

# Number and Nationality of Spanish-Speaking Immigrant Respondents for U.S. Cognitive Testing Studies<sup>1</sup>

Patricia L. Goerman<sup>1</sup>

<sup>1</sup>Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road,  
Washington, DC, 20233

## Abstract

A basic question that researchers must answer in any cognitive testing project is: How many respondents should be included and what characteristics should they have? These questions are particularly relevant when dealing with the Spanish-speaking population in the U.S., a group made up of people from many different national origins. This paper examines strategies for choosing the number and type of Spanish-speaking respondents used in two U.S. Census Bureau cognitive interview projects. On the whole, it appears that less national-origin diversity among respondents is necessary when testing basic demographic questions than when testing questions related to more specialized subject areas. Other factors that need to be considered are regional diversity within the U.S., and the subject matter being tested in the survey. The paper concludes with recommendations for more systematic research.

**Key Words:** Pretesting non-English surveys, Cognitive interviewing and Spanish-speaking respondents, respondent characteristics

## 1. Introduction

The question of number and types of respondents with whom to test survey questions is a basic issue that arises at the beginning of every cognitive testing project. Many agencies or organizations have a standard number of cognitive interview respondents that are used in English-language studies, ranging anywhere from five to hundreds of interviews (Willis, 2005; Brick and Blair, 2009).

The type of questions being tested often adds an important dimension to the discussion. For example, when testing a question such as the Census Bureau's "race" question, which in the 2010 Census contained a total of 15 check boxes representing different groups, as well as a separate "Hispanic origin" question, it may not be possible to test with a small number of respondents and get the most desirable results. In this type of situation, researchers are confronted with the decision of whether to include respondents who would be likely to choose each category (or not fit into any categories) and they must also consider how many people from each group are needed to ensure that the question is "working" properly. If a group of unrelated questions is being tested, the researchers must also decide whether recruitment for a given question should be prioritized over others. Available time and resources are often a deciding factor in this type of situation.

Things become even more complicated when testing a survey translation in a language in which the target population is comprised of people from many national origin

---

<sup>1</sup> This report is released to inform interested parties of research and to encourage discussion. The views are those of the author and not necessarily those of the U.S. Census Bureau.

backgrounds, such as the Spanish-speaking population in the United States. In this case researchers need to make sure that a term used by people from one country does not have a different or inappropriate meaning to those from another. Little, if any, research exists on the ideal number and types of respondents to use in this type of research.

### **1.1 Review of the Literature on Number of Cognitive Interview Respondents**

Very little empirical research exists on the ideal number of cognitive interview respondents in survey pretesting. In his book on cognitive interviewing, Willis (2005) reports that a typical number of respondents per round of testing at many institutions is between five and 15 people (p. 7).

Based on qualitative interviews with six experienced cognitive interviewers from federal agencies, government contracting agencies and universities, Brick and Blair (2009) report that many researchers used to do two rounds of nine interviews per project, but now often do anywhere from 30-60 interviews to hundreds of interviews on larger projects. However, the numbers vary a great deal across projects, researchers and institutions.

Blair et al. (2006) conducted empirical research in an attempt to quantify the effect of cognitive interview sample size on cognitive interview findings. They embedded problems in a series of English-language survey questions and conducted a total of 90 cognitive interviews. They then drew sample sizes between five and 50 and found that small numbers of interviews failed to detect many serious problems with the questions. All of the cognitive interviews in their study were conducted using previously scripted probes, which many would argue is not an ideal method when working with experienced cognitive interviewers. If allowed to deviate from a script, experienced interviewers might be able to identify greater numbers of problems in an interview than they could by using scripted probes alone. Nevertheless, the researchers found that while many high impact problems were detected in smaller sample sizes, other high impact problems continued to be uncovered in relatively large sample sizes, up to 50 interviews and beyond. While they were unable to quantify an ideal number of interviews per project based on their research, they do recommend conducting greater numbers of interviews than are often currently done in the industry.

What is often left out of this discussion is cultural variation among the target population. In terms of types of respondents, Willis (2005) recommends that respondents have as wide a variety of characteristics as possible in order to help researchers identify problems that might be particular to a given group (p. 140). However, he does not provide guidance in terms of choosing the type of variation in characteristics or the number of respondents with a given characteristic.

