A Longitudinal Perspective on the Effects of Household Language on Data Quality in the American Community Survey

Heather Kitada Smalley

Introduction

Linguistic diversity in the United States is dynamic and reflective of changes in immigration patterns. As the premier statistical organization for the federal government, the US Census Bureau is tasked with collecting data on language use for people living within the United States for diverse applications from sociolinguistic studies to support of legislation. As a result of studies on language use, the Census Bureau plans to offer the 2020 Decennial Census in seven new languages (Arabic, French, Haitian Creole, Japanese, Polish, Portuguese, and Tagalog) to join the already used English, Chinese, Korean, Russian, Spanish, and Vietnamese versions (Prior, 2019; Wang, 2019). With these new translations comes the need to study the effects of language on data quality.

The goal of this chapter is to examine the effects of household language on data quality in the American Community Survey (ACS) via the Public Use Microdata Sample (PUMS) from 2006 through 2017. This research combined multiple fields of study including sociolinguistics, mode effect, statistical modeling for complex surveys, and big data. We present novel data visualization tools that highlight temporal and spatial trends, as well as statistical models that account for the complexities of the sample. All data and analysis methods are freely available, reproducible, and accessible through the American FactFinder and R interfaces.

Background

Although English is the most commonly used and de facto language for governmental purposes, the United States has no official language. Questions regarding languages spoken and the degree of English proficiency have been included in the census in some form since 1890 and have evolved over time to mirror legislative needs (Shin & Kominski, 2010). Section 203 of the Voting Rights Act of 1965 mandated the creation of voting materials in minority languages (Ortman & Shin, 2011). This Act was reinforced in 2000 by Executive Order 13166, which aimed to bridge language barriers to accessing federal programs for individuals with limited English proficiency (LEP; Pan, Leeman, Fond, & Goerman, 2014). These types of legislation necessitated the creation of questions in the census and the ACS to study language trends and distributions.

Forecasting models confirm the increasing trend in the number of non-English-speaking households and the resulting need for linguistic support. The diversity and frequency of languages spoken parallel immigration patterns, reflecting a transition from immigrants speaking predominantly English and Indo-European languages in the late 19th to early 20th centuries (Stevens, 1999) to continually increasing numbers of Spanish and Asian/Pacific Islander language speakers starting in the middle of the 20th century (Bean & Stevens, 2003). Using Census Bureau National Population Projections along with assumptions for population growth and levels of international migration, Ortman and Shin (2011) forecasted language trends using both linear and logistic models. These models suggested that (1) English would continue to be the majority language spoken; (2) Spanish, Portuguese, Russian, Hindi, Chinese, Vietnamese, Tagalog, and Arabic language prevalence would increase, with Spanish continuing to be the most frequently spoken non-English language; and (3) French, Italian, German, Polish, and Korean would decline. In addition to migration patterns, English language acquisition, transmission, and proficiency are often viewed as key indicators of an immigrant's and their descendants' social and cultural assimilation in the United States (Akresh, Massey, & Frank, 2014; Alba, Logan, Lutz, & Stults, 2002; Mouw & Xie, 1999; Ortman & Stevens, 2008; Rumbaut, 1997; Rumbaut, Massey, & Bean, 2006). Yet holistic translation techniques need to be applied to surveys to acquire high-quality data.

Functional equivalence in multilingual survey instruments is of paramount concern (Genkova, 2015; Johnson, 1998). Translations require care at the lexical (wording), syntactic (grammar and naturalness in target language), and pragmatic (sociocultural context and appropriateness) levels (Pan, Sha, Park, & Schoua-Glusberg, 2009). Two primary techniques for survey translation include adoption and adaptation (Harkness, Pennell, & Schoua-Glusberg, 2004; Harkness & Schoua-Glusberg, 1998). The goal of the adoption method is to obtain the most direct translation from the source to the target language alone. It does not allow for differences in cultural interpretation. By contrast, the adaptation method not only involves lexical translation but also incorporates flexibility that allows for changes to achieve a similar stimulus to the desired question construct to ensure the intended meaning is preserved for diverse respondents. The Census Bureau advocates for the committee approach in their Translation Guidelines to aid in the goal of attaining functional equivalence between the source and target language versions (Pan & de la Puente, 2005). However, this goal can be hindered by the fact that the source survey materials are developed in English and are closed to modifications during the translation process (Pan & Fond, 2014). Thus, during the translation process, cognitive interview pretesting is used to assess the efficacy of the instruments (Goerman, Caspar, Sha, McAvinchey, & Quiroz, 2007; Pan, 2004; Pan, Landreth, Park, Hinsdale-Shouse, & Schoua-Glusberg, 2010; Park, Sha, & Pan, 2014; Sha, Pan, & Lazirko, 2013). This method has illuminated conceptual problems in the Spanish-language translation of the ACS for Spanish-speaking respondents (Carrasco, 2003). Furthermore, intrinsic differences in cultural communication norms may affect total survey error (TSE) and ultimately result in biased results.

