper 2011-4. Lausanne: FORS.
- Fitzgerald, Rory, Sally Widdop, Michelle Gray, and Debbie Collins. 2011. "Identifying sources of error in cross-national questionnaires: application of an error source typology to cognitive interview data." *Journal of Official Statistics* 27: 569-599.
- Forsyth, Barbara H., Martha Stapleton Kudela, Kerry Levin, Deirdre Lawrence, and Gordon B. Willis. 2007. "Methods for translating an English-language survey questionnaire on tobacco use into Mandarin, Cantonese, Korean, and Vietnamese." *Field Methods* 19:264-283.

- Fox, Richard J., Melvin R. Crask, and Jonghoom Kim. 1988. "Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response." *Public Opinion Quarterly* 52:467-491.
- Freed, Anne O. 1988. "Interviewing through an interpreter." *Social Work* 33:315-319.
- Gerber, Eleanor, and Melinda Crowley. 2005. "Report on Cognitive Testing of a Shortened Sequence of Hispanic Origin, Race, and Modified Ancestry Questions: Content Development for the 2005 National Content Test." *Report for the Statistical Research Division, Washington, DC: U.S. Census Bureau.*
- Goerman, Patricia L. 2006a. Adapting cognitive interview techniques for use in pretesting Spanish language survey instruments. Statistical Research Division Study Series (Survey Methodology #2006-03). Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/srd/papers/pdf/rsm2006-03.pdf>.
- Goerman, Patricia L. 2006b. "An Examination of Pretesting Methods for Multicultural, Multilingual Surveys: The Use of Cognitive Interviews to Test Spanish Instruments." Conducting Cross-National and Cross-Cultural Surveys. ZUMA Nachrichten Spezial Band 12:67-79.
- Goerman, Patricia, and Rachel Casper. 2010. "Managing the Cognitive Pretesting of Multilingual Survey Instruments: A Case Study of Pretesting of the U.S. Census Bureau Bilingual Spanish/English Questionnaire" in *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Harkness, J., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P. P., Pennell, B. and Smith, T.W. 75-90. John Wiley and Sons.
- Goerman, Patricia L., Rachel Caspar, Mandy Sha, G. McAviney, and Rosanna Quiroz. 2007a. Census Bilingual Questionnaire Research Final Round 1 Report. Statistical Research Division Research Report Series (Survey Methodology #2008-1). Washington, DC: U.S. Census Bureau.
- Goerman, Patricia L., Rachel Caspar, Mandy Sha, G. McAviney, and Rosanna Quiroz. 2007b. Census Bilingual Questionnaire Research Final Round 2 Report. Statistical Research Division Research Report Series (Survey Methodology #2007-27). Washington, DC: U.S. Census Bureau.
- Goerman, Patricia L., Leticia Fernández, and Rosanna Quiroz. 2013a. "Translation of U.S. Educational Level Survey Questions into Spanish: Is Adaptation the Solution?" In *JSM Proceedings*, Statistical Computing Section. Alexandria, VA: American Statistical Association 4289-4303.
- Goerman, Patricia L., Rosanna Quiroz, Georgina McAviney, Leticia Reed and Sonia Rodríguez. 2013b. Census American Community Survey Spanish CAPI/CATI Instrument Testing Phase I - Round 1 Final Report. *U.S. Census Bureau Statistical Research Division Report Series*. <https://www.census.gov/srd/papers/pdf/ssm2013-21.pdf>

- Grice, H. Paul. 1975. Logic and conversation. In *Syntax and semantics 3: Speech acts*, ed. Peter Cole and Jerry L. Morgan, 41-58. New York: Academic Press.
- Groves, Robert M., Floyd J. Fowler Jr, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey methodology*. Wiley.
- Harkness, Janet. 2003. Questionnaire translation. In *Cross-cultural survey methods*, edited by J. A. Harkness, F. J. R. van de Vijver, and P. P. Mohler 35-56. Hoboken, NJ: John Wiley and Sons.
- Harkness, Janet A., and Alisú Schoua-Glusberg. 1998. Questionnaires in translation. In *ZUMA-Nachrichten Spzial No. 3. Cross-cultural survey equivalence*, edited by J. A. Harkness 87-127. Mannheim: ZAMA
- Harkness, Janet A., Peter Ph Mohler, and Fons JR van de Vijver. 2003. Comparative research. In *Cross-cultural survey methods*, edited by Janet A. Harkness, Fons JR Van de Vijver, and Peter Ph Mohler 3-16. Hoboken, NJ: John Wiley and Sons.
- Harkness, Janet. A., Ana Vilar, and Brad Edwards. 2010. Translation, adaptation and design. In *Survey methods in multinational, multiregional and multicultural contexts* edited by Janet A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, B. E. Pennell, and T. W. Smith 117-140. Hoboken, NJ: Wiley.
- Houtkoop-Steenstra, Hanneke. 2000. *Interaction and the standardized survey interview: The living questionnaire*. Cambridge University Press.
- IOM (International Organization for Migration). 2010. World migration report 2010 - The Future of Migration: Building Capacities for Change. Downloaded January 26, 2013: http://publications.iom.int/bookstore/index.php?main_page=product_info&products_id=653
- Kapborg, Inez and Carina Bertero. 2002. "Using an interpreter in qualitative interviews: Does it threaten validity?" *Nursing Inquiry*, 9:52-56.
- Kim, Jennifer, and Jonathan Zapata. 2012. "2010 Census: Language Program Assessment Report." 2010 Census Program for Evaluations and Experiments. 2010 Census Planning Memoranda Series #204.
- Kissam, E., E. Herrera, and J. M. Nakamoto. 1993. Hispanic response to Census enumeration forms and procedures." Contract report submitted to the U.S. Census Bureau.
- Leeman, Jennifer, Marissa Fond, and Kathleen T. Ashenfelter. 2012. Cognitive and Usability Pretesting of the Online Version of the Puerto Rico Community Survey in Spanish and English. *Research and Methodology Directorate, Center for Survey Measurement Study Series (Survey Methodology #2012-09)*. Washington, DC: U.S. Census Bureau.