Fernández, et.al (2009b) have found that even when researchers do not have the time or resources to include respondents with every desired characteristic, it is worthwhile to include as many different types of respondents as possible. Fernández, et al. were tasked with cognitively testing five different experimental versions of the Census Bureau's race and Hispanic origin questions under a very limited timeline. It was not going to be possible to interview numerous people who would be likely to mark every box in multiple versions of the Census Bureau's lengthy race question in the time allowed. The researchers therefore conducted between eight and 14 interviews with people of varying characteristics with each of the five questionnaire versions. The interviews were supplemented with an expert review of each of the forms. The researchers found that cognitive interviews confirmed many (but not all) of the predictions made by expert

reviewers in terms of where problems might arise with the forms. However, additional issues were found by cognitive interviews alone. These were particular to respondents of unique backgrounds and were things that had not been predicted by expert reviewers.

The question that remains is how many people with a given characteristic should be included in the respondent pool. There is a decided lack of research that addresses this issue. There is even less research on the number of respondents to use in the testing of translations through cognitive interviews. Spanish-speaking respondents in the United States are a group that is comprised of people from many different national origin backgrounds. There can be differences in terminology used across countries and presumably respondents will be more or less acculturated in the United States.

Researchers doing cognitive testing with Spanish speakers in the United States not only must decide how many interviews to conduct, but which national origin groups to include and how many respondents from each one will be sufficient. Some factors that must be considered are other respondent characteristics that are of interest such as age, gender, educational level, and acculturation in the United States. In addition, researchers are sometimes testing questions that are relevant to respondents with very specific characteristics, such as smokers, people who have a disability or people who own their own homes. Diversity in terms of this type of respondent characteristic is often necessary. Finally, time and resources available for the project are a major factor. It is often the case that to include a sufficient number of respondents from a given national origin background in Spanish-language projects, the researchers may need to travel to a particular part of the United States. This can be costly and time consuming.

## **2. The Research Problem**

This paper examines the issue of choosing the number and types of respondents to include in Spanish-language cognitive testing studies in the United States. The paper presents experiences from two recent U.S. Census Bureau studies in which Spanish-language survey translations were cognitively tested. The paper discusses how the number and types of respondents were chosen for each study and how the results differed in our findings based on the national origin of respondents and types of questions tested. The paper concludes with a discussion of lessons learned and proposals for future empirical research to help determine the ideal number and types of respondents for testing survey translations.

### **3. Study 1: The Pretesting of the Decennial Census Bilingual Questionnaire**

The first project is the cognitive testing of the Census Bureau's first bilingual (Spanish and English) paper questionnaire which was used in the 2010 Census. The cognitive testing for this project was carried out in collaboration between Census Bureau and RTI International researchers. This was one of the first large-scale Spanish-language cognitive testing projects conducted at the Census Bureau and it took place between 2005 and 2007.

The fact that this work was carried out as a contracted project brings up an important issue. When structuring a typical firm, fixed price contract, the number of cognitive

interviews to be conducted must be planned in advance and specified in the contract. There is often no room to continue doing interviews until a “saturation point” is reached and interviewers deem that the same problems are arising and that most unique problems have been identified. Deciding the number of interviews that will be done in advance can be problematic if issues are still arising or if researchers deem that a change in approach is warranted mid-project.

Because ample time and resources were available and because we wanted to take the diversity of the Spanish-speaking population in the U.S. into account, we initially planned to conduct between 40 and 50 interviews in each of two iterative rounds of testing. My discussion here focuses on the first round of testing, for which 44 Spanish language interviews were ultimately conducted (Goerman, et al., 2007).

### **3.1 Decennial Project: Questionnaire Content**

In terms of questionnaire content, the Decennial Census contains mostly basic demographic questions such as questions about name, age, sex, Hispanic origin, and race. There are a few exceptions, such as questions about the number of residents in the household, whether the residence is owned or rented and some questions to see if anyone has been inadvertently counted or omitted from the form. We were therefore not concerned about choosing respondents who would follow specialized skip patterns, although we wanted to seek variation in terms of the number of residents in each household.