A TSE frame describes and classifies sources of variability that contribute to differences between a population parameter of interest and the estimated statistic obtained from the survey (Weisberg, 2009). TSE is first partitioned into sampling and nonsampling errors, the latter of which is further broken down into coverage error, nonresponse error, measurement error, and processing error (Groves, 1987, 2004). Of these, nonresponse error and measurement error are especially vulnerable to shifts in cultural perceptions of survey studies. Pan (2003) argued that with “the increase of cultural diversity in survey population, cultural factors, including cultural value systems and social circumstances of personal experience, have been recognized as a strong influence on survey quality and participation” (p. 2). In regard to unit nonresponse, degree of social responsibility, perceived legitimacy of society institutions, and social cohesion can affect a respondent’s participation (Groves & Couper, 2012). These factors can have strong negative effects on participation for immigrants who have little or no experience with surveys in their home country (Pan, 2004). Moreover, answering surveys is inherently a social activity that is governed by social and cultural communication standards, the differences of which are made more apparent when coupled with survey mode effect.

Survey modes of administration vary in degree of cognitive tasks and social interaction (de Leeuw, 1992; Sudman, Bradburn, Schwarz, & Gullickson, 1997), which contribute to systematic differences in response distribution among modes. This phenomenon (known as mode effect) is most commonly found to be significant in comparisons between self-administered and interview-type modes and is predominantly due to the presence of an interviewer (Bowling, 2005; Tourangeau & Smith, 1996). When engaging with an interviewer, respondents may feel compelled to provide perceived socially desirable responses (Walker & Restuccia, 1984). However, social desirability is subjective and related to a respondent's cultural experiences. A key assumption of survey research is that respondents are able to "express their opinions and preferences openly and directly" (Pan, 2003, p. 7). Yet, within the continuum of directness, "Western cultures tend to be direct in expressing their opinions" (Pan, 2003, p. 7), whereas respondents who come from cultures that value indirect communication (e.g., some Asian and African cultures) may become uncomfortable when forced to give direct answers in surveys. Cross-cultural studies on social desirability have used individualism–collectivism, expressiveness, and self-disclosure frameworks to describe respondents' willingness to engage and share personal information with an interviewer who is a stranger (Johnson & Van de Vijver, 2003). In some cases, the effects of social desirability and willingness to respond may result in a modified response or a lack of response to questions that are perceived to be irrelevant. Thus, self-administered modes may appear favorable because they allow respondents to have a sense of anonymity and the security to provide more honest answers. Nonetheless, self-administered modes may come with increased item nonresponse because of complex skip patterns without the aid of an interviewer to help with correct navigation. Hence, mode choice comes with trade-offs between costs, nonresponse, and measurement errors. The balance of these trade-offs has led to the increased popularity of mixed-mode studies (De Leeuw, Hox, Dillman, & European Association of Methodology, 2008; Dillman, Smyth, & Christian, 2014) that aim to balance mode characteristics, resulting in better quality data. This study presents a rare opportunity to examine the effects of both linguistic diversity and survey mode of administration.

Data Application

The ACS and its Public Use Microdata Area (PUMA) data are well used and documented elsewhere (US Census Bureau, 2014). The complex design and methodology components described in this section summarize elements that are relevant to the subsequent analyses presented.

About the ACS

The ACS is an annual product of the US Census Bureau that has provided social, demographic, economic, and housing data at both the individual and household levels since its inception in 2005. Originally these data were collected only on the long form of the decennial census, once every 10 years. The increased frequency of these surveys allows for a better understanding of trends and improved time series data. Estimates from the ACS can be obtained for 1- and 5-year increments; 3-year estimates are also available for years between 2007 and 2013. Laypeople, unaffiliated with the US Census Bureau, may obtain these estimates at the aggregated level through the American FactFinder and may obtain individual questionnaire-level data for people or housing units via the PUMS. The PUMS represents a subsample of responses from the ACS, where a single year of data records is approximately 1 percent of the US population. Individual records have been de-identified to protect personally identifiable information. The PUMS data are often used by researchers and policy makers for analyses because of their granularity and flexibility (Kinney & Karr, 2017).

The smallest obtainable geographic units within the PUMS dataset are the artificial boundaries known as the PUMA. PUMAs are built on census tracts, and counties and have been designed to partition a state such that at least 100,000 people are contained within them. In fact, "nearly every town and county in the country" is represented by a respondent in the PUMS files (US Census Bureau, 2018). The ACS is composed of two separate samples from housing units and group quarters. The Census Bureau's Master Address File is used to construct the sampling frame for the ACS (Bates, 2013). Viable housing units are sampled independently from each of the 3,143 counties using a stratified sampling technique. A two-stage sampling process is used to obtain responses from housing units. In the first stage, blocks are assigned to the sampling strata, sampling rates are calculated, and the sample is selected. The second stage of sampling serves to capture data from those who have not responded to the previous mode of contact by using differential subsampling rates based on expected rates of completed interviews at the tract level, mailability of the address, and harder-to-reach populations (Asiala, 2005; US Census Bureau, 2012).