- Levin, Kerry, Gordon B. Willis, Barbara H. Forsyth, Alicia Norberg, Martha Stapleton Kudela, Debra Stark, and Frances E. Thompson. 2009. "Using Cognitive Interviews to Evaluate the Spanish-Language Translation of Dietary Questionnaire." *Survey Research Methods* 3:13-25.
- Matir, Javier F. Sevilla and Deanna R. Willis. 2004. "Using bilingual staff members as interpreters." *Family Practice Management* 11:34-36.
- McKay, Ruth B., Martha J. Breslow, Roberta L. Sangster, Susan M. Gabbard, Robert W. Reynolds, Jorge M. Nakamoto, and John Tarnai. 1996. "Translating survey questionnaires: Lessons learned." *New Directions for Evaluation* 70:93-104.
- Pan, Yuling. 2004. "Cognitive Interviews in Languages Other Than English: Methodological and Research Issues." In *Joint Statistical Meeting Proceedings*, Alexandria, VA: American Statistical Association.
- Pan, Yuling. 2007. Development of guidelines on the use of interpreters in survey interviews. *Statistical Research Division Study Series (Survey Methodology #2007-4)*. Washington, DC: U.S. Census Bureau.
- Pan, Yuling. 2008. "Cross-cultural Communication Norms and Survey Interviews." In Sun, H. and Kádár, D. (eds.), *It's the Dragon's Turn – Chinese Institutional Discourse(s)*, 17-76. Berne: Peter Lang.
- Pan, Yuling and Marissa Fond. 2010. "Developing multilingual questionnaires: A sociolinguistic perspective." In *Proceedings of the 2010 Joint Statistical Meetings*, Alexandria, VA: American Statistical Association.
- Pan, Yuling and Marissa Fond. 2012. "Beyond words: In search of effective methods for translating survey letters." Paper presented at the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations. New Orleans, LA. Oct. 31-Nov. 3, 2012.
- Pan, Yuling and Marissa Fond. (forthcoming). "Evaluating Multilingual Questionnaires: A Sociolinguistic Perspective." *Survey Research Methods*.
- Pan, Yuling and Ashley Landreth. 2009. "Conveying Translated Informed Consent Concepts: Effects of Language and Culture on Interpretation of Legally Required Messages." In *Joint Statistical Meeting Proceedings*, Alexandria, VA: American Statistical Association
- Pan, Yuling and Stephen Lubkemann. 2013. "Observing Census Enumeration of Non-English Speaking Households in the 2010 Census: Evaluation Report." *Research and Methodology Directorate, Center for Survey Measurement Study Series (Survey Methodology #2013-02)*. U.S. Census Bureau. Available online at <http://www.census.gov/srd/papers/pdf/ssm2013-02.pdf>.