### **3.2 Decennial Project: Respondent Recruitment Plan**

The recruitment plan was created in collaboration between the Census Bureau and the contractor. In the end, the plan called for respondents to be recruited in about equal numbers from four different groups: 1) Mexico, 2) Central America, 3) South America, and 4) an aggregate group of Cuba, the Dominican Republic, Puerto Rico and Spain. Exact countries were not specified in the Central and South American categories and the recruiters were not given specific numbers to look for from each country in the aggregate group. They were basically asked to seek as much diversity as possible. The recruiters were also asked to seek diversity in terms of age, gender, and educational level. In order to expedite finding respondents from different national origin backgrounds, respondents were recruited in four sites, Chicago, Miami, Los Angeles and Raleigh-Durham, North Carolina.

### **3.3 Decennial Project: National Origin of Respondents**

The final national origin breakdown of respondents was a bit uneven across the four groups (see Table 1). There were 12 respondents from Mexico, 14 people from four different Central American countries, 10 people from four different South American countries and eight people from the aggregate group, which in the end only included two different countries. We did not collect information about the length of time the respondents had resided in the U.S. or whether they lived among people from different national origin backgrounds, but we did seek to interview respondents who spoke little or no English.

**Table 1:** Project 1: National Origins of Respondents in Decennial Census Study

<i>Group</i>	<i>Country of origin</i>	<i>Number</i>	<i>Total</i>
1. Mexico	Mexico	12	12
2. Central America	Guatemala	5	14
	Honduras	4	
	El Salvador	4	
	Nicaragua	1	
3. South America	Peru	4	10
	Colombia	4	
	Venezuela	1	
	Argentina	1	
4. Aggregate	Cuba	5	8
	Puerto Rico	3	
Total:			44

### 3.4 Decennial Project: Results by Nationality

We found that in testing the basic demographic questions included in the Decennial Census form, few major linguistic differences were noted across the national origin groups. Most of our findings of difficult terms and concepts held true across all groups. For example, the bilingual Census form contains a question on household tenure (see Figure 1). One of the response options read: “Rented for cash rent” in English and “Alquilada por pago en efectivo” in Spanish. We found that respondents of many national origin groups were confused by the reference to “cash” or “pago en efectivo” and they often reported that they paid their rent “by check.” In fact, a second round of testing that included English speakers demonstrated that this problem was not restricted to only Spanish speakers and the term “cash” was eventually dropped from both the English and Spanish versions of the question.

This same response option contained a finding that was particular to one national origin group. We found that Mexican origin respondents tended to use the term “rentada” instead of “alquilada” to refer to the concept of “rented.” Because these same respondents did understand the term “alquilada” and this term seemed to be well understood (and used) by people of other national origin backgrounds, we did not recommend that a change be made.

<p><b>3</b> Is this house, apartment, or mobile home . . .</p> <p>Mark <input checked="" type="checkbox"/> ONE box.</p> <p><input type="checkbox"/> Owned by you or someone in this household with a mortgage or loan?</p> <p><input type="checkbox"/> Owned by you or someone in this household free and clear (without a mortgage or loan)?</p> <p><input type="checkbox"/> Rented for cash rent?</p> <p><input type="checkbox"/> Occupied without payment of cash rent?</p>	<p><b>3</b> ¿Es esta casa, apartamento o casa móvil . . .</p> <p>Marque <input checked="" type="checkbox"/> UNA casilla.</p> <p><input type="checkbox"/> Propiedad suya o de alguien en este hogar con una hipoteca o préstamo?</p> <p><input type="checkbox"/> Propiedad suya o de alguien en este hogar libre y sin deuda (sin una hipoteca o préstamo)?</p> <p><input type="checkbox"/> Alquilada por pago en efectivo?</p> <p><input type="checkbox"/> Ocupada sin pago de alquiler en efectivo?</p>
--	--

**Figure 1:** Snapshot of the “tenure” question from the 2005 test version of the Census Bureau’s bilingual census form.

There was an additional item on the questionnaire that had content that was problematic both across groups and for a specific group. This was true of the “overcount” question; a question designed to flag people that may have been erroneously included on the census form at a given address (see Figure 2).