A mixed-mode methodology and a schedule of multiple contacts are used to improve data quality. The Census Bureau monitors the ACS quality measures, which include sample size, coverage rates, response rates, and item allocation rates, to ensure accuracy and reliability of the data (US Census Bureau, 2002, 2004, 2015). Each ACS iteration comprises 12 monthly independently sampled panels with overlapping cycles of data collection, each of which lasts for 3 months. During these 3 months, three sequential phases of data collection are deployed: mail and Internet, phone, and personal visits. Given that mail and Internet modes are the most cost-effective options, respondents are encouraged to respond through several contacts via these methods. The mail phase consists of up to six postal mailing attempts: prenotice letter, initial mail package, first reminder postcard, replacement mail package (containing an ACS questionnaire), second reminder postcard, and an additional postcard. These multiple mailing attempts, along with a statement regarding one's legal obligation to answer the survey, have been shown to improve response rates (Dillman, 1978). The first three mailings' (prenotice, initial mail package, and first reminder postcard) respondents are given directions on how to log in and respond to the survey via the Internet. It is not until the replacement mail package that respondents are provided with a physical printed copy of the survey and a prepaid envelope in which to return it. Each mailing also includes information about the toll-free telephone questionnaire assistance (TQA), which can be used if a respondent has any questions or needs help completing their survey. If a sampling unit still has not responded to the survey through the mail or Internet and the household has a valid associated phone number, they are eligible to receive the questionnaire over the phone. Computer-assisted telephone interviews (CATIs) are used to automate the data collection process, prevent out-of-range responses, and navigate question skips. Finally, if a unit still fails to respond, they may be selected for the personal visit phase. In this final phase, trained interviewers equipped with laptops are sent into the field to conduct computer-assisted personal interviews (CAPIs). Furthermore, although multiple modes are implemented throughout survey administration to mitigate the weaknesses inherent in each mode, there may be concerns about subsequent mode effects resulting in instability. The estimation and impact of these effects are further complicated by the use of multiple languages within different modes.

The ACS Language Assistance program has been developed to improve accessibility for the ACS and the quality of data obtained from non-English-speaking households. It is standard for all initial mailing materials of the ACS within the United States to be sent in English but also to provide resources for additional support of other languages (Table 3-1). The prenotice letter is accompanied by a multilingual informational brochure with text in English, Spanish, Russian, Chinese, Korean, and Vietnamese. The multilanguage brochure has been shown to significantly improve response rates in experiments for these supported language groups (Joshipura, 2010). The TQA number is also provided so that respondents can receive help directly from an in-language speaker to answer the survey in each of these languages. If a respondent calls the TQA and speaks to an agent during business hours, they may be prompted to answer the questionnaire over the telephone using an automated survey instrument. It should be noted that even though a respondent in this scenario answers the questionnaire over the telephone, they are considered a "mail" response because they were initially part of that group. In addition, if a respondent accesses the ACS online, they have the ability to answer in either English or Spanish. Similarly, if a respondent receives a physical paper survey, the questionnaire is in English, but there is a message on the cover in Spanish that instructs respondents how to receive a paper questionnaire in Spanish. Historically, these requests for Spanish language questionnaires have comprised less than 1 percent of those in the mail phase, which is approximately 200 questionnaires per panel (Fish, 2013). Furthermore, additional support may be requested for Chinese and Korean speakers in the form of language assistance guides. These guides contain full translations of the questionnaires, which are useful for both respondents and interviewers. Bilingual interviewers are hired for the CATI and CAPI phases. Although the CATI and CAPI instruments are in English and Spanish,

Table 3-1. American Community Survey modes of survey administration and languages per mode

Mode	Language of Questionnaire/Interview
Internet (via mail sample)	English or Spanish
Mail	English or Spanish (if requested)
Telephone[a] (via mail sample)	English, Chinese, Korean, Russian, Spanish, and Vietnamese
CATI and CAPI	Instrument in English and Spanish Personal interviews provided in Arabic, Chinese, English, French, German, Greek, Haitian Creole, Italian, Japanese, Korean, Navajo, Polish, Portuguese, Russian, Spanish, Tagalog, Urdu, and Vietnamese[b]

CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

[a] Support provided by calling TQA (telephone questionnaire assistance).

[b] This list depends on the capabilities of bilingual interview staff.