- Pan, Yuling and Stephen Lubkemann. (forthcoming). "Standardization and Meaning in the Survey of Linguistically-Diversified Populations: Insights from the Ethnographic Observation of Linguistic Minorities in 2010 Census Interviews." In R. Tourangeau, N. Bates, B. Edwards, T. P. Johnson, and K. M. Wolter (eds.) *Hard-to-Survey Populations*. Cambridge: Cambridge University Press.
- Pan, Yuling, Marjorie Hinsdale, Hyunjoo Park, and Alisú Schoua-Glusberg. 2006. "Cognitive Testing of Translations of ACS CAPI Materials in Multiple Languages." *Statistical Research Division Research Report Series (Survey Methodology #2006/09)*. Washington, DC: U.S. Census Bureau.
- Pan, Yuling, Marjorie Hinsdale, Hyunjoo Park, and Alisú Schoua-Glusberg. 2008. "Cognitive Testing of ACS Multilingual Brochures." In *Statistical Research Division Research Report Series (Survey Methodology #2008-6)*. Washington, DC: U.S. Census Bureau.
- Pan, Yuling, Ashley Landreth, Marjorie Hinsdale, Hyunjoo Park, and Alisú Schoua-Glusberg. 2007. "Methodology for Cognitive Testing of Translations in Multiple languages." In *Joint Statistical Meeting Proceedings*, Alexandria, VA: American Statistical Association.
- Pan, Yuling, Ashley Landreth, Marjorie Hinsdale, Hyunjoo Park, and Alisú Schoua-Glusberg. 2010. "Cognitive interviewing in non-English languages: A cross-cultural perspective." In Janet Hardness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars. Lyber, Peter Ph. Mohler, Beth-Ellen Pennell, Tom W. Smith (eds.), *Survey methods in multinational, multiregional, and multicultural contexts*, 91–113. Berlin: Wiley Press.
- Pan, Yuling, Jennifer Leeman, and Marissa Fond. (2013). "Development of Census Bureau Survey Interpretation Guidelines." Center for Survey Measurement Study Series (Survey Methodology #2013-27). U.S. Census Bureau. Available online at <http://www.census.gov/srd/papers/pdf/ssm2013-27.pdf>.
- Pan, Yuling, Mandy Sha, Alisú Schoua-Glusberg, and Hyunjoo Park. 2009. "2010 Census Language Program: Pretesting of Census 2010 Questionnaire in Five Languages." *Statistical Research Division Research Report Series (Survey Methodology #2009-01)*. Washington, DC: U.S. Census Bureau.
- Pascale, Joanne, Patricia Goerman, and Katherine Drom. (2012) "2010 ACS Content Test Evaluation: Behavior Coding Results." 2012 American Community Survey Research and Evaluation Report Memorandum Series #ACS12-RER-12. Available online at: https://www.census.gov/acs/www/Downloads/library/2013/2013_Pascale_01.pdf
- Pascale, Joanne, Jonathan Rodean, Jennifer Leeman, Carol. Cosenza, and Alisu Schoua-Glusberg. (Forthcoming). Preparing to measure health coverage in federal surveys post-reform: Lessons from Massachusetts. *Inquiry*.

- Potaka, Lynn, and Suzanne Cochrane. 2004. Developing bilingual questionnaires: Experiences from New Zealand in the development of the 2001 Maori language survey. *Journal of Official Statistics*, 20: 289-300.
- Presser, Stanley, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. 2004. "Methods for testing and evaluating survey questions." *Public opinion quarterly* 68:109-130.
- Saville-Troike, Muriel. 1989. *The ethnography of communication*. Cambridge: Basil Blackwell.
- Schwede, Laurel. 2003. "Complex Households and Relationships in the Decennial Studies and in Ethnographic Studies of Six Race/Ethnic Groups." Census 2000 Testing, Experimentation and Evaluation Program Ethnographic Research Report. U.S. Census Bureau. <http://www.census.gov/pred/www/rpts/Complex Households Final Report.pdf>
- Schwede, Laurel, and Rodney Terry. 2013, "Comparative Ethnographic Studies of Enumeration Methods and Coverage across Race/Ethnic Groups." *2010 Census Program for Evaluations and Experiments Evaluation. 2010 Census Planning Memoranda Series, No. 255*. March 29, 2013.
- Schwede, Laurel, Rae Lesser Blumberg, and Anna Y. Chan, eds. 2005. *Complex ethnic households in America*. Rowman and Littlefield.
- Smith, Tom W. 2004. "Developing and evaluating cross-national survey instruments." *Methods for testing and evaluating survey questionnaires* 431-452.
- Terry, Rodney L., and Marissa Fond. 2013. "Experimental US Census Bureau Race and Hispanic Origin Survey Questions Reactions From Spanish Speakers." *Hispanic Journal of Behavioral Sciences* 35:524-541.
- U.S. Census Bureau. 2003. "Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses." Online, Available: <http://www.census.gov/srd/pretest-standards.pdf>
- Willis, Gordon. 2005. *Cognitive interviewing: A tool for questionnaire design*. Thousand Oaks, CA: Sage.
- Willis, Gordon, Martha Stapleton Kudela, Kerry Levin, Alicia Norberg, Debra S Stark, Barbara H. Forsyth, Pat Dean Brick, David Berrigan, Frances E. Thompson, Deirdre Lawrence, and Ann M. Hartman, 2010. E Evaluation of a Multistep Survey Translation Process. In Janet Hardness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars. Lyber, Peter Ph. Mohler, Beth-Ellen Pennell, Tom W. Smith (eds.), *Survey methods in multinational, multiregional, and multicultural contexts*, 141-156. Berlin: Wiley Press.

Willis, Gordon, and Elaine Zahnd. 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.

Yammarino, Francis J., Steven J. Skinner, and Terry L. Childers. 1991. "Understanding Mail Survey Response Behavior." *Public Opinion Quarterly* 55:613-639.