We found that Spanish speakers of many national origin backgrounds had difficulty with the response option “Para quedarse en una residencia estacional o una segunda residencia,” the translation for “To stay at a seasonal or second residence.” Many people, across nationality groups, interpreted “estacional” or the original translation for “seasonal” to mean something like “stationary” or “permanent,” causing the response option to have the opposite of the intended meaning.

Mexican origin respondents had a particular understanding of a term in this same response option. A number of these respondents reported that the term “residencia” or “residence” made them think of a mansion or a grand home. Taking the “seasonal” misinterpretation into account, we found that some respondents were therefore interpreting the response option as asking if the person sometimes lives or stays in a “permanent mansion or second home.”

These problems were deemed significant enough to warrant a change to both terms. The new phrase “vivienda de temporada,” or “temporary/seasonal dwelling,” was tested with respondents from various national origin groups in a second round of testing. The new wording was found to have eliminated the problems for all national origin groups and the Census Bureau went forward with this change.

**11** Does Person 1 sometimes live or stay somewhere else?  
 No  Yes - Mark  all reasons that apply.  
 To attend college  For a child custody arrangement  
 To be closer to work  While in jail or prison  
 While in the military  While in a nursing home  
 To stay at a seasonal or second residence  For another reason  
 → If more people live here, continue with Person 2.

**11** ¿Vive o se queda a veces la Persona 1 en algún otro lugar?  
 No  Sí - Marque  todas las razones que apliquen.  
 Para asistir a la universidad  Por un arreglo de custodia de niños  
 Para estar más cerca del lugar de trabajo  Mientras está en la cárcel o prisión  
 Mientras está en el servicio militar  Mientras está en un hogar de convalecencia  
 Para quedarse en una residencia estacional o una segunda residencia  Por alguna otra razón

**Figure 2:** Snapshot of the “overcount” question from the 2005 test version of the Census Bureau’s bilingual census form.

All of the other findings in this test were general and cut across the national origin groups. This was probably due to the nature of the questions, most of which were demographic and rather straightforward and simple. When looking at a study with more specialized terminology, such as the American Community Survey, we find that there are more national-origin specific findings.

#### 4. Study 2: The pretesting of the Spanish CATI/CAPI versions of the American Community Survey (ACS)

The second study was the cognitive testing of the Computer Assisted Telephone Interview (CATI) and Computer Assisted Personal Interview (CAPI) versions of a segment of the Census Bureau's American Community Survey (ACS) in Spanish. This research was also done in collaboration between Census Bureau and RTI International researchers. This project is ongoing. Because the ACS is a comparatively lengthy instrument, it was broken down into related segments for testing. This paper discusses the testing of one segment of the questionnaire which took place in 2008 and 2009 (Goerman, et al., forthcoming).

#### **4.1 ACS Project: Survey Content**

The ACS is a survey that contains questions on basic demographics about the people living in a household as well as questions about the characteristics of housing. There are also a number of "person-level" questions that go beyond basic demographics and cover topics such as citizenship, education, health insurance coverage, employment, and so on. The tested segment of the survey discussed in this paper relates to housing questions. We were testing a total of 31 questions on different topics such as the size and type of dwelling a person lives in, kitchen facilities, plumbing, heating and cooling and costs associated with utilities. Because many of the questions are more complex and contain more specialized vocabulary than basic demographic questions, we anticipated that different terms might be used across national origin groups.

The segment of the survey being tested was lengthy and covered many disparate topics. In addition, many of the questions being tested had large numbers of response options. For example, a question asking about the type of dwelling people lived in contained ten response options. Some questions on unrelated topics were similarly lengthy. A question about the type of heating fuel used in the home contained eight response options. It would have been ideal to include respondents who would choose each of the response options for each of these different questions in order to see how people in each situation interpreted the questions and terms but given the scope, time constraints and cost of the project, this was not possible. In order to test the entire Spanish version of the instrument, the survey was broken into six segments and each segment was tested in two iterative rounds of 25-30 interviews. Recruiting respondents who would choose every response option in every question would have expanded the project to the point of being cost-prohibitive.