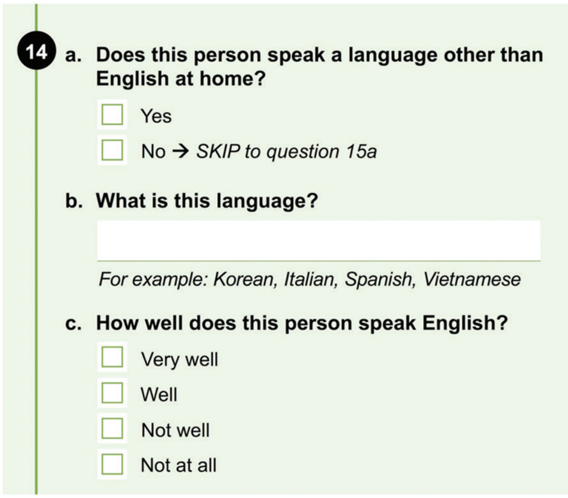
bilingual staff have been able to conduct interviews in more than 30 languages other than English, including Arabic, Chinese, French, German, Greek, Haitian Creole, Italian, Japanese, Korean, Navajo, Polish, Portuguese, Russian, Spanish, Tagalog, Urdu, and Vietnamese. The efficacy of the ACS Language Assistance program, with regard to bridging language barriers, is of considerable interest. In 2005, Griffin (2006) found that bilingual interviewers were well used, interviewing 86 percent of all Spanish-speaking households and 8 percent of Chinese-speaking households who received the CAPI mode.

Language Use and Data Quality

The consistent use of three language questions in the decennial censuses and the ACS has provided useful time series data on the dynamic state of language use in the United States. These questions are part of the person-level sections of the ACS. As shown in Figure 3-1, the first question asks, "Does this person speak a language other than English at home?" with a binary response choice of "yes" or "no." If the respondent answers "yes," the following question asks, "What is this language?" and is accompanied by a one-word open-ended write-in box. Finally, the questionnaire asks, "How well does this person speak English?" with a 4-point Likert response scale with "very well," "well," "not well," and "not at all" as options. Although Singer and Ennis (2002) found that respondents' self-assessment of their proficiency was highly variable, for each housing unit, values were obtained by aggregating the responses from individuals living in the unit.

This chapter explores the effects of three factors on data quality: the household language (HHL), whether the unit is limited English-speaking status (LNGI), and what mode the unit used to respond to the ACS (RESMODE). Although thousands of languages are spoken in the United States, the HHL variable is condensed into five major language groups: English, Spanish, Other Indo-European, Asian and Pacific Island, and a final group that encompasses other languages. Let us further classify the latter four groups as language-other-than-English (LOTE) households. A housing unit may then be categorized as limited English-speaking status, formerly known as "linguistically isolated" until 2010, if no member of the household 14 years old or older (1) speaks only English or (2) speaks a non-English language and speaks English "very well." This distinction is important because it indicates housing units that need additional assistance with English outside of their homes. Because of the varying language support provided for the ACS, we

Figure 3-1. A reproduction of the language questions from the 2017 American Community Survey



14 a. **Does this person speak a language other than English at home?**

- [ ] Yes
- [ ] No → *SKIP to question 15a*

b. **What is this language?**

For example: Korean, Italian, Spanish, Vietnamese

c. **How well does this person speak English?**

- [ ] Very well
- [ ] Well
- [ ] Not well
- [ ] Not at all

Source: US Census Bureau (2019).

would expect to see differential language distributions across the modes of survey administration coded as mail, CATI/CAPI, and Internet. McGovern and Griffin (2003) demonstrated that "linguistically isolated" households are less likely to respond by mail than households speaking English only. This finding is especially true for Spanish linguistically isolated households, which respond at greater rates when interview modes are used. Ideally, the effects of different questionnaires and interview translations used by the ACS could be studied; however, currently the PUMS does not contain a variable that distinguishes in which language the survey was completed. A proxy variable can be created for this by assuming that LOTE households that have LEP will choose to respond in their preferred language when it is available. Although this is not necessarily true, research suggests that a respondent may be more likely to respond if the mode of communication is in their language (Chan & Pan, 2011).

In this setting, we focus on data quality by assessing item nonresponse and response distribution. The occurrence of an item nonresponse in the ACS data record can be deduced by whether a value needed to be imputed to create a complete data record. Two types of imputation methods can be used: assignment and allocation. In the assignment case, the missing value can be derived by taking logical steps from other provided responses within the questionnaire. If logical assignment cannot be used, allocation can be performed, which uses hot-deck or nearest neighbor imputation (Chen & Shao, 2000; Lohr, 2019). Allocation indicators for each item are built into the PUMS data and are used to calculate their respective item allocation rates. Furthermore, examining whether distributions vary across items and across different modes of administration may provide evidence for mode effect that can be detrimental to longitudinal comparisons and even render trends inestimable.

Methods

This chapter uses the 1-year national PUMS data records at the household level from 2006 through 2017 to study trends in language diversity and prevalence over time and space, the effects of non-English-speaking households on data quality, and how they interact with the effect of survey mode of administration. We combined 12 years of PUMS data to expand on the work of McGovern and Griffin (2003), which originally used data from the Census 2000 Supplemental Survey and the 2001 Supplementary Survey to ask (1) which languages have the greatest numbers of linguistically isolated households, (2) how linguistically isolated households were interviewed, and (3) how complete the data collected from linguistically isolated households were. The novelty and contribution of our work lie in the graphical tools and statistical modeling techniques that account for the complexities of the sample to identify and test for trends within these data.