In the end we chose to recruit respondents based on differences in national origin and other demographic characteristics and ask these people about their interpretation of the terminology in the questions regardless of whether a given response option applied to them. Many useful findings came from this research but unfortunately we did run into some problems as well. For example, we did not have respondents who used each of the types of heating fuel contained in one of the questions. This turned out to be problematic because if a respondent did not understand the term used to describe a type of heating fuel but the person did not use this type of fuel, it was difficult to judge whether revision was in order. It was possible that people who did use that type of fuel would understand the term. In the end, this was a strong limitation to our research.

#### **4.2 ACS Project: Respondent Recruitment Plan**

As in the first study, respondent recruitment was done by a contractor and the recruitment plan was constructed in collaboration between Census Bureau and RTI researchers. As previously mentioned, because this was a large scale project, we were not able to have as many respondents per segment as had been the case with the bilingual questionnaire

project. We tested each segment of the survey with 25-30 respondents in each of two iterative rounds of testing.

Because the ACS is routinely conducted in Spanish in Puerto Rico and a slightly different version of the instrument is used there, an important goal of the project was to test both the “stateside” and Puerto Rico versions of the survey. We also planned to include a small number of English-language interviews so that we could see if issues identified were specific to the translation or cross-cutting problems that also existed in the English version. This would help us to decide when it was necessary to make revisions to just one language version or whether a change was needed in both versions. See Goerman and Caspar (2010) for more information on the method of testing the original wording along with a survey translation.

In order to allow for diversity in the respondent pool, respondents were recruited in two sites, Chicago, Illinois and Raleigh-Durham, North Carolina. Only two sites were chosen for the testing due to budgetary considerations. In addition, the decision was made to recruit Puerto Rican respondents with little or no English speaking ability, who were living stateside, rather than spending additional money to interview people in Puerto Rico. This decision was later deemed to have caused some problems, which are discussed further below.

#### **4.3 ACS Project: Previous Respondent Report**

Because the ACS is an ongoing survey and it contains questions about language use and national origin, we were able to use previous ACS data on Spanish-speaking respondents as a guide to help us choose the types of cognitive interview respondents to include. The survey sponsor created a report detailing the characteristics of respondents who had reported speaking Spanish and speaking English less than “well” in the 2006 ACS. The report gave a great deal of information about the target population that was likely to receive the survey in the future. We had information about the language spoken at home, English-speaking ability, national origin, educational levels, and typical states where people were residing. This information was meant to guide the creation of a recruitment plan and to facilitate choosing the easiest locations to complete the cognitive interviews. The report showed that the vast majority, between 60-70 percent, of ACS Spanish-speaking respondents were of Mexican origin. The next largest groups were from El Salvador, Guatemala, Cuba, Honduras and Puerto Rico but in much smaller numbers.

Because the budget only allowed for a total of 25-30 respondents per round of interviews and we needed to include English speakers as a baseline and Puerto Rican respondents to test the Puerto Rico form, we realized that it would not be possible to include numerous respondents from many other groups. The recruiters were ultimately asked to focus on Mexican and Puerto Rican origin respondents and to include as much variety among the other groups as possible with the remaining cases. This approach was later deemed to have some problems, which are discussed further below.

#### **4.4 ACS Project: National Origin of Respondents**

A total of 27 respondents were included in the first round of testing of the housing segment of the ACS instrument (see Table 2). The largest groups were from Mexico and Puerto Rico, with eight respondents each. There were also five English speakers included. Finally, there were three respondents from Peru and one respondent each from Guatemala, Costa Rica and Honduras.



The data-driven approach to choosing respondents encouraged us to focus on Mexican and Puerto Rican-origin respondents but we later questioned whether the findings based on just one respondent from other groups, such as Guatemala, Costa Rica and Honduras were enough to feel confident that the instrument had been reasonably tested with sufficient diversity of respondents.