Comma-separated values (CSV) files for each year of PUMS data for households and individuals are approximately 1 and 4 gigabytes, respectively. Thus, combining several years of data quickly exhausts the capabilities of many statistical computing software tools. Because of the size of these data, we used data wrangling and split-apply-combine techniques for big data (Wickham, 2011). In addition, thoughtful consideration was given to constructing data visualization tools to illuminate spatial and temporal trends, particularly for subgroups, in an exploratory data analysis (EDA) (Tufte, 2001). Choropleths were created by joining Census TIGER/Line shapefiles (Walker, 2019) using GEOIDs geographic identifiers at the PUMA level with 2017 PUMS data to visualize the language diversity distribution across the United States. Finally, all statistical modeling was done in R with the survey package to incorporate the complex survey design and weighting structure (Lumley, 2011). A survey design object must first be declared to employ further survey model functionality. This object contains the data as well as the sampling design and weights. Although a household weighting factor variable (WGTP) was contained in the dataset for calculating aggregate statistics, it alone was not sufficient for estimating standard errors. These household weights align demographic characteristics with those determined by the Population Estimates Program of the Census Bureau. Thus, to compute the proper standard errors to use for inference, such as hypothesis testing and confidence interval construction in this complex setting (Binder, 1983), we used a replicate weight methodology. This methodology is akin to resampling techniques, such as the bootstrap, that enable the estimation of variability for a statistic by obtaining multiple samples from a single sample, while still retaining information about the complex survey design (Asparouhov & Muthén, 2010). Eighty columns of replicate weights (WGTP1–WGTP80) were provided with the PUMS data using the successive differences replication (SDR) method (Fay & Train, 1995; Judkins, 1990). The standard error equation for a statistic X using the SDR method is given by

$$SE(X) = \sqrt{C_r \sum_{r=1}^{R} (X_r - X)^2}$$

where there are R replicate estimates of the statistic X_r and $C_r = 4/R$ is a multiplier that scales the variance. For the PUMS data, $R = 80$ and $C_r = 4/80 = 0.05$, which is referred to as the scale in R. Although we used SDR to construct the weights, it is available in neither SAS nor R. However, the jackknife method for variance estimation can be used instead because it is similar and widely available (Dirmyer, 2017; Keathley, Navarro, & Asiala, 2010). This weighting scheme can be coded into R for the PUMS data to define the survey design object (Figure 3-2). Note the use of regular expressions, or regex, to manipulate strings (Friedl, 2006). In this case, regex is used to identify column names for the 80 replicate weights.

Once the survey design object has been defined, the effects are estimated, tested, and modeled with functions built into the survey package, such as svytotal, svyby, svychisq, and svyglm. It should be made clear that the data collected were not from experiments, that is, respondents were not randomly assigned to modes or languages. Therefore, the results of all modeling should be interpreted with caution. We sought to understand patterns inherent in the sample without making causal or broad inferences. For instance, the Rao-Scott adjusted chi-squared test was used to assess the significance of the

Figure 3-2. Segment of R code specifying a survey design object using the survey package

```
svrepdesign(weights = ~WGTP,
 repweights = 'WGTP[1-9]+',
 scale = 4/80,
 rscales = ncol('WGTP[1-9]+'),
 mse = T,
 combined.weights = T,
 type = 'JK1',
 data = PUMS)
```

difference in mode of response across language groups, while accounting for the complex nature of the sample (Rao & Scott, 1981, 1984; Scott, 2007). Although contingency tables are useful for determining associations between two categorical variables, they do not allow for modeling relationships involving multiple covariates simultaneously. Survey generalized linear models (GLMs) are used to model the main effects of mode and household language, as well as their interactions, which are all treated as factors. Interactions in statistical models occur when the effect of one or more variables depends on the level of another variable. In addition, survey GLMs are different from traditional GLMs because they account for weighting and complex sampling in coefficient and standard error estimation. Survey GLMs can be used for both numeric and binary responses, and both numeric and binary approaches can be used to model the allocation indicators or overall allocation scores. When modeling allocation indicator variables, a survey logistic regression technique was used with a quasibinomial family (Morel, 1989). In addition, we computed an allocation rate for each individual household by taking the ratio between the number of imputed values and the number of items with eligible responses (i.e., non-NA values). These rates were used as a response variable and are considered to be independent within and across years.

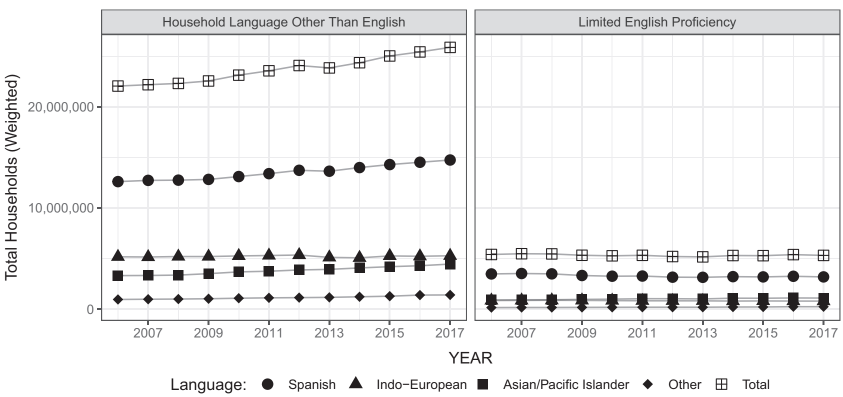
Results

Exploratory Data Analysis

First, using the most currently accessible data, we sought to understand the trends in language diversity and prevalence from 2006 through 2017 and the distribution of non-English speakers across the country. Using weighted values, we estimated that the number of households speaking a LOTE increased 17.3 percent from 2006 through 2017, for a total of 25.8 million households (Figure 3-3, Left). This continual increase is largely driven by a 16.9 percent increase in the number of Spanish-speaking households (to 14.7 million), which is by far the largest LOTE group. Although the Indo-European languages group remains the second largest non-English-speaking group, its membership has stagnated at around 5.2 million households. Conversely, the number of Asian/Pacific Islander language households has shown the greatest increase (34.3 percent, to 4.4 million). In addition, if we consider only households that have LEP status, we see very different trends (Figure 3-3, Right). Overall, the number of LEP households has been relatively stable around 5.3 million. However, both Spanish and Indo-European LEP numbers experienced slight 8 percent decreases to 3.2 million and 0.8 million households, respectively. In contrast, the number of Asian/Pacific Islander LEP households increased by 20 percent, to 1.1 million, surpassing Indo-European as the second most common LOTE-LEP language group. Although we computed 95 percent confidence intervals for all of the total estimates and illustrated significant differences between all estimates, we omitted them from the plot for ease of comparison and trend identification.

Choropleth maps using US PUMAs revealed spatial relationships for both language and LEP in 2017. The most popular language spoken other than

Figure 3-3. Trends in language diversity and prevalence and distribution of non-English speakers across the country, 2006–2017



(Left) Weighted totals of non-English-speaking households from 2006 through 2017 show that the overall increase was driven predominantly by increases in Spanish and Asian/Pacific Islander language speaking households.
(Right) Considering only households with LEP status, the weighted totals of households split by language group appear to be relatively constant from 2006 through 2017.

English was computed for each PUMA (Figure 3-4). It is immediately evident that Spanish is the most common LOTE across PUMAs. There is also a clear pattern of Indo-European languages being most commonly spoken in many PUMAs starting in the Northeast and continuing throughout the Midwest.

Furthermore, the distribution of LEP household counts by PUMA has a strong left skew: LEP households are mainly concentrated along the Southern border and in large metropolitan cities (Figure 3-5). These graphics may help provide an indication of areas in need of linguistic support.

Statistical Modeling

The statistical models we used incorporated weighting and a complex sampling design, which highlighted the significant effect that household language has on both mode choice and data quality via allocation rates. The distribution of the four sequential modes (Internet, mail, CATI, and CAPI), which are associated with varying degrees of language assistance, follows a natural pattern for the language groups. For instance, Chinese speakers are more likely to respond to modes that offer Chinese assistance (Chan & Pan, 2011). We calculated conditional probability distributions for each language group within each year to emphasize these differences (Figure 3-6). In general, these distributions show mail to be the most popular mode of response until 2013, when it was overtaken by the Internet mode; however, this is not the case

Figure 3-4. Choropleth map of Public Use Microdata Areas colored by the most popular language group other than English, using 2017 Public Use Microdata Sample data

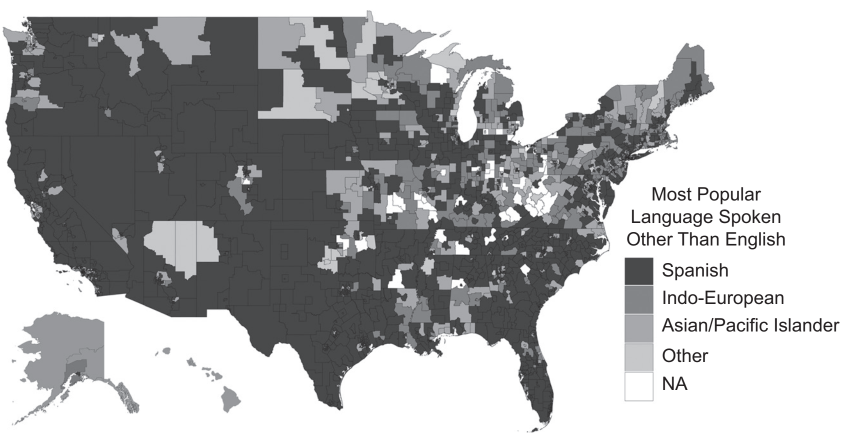
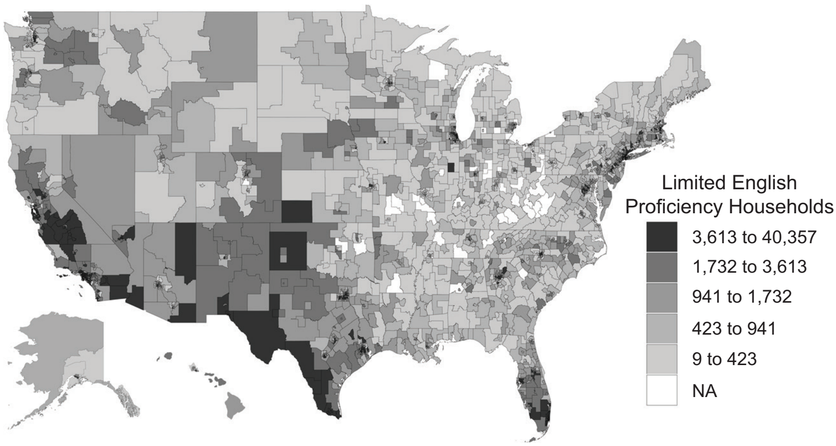