**Table 2:** Project 2: ACS Respondent National Origins

<i>Country of origin</i>	<i>Number of respondents</i>
Mexico	8
Puerto Rico	8
Peru	3
Guatemala	1
Costa Rica	1
Honduras	1
U.S. (English)	5
Total:	27

#### **4.5 ACS Project: Results by Nationality**

As in the Decennial bilingual questionnaire project, there were a great many findings in this project that cut across national origin groups. At the same time, this segment of the survey contained more specialized questions about topics such as heating, cooling and kitchen facilities and we saw a correspondingly higher number of findings that differed by national origin groups. This was despite the fact that we had not included a large number of respondents from different national origin groups in the testing.

When we tested questions containing more complex or specialized terminology than had been included in the demographic questions, we found some important differences. For example, a question about the type of dwelling in which a respondent lived read, “Using the Card D, which best describes this building?” In Spanish it read “Usando la Tarjeta D, ¿cuál describe mejor este edificio?” One of the response options read: “Bote, vehículo recreativo, van, etc.,” meaning “Boat, RV, van, etc.” We found that the term “bote” or “boat” was problematic for a number of respondents in that they interpreted it to mean different types of boats, some of which could not be used as a residence.

We found that Mexican-origin respondents in particular interpreted the term “bote” to mean a trash can or other type of container. When these respondents were asked to discuss the response option, we found that the term did not evoke the concept of a boat or ship at all. We deemed this to be a serious enough issue to warrant testing an alternative term. In the end we tested the modified response option “bote o barco,” which includes another term used to describe a boat. We kept the original “bote” wording since some respondents from other countries, particularly Puerto Rico, had expressed the idea that this was a preferred term for them. In the end this new wording tested well and was added to the instrument.

We found a second example of a term that performed differently for different national origin groups in a question about kitchen facilities in the home. The question read “Does this <mobile home/house/apartment/unit> have a stove or range?” The question was originally translated as “¿Tiene <esta casa móvil/esta casa/este apartamento/esta vivienda> una estufa?”

Two of the three Peruvian respondents expressed confusion when they heard this question. They interpreted the term “estufa” or “stove” to be referring to whether they had a space heater. They expressed confusion that this question would appear in a series of questions about kitchen facilities and they responded incorrectly, saying “no” when probing revealed that they did in fact have a stove in their kitchens.

For our second round of testing, we attempted the same type of solution as discussed above for the term “boat.” We added a synonym to the question while maintaining the “estufa” terminology that had worked well for other respondents. The question now read: “¿Tiene <esta casa móvil / esta casa / este apartamento / esta vivienda> una estufa u hornilla?” This can be loosely translated as asking whether the home has a stove or oven. We found that this new term “hornilla” introduced even more confusion for many of the national origin respondents included, so in the end we recommended changing the final question wording to read: “¿Tiene <esta casa móvil / esta casa / este apartamento / esta vivienda> una estufa para cocinar?” or “Does this <mobile home / house / apartment / unit> have a stove *for cooking*?” We were unable to test this modified wording but we believed that it would perform better than the two phrases that we had tested.

An additional example of a national-origin-specific interpretation arose when we tested a question about whether the housing unit had a toilet. The English-language question read: “Does this <mobile home / house / apartment / unit> have a flush toilet?” This was translated as: “¿Tiene <esta casa móvil / esta casa / este apartamento / esta unidad> un inodoro?” In the case of this question we found that respondents from most of the national origin backgrounds we had included did not use or understand the term inodoro (the translation for toilet). We found that Peruvian respondents used the term “water,” Costa Ricans used the term “sanitario” and Mexicans called it a “taza de baño.” We found that the only national origin group that used and understood the term “inodoro” easily was the Puerto Rican group.

Through cognitive testing with different national origin groups we have often found that different respondents express a preference for different terms but still understand a more general term. Often the goal is to find the term that more people understand in a similar way whether or not it is their preferred term. In the case of inodoro, many respondents were unable to answer the survey question without asking for clarification from the interviewer, so we deemed this a serious problem in need of revision.

In our second round of testing, we tested the phrase “inodoro o toilet” but this did not work well. Many people who did not understand that term “inodoro” also did not know the English term “toilet.” In the end, we recommended keeping the inodoro terminology the same on the Puerto Rico version of the instrument and changing the stateside version to read “inodoro o taza de baño.” These were the two terms best understood by the majority of respondents in our sample.