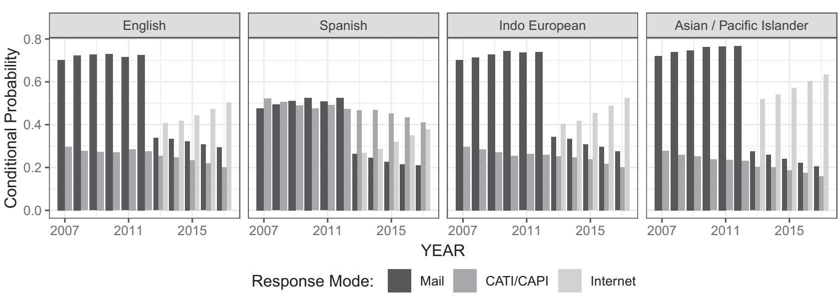
Figure 3-5. Choropleth map of Public Use Microdata Areas shaded by the number of limited English proficiency households, using 2017 Public Use Microdata Sample data



for Spanish-speaking households. For these households, the interview methods (CATI and CAPI) are used at a much higher rate than for the other language groups. This difference drives the significance in the Rao-Scott adjusted chi-squared tests comparing response modes against household language for all years, which all resulted in *p* values less than .0001.

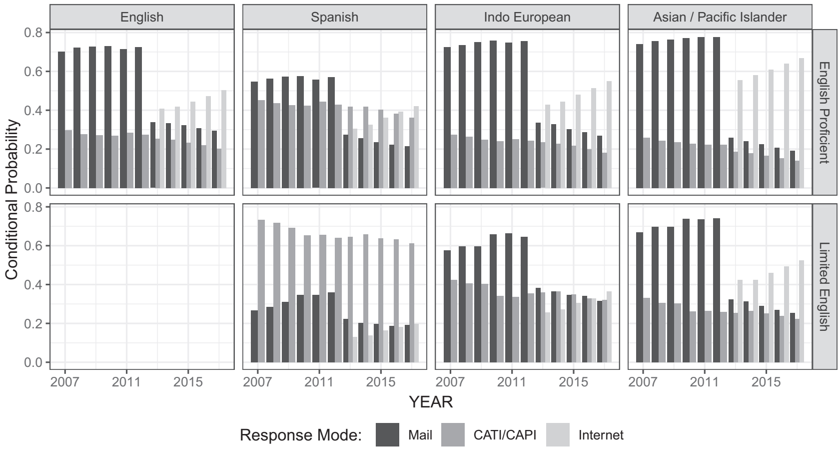
English proficiency was then added to induce additional dimensionality and perspective on conditional mode distribution across language groups (Figure 3-7). The interaction of English proficiency and household language

Figure 3-6. Conditional distributions for response mode across each language group and year have significant differences



CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

Figure 3-7. Conditional distributions on both household language and English proficiency exhibit significantly different modes of response distributions

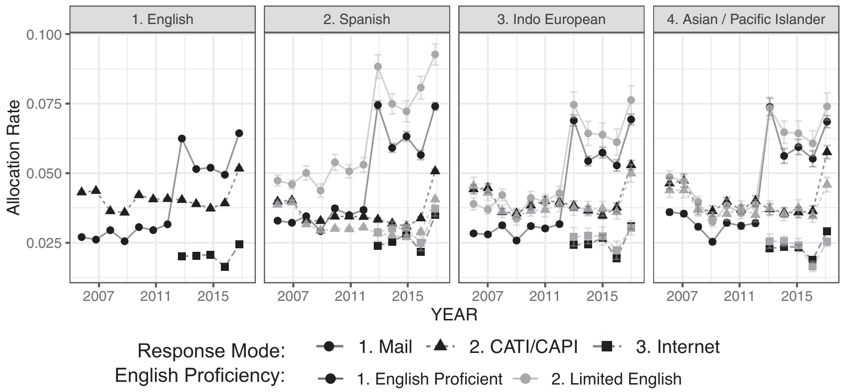


CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

proves to be significant in predicting mode of response. For instance, Spanish LEP households favor interview modes or, rather, do not respond to the self-administered modes of mail and Internet with a clear majority.