#### **4.6 Contact between National Origin Groups in the U.S.**

One final issue that arose out of our testing was that of contact between groups and acculturation in the U.S. While we had a relatively large number of Puerto Rican respondents in our sample, in order to save money on travel we had interviewed only Puerto Rican respondents living stateside, who did not speak English. First of all, we found that it was difficult and time-consuming to find Puerto Rican respondents on the mainland who had minimal English-speaking ability. When we did find non-English speaking Puerto Rican respondents, it turned out that they were often living in a community of Spanish speakers of other backgrounds, particularly Mexican origin respondents. During the cognitive interviews, they often expressed the idea that “Mexicans use that word.” So, in essence they understood many things the same way that Mexican-origin respondents did.

Our conclusion was that since one of our goals had been to test the instrument used in Puerto Rico with Puerto Rican respondents, it really would have been better to travel to Puerto Rico and find people who were part of the target population. This situation also brings up the issue of acculturation, not only in terms of mainstream U.S. English-speaking culture but also among different national-origin groups living in the U.S. Clearly, Spanish speakers who live in the U.S. are surrounded by many influences and they probably do not speak or understand things in exactly the same way they may have while living in their countries of origin. Of course, this will vary by length of time in the U.S. and the composition of the community in which a given person lives. This is something to keep in mind in Spanish-language cognitive testing in the U.S. It is interesting to consider whether regional representation of Spanish speakers across the U.S. maybe just as important as variation in national origin background.

## **5. Conclusions and Discussion**

These two projects have brought up a number of issues. First of all, we found that when testing basic demographic questions without a lot of specialized vocabulary, there may be relatively few differences across Spanish speakers of different national origin backgrounds. More specialized vocabulary, such as terms used to describe household appliances and activities, has a greater tendency to differ across groups. Often these differences can be unexpected.

Another important thing to note is that while a translator can do research about how terms are used in different countries, it takes testing with real respondents to find out how the terms are actually interpreted. First of all, contact among different groups within the U.S. may have encouraged them to use a common term in a given region. Even when that is not the case, there maybe terms that are commonly understood across groups despite the fact that different people would use a different term in their everyday life.

Based on these two studies alone, it is not possible to give a “magic number” of national origin groups or respondents to include, nor is it possible to say which specific groups should be included in a given cognitive testing project. In these two studies, we did see that certain terms were problematic for Mexican origin respondents or Peruvian respondents. It should also be noted that not every respondent from a given group experienced the same problems, possibly in part due to having contact with Spanish speakers of different national origins in the U.S. For example, if we had only included the one Peruvian respondent who did understand the term “estufa” to mean a “stove” in the ACS study, we may not have realized that this term is more commonly used to refer to a space heater in Peru.

If more people from the other groups had been included or if different groups had been included, we would most likely have found even more issues. Time and resources are always a factor in how many respondents can be included in a given study.

Based on these two projects, we recommend that respondents of different national origin backgrounds be recruited for U.S. Spanish cognitive testing projects to the extent possible. It is also important, where possible, to include at least several respondents from each group in question. An additional goal should be to interview Spanish speakers living in a variety of different regions across the U.S. since people's vocabulary and interpretation of terms is strongly influenced by their social context. Finally, it is important to include respondents who are expected to choose different response options in complex questions to see if their interpretation of questions and terms varies. Whether all of these issues can be addressed in a given study will depend on time and resources. Additional research is needed to determine which factors should be prioritized over others, if any.

Whether or not a researcher can include many groups or even a large number of respondents from each group in a given study, the two studies described here show that major flaws and problems with a translation can be discovered even with small numbers of respondents. Cognitive testing is important whether or not a major undertaking is possible for a given survey.

## **6. Areas for Future Research**

After an examination of these two studies, a number of questions remain to be answered. First of all, how can we choose which specific national origin groups to include in a given study? How many respondents of each national origin group is "enough"? Should repeat rounds of testing include the same or different national origin speakers? To what extent should geographic variability in recruiting be included in survey testing? Should including respondents who vary in their characteristics related to the subject matter being tested be prioritized over national origin diversity?

Empirical research is needed to help answer these questions. In fact, many of these questions could also be applied to English-language cognitive testing in the U.S. Are we interviewing enough English speakers of diverse backgrounds and in enough diverse regions of the country to have confidence that most major issues are being caught? What about non-native English speakers who fill out surveys in English?

Similar to the work done by Blair, et al. (2006), it would be useful to conduct a large number of cognitive interviews with Spanish speakers of different national origins and examine the types of findings made possible with varying numbers of interviews. New research could also compare expert review findings to the findings from various numbers of cognitive interviews in Spanish such as was done in the Fernández et al. (2009b) study in English.

It is inevitable that different projects will have different target audiences and use different types of terminology, so it may be difficult to arrive at one "answer" in terms of the ideal number and types of Spanish-speaking respondents to include in U.S. cognitive testing studies. A calculation of cost and additional time required per interview must always be

included in any decision about numbers and types of respondents. Additional research would help to illuminate many of these issues.

### Acknowledgements

I would like to thank Debbie Griffin, Manuel de la Puente, Herman Alvarado, Matthew Clifton, Leticia Fernández and Mikelyn Meyers of the U.S. Census Bureau and Rachel Caspar, Mandy Sha, Rosanna Quiroz, Georgina McAvinchey Leticia Reed, and Sonia Rodriguez from RTI International for their work on the larger projects that I discuss in this paper. I would also like to thank Terry DeMaio, Debbie Griffin, Yuling Pan and Tommy Wright for their review of this paper.

### References

- Blair, J. and Brick, P.D. (2009). "Current Practices in Cognitive Interviewing." Paper presented at the 64th Annual Conference of the American Association for Public Opinion Research (AAPOR) Hollywood, Florida, May 14-17, 2009 and submitted to the 2009 JSM Proceedings, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association: pp-pp. 5691-5700.
- Brick, P.D. and Blair, J. (2009). "Current Practices in Cognitive Interviewing." Presentation given at the 64th Annual Conference of the American Association for Public Opinion Research (AAPOR) Hollywood, Florida, May 14-17, 2009.
- Blair, J. Conrad, F., Ackermann, A.C., and Claxton, G. (2006). "The Effect of Sample Size on Cognitive Interview Findings" Paper presented at the 61st Annual Conference of the American Association for Public Opinion Research (AAPOR) Montreal, Canada May 18-21, 2006 and submitted to the 2006 JSM Proceedings, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association: pp-pp. 4041-4046.
- Fernández, L., Gerber, E., Clifton, M., Higbie, G., and Meyers, M. (2009a). "Cognitive Pretesting of 2010 Alternative Questionnaire Experiment (AQE) Race and Hispanic Origin Treatment Panels." Report submitted to the U.S. Census Bureau, Statistical Research Division Research Report Series. Online, Available: <http://www.census.gov/srd/papers/pdf/rsm2009-08.pdf>.
- Fernández, L., Goerman, P., Clifton, M. and Meyers, M. (2009b). "Overlap and Gaps between Experts and Respondents: What we learn from both and from each about race/ethnicity questions" Presentation given at the 64th Annual Conference of the American Association for Public Opinion Research (AAPOR) Hollywood, Florida, May 14-17, 2009.
- Goerman, P.L., Quiroz, R., McAvinchey, G., Reed, L., and Rodriguez, S. (Forthcoming). Cognitive Testing of the Spanish American Community Survey CATI/CAPI Instrument, Phase 2, Housing Section. Report to appear in the U.S. Census Bureau Statistical Research Division Report Series.
- Goerman, P.L. and Caspar, R.A. (2010). "A Preferred Approach for the Cognitive Testing of Translated Materials: Testing the Source Version as a Basis for Comparison." *International Journal of Social Research Methodology*. 13: 4. Pp. 303-316, First published on: 16 November 2009 (iFirst).
- Goerman, P., Caspar, R., Sha, M., McAvinchey, G., & Quiroz, R. (2007). Census

- bilingual questionnaire research. Final round 1 report. Study series (survey methodology #2008-1). Statistical Research Division, US Census Bureau. Retrieved May 8, 2009, from <http://www.census.gov/srd/papers/pdf/ssm2008-01.pdf>
- Willis, G. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications, Inc.: Thousand Oaks, CA.