Finally, when modeling allocation rates, we found many significant effects for household language and English proficiency levels for both main effects and their interaction. Lower allocation rates are viewed as favorable because they suggest that there is less need for imputation and thus fewer missing data. Estimated marginal means with 95 percent confidence intervals show surprising patterns in allocation rates over time (Figure 3-8). When comparing mail and the CATI/CAPI modes that have existed since the start of the ACS, we see that neither mode is dominant across all language groups and times. Initially, respondents from the mail mode have the lowest average allocation rates, but the lowest average allocation rate shifts to CATI/CAPI after 2012. However, again we observe that Spanish LEP households have the highest allocation rates in the mail mode compared with all other groups and combinations throughout the study from 2006 through 2017. Lower allocation rates for this group may be observed in the CATI/CAPI group because of the aid of bilingual interviewers. In addition, the data first include the Internet response mode in 2013, which clearly has the best allocation rate.

Figure 3-8. Average allocation rates across household language groups split by mode of response and level of English proficiency with bands for 95 percent confidence intervals



CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.
Note: Mode types are distinguished with different symbols, whereas the level of English proficiency is shaded.

This finding could be because the Internet questionnaires have programmed skips to help respondents navigate the survey properly.

English language with the mail mode of response was used as the baseline group for comparison in all models (Tables 3-2 and 3-3). From 2006 through 2012, the CATI/CAPI mode had significantly higher allocation rates, which reversed from 2013 through 2017, when its allocation rates were significantly lower. The effect of Internet mode on allocation rates was significantly lower in all years. In addition, the main effects for all language groups other than English showed significantly higher allocation rates than their English counterparts.

The effects of the interactions between household language and response mode on allocation rates were less consistent, but all estimates had negative point estimates. This result affirms the efforts of the Census Bureau to provide sufficient language assistance, especially by hiring and training bilingual interviewers to better acquire responses. Moreover, although it would have been informative to include an indicator for English proficiency to test for its main effect, two-way interactions, and the three-way interaction, this model did not converge.

Table 3-2. Coefficient estimates for main effects and interactions of household languages and response modes both treated as factors from 2006 through 2012

Coefficients		Effect Estimates						
Response Mode	**Household Language**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
(Intercept)		0.027	0.026	0.029	0.026	0.031	0.030	0.031
CATI/CAPI	—	0.016 ***	0.017 ***	0.007 ***	0.010 ***	0.012 ***	0.011 ***	0.009 **
—	Spanish	0.008 ***	0.008 ***	0.007 ***	0.006 ***	0.009 ***	0.008 ***	0.008 ***
—	Indo-European	0.003 ***	0.003 ***	0.003 ***	0.001 **	0.002 ***	0.002 ***	0.001 ***
—	Asian/Pacific Islander	0.012 ***	0.012 ***	0.004 ***	0.002 ***	0.003 ***	0.003 ***	0.002 ***
—	Other	0.008 ***	0.009 ***	0.007 ***	0.005 ***	0.006 ***	0.007 ***	0.008 ***
CATI/CAPI	Spanish	–0.012 ***	–0.012 ***	–0.011 ***	–0.010 ***	–0.018 ***	–0.015 ***	–0.015 ***
CATI/CAPI	Indo-European	–0.001	–0.002 **	–0.003 ***	–0.002 *	–0.006 ***	–0.003 ***	–0.003 ***
CATI/CAPI	Asian/Pacific Islander	–0.010 ***	–0.010 ***	–0.003 **	–0.002 *	–0.007 ***	–0.006 ***	–0.004 ***
CATI/CAPI	Other	–0.004 *	–0.007 ***	–0.005 **	–0.005 **	–0.010 ***	–0.011 ***	–0.010 ***

CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

Significance codes for *p* values: *** = 0; ** = .001; * = .01; '.' = .05; ' ' = .1.

Note: Standard errors were computed with a jackknife procedure and repeated weights.

Table 3-3. Coefficient estimates for main effects and interactions of household languages and response modes both treated as factors from 2013 through 2017

Coefficients		Effect Estimates				
Response Mode	**Household Language**	**2013**	**2014**	**2015**	**2016**	**2017**
(Intercept)		0.062	0.051	0.052	0.049	0.064
CATI/CAPI	—	−0.022 ***	−0.013 ***	−0.015 ***	−0.010 ***	−0.013 ***
Internet	—	−0.042 ***	−0.031 ***	−0.031 ***	−0.033 ***	−0.040 ***
—	Spanish	0.014 ***	0.010 ***	0.013 ***	0.011 ***	0.013 ***
—	Indo-European	0.007 ***	0.004 ***	0.006 ***	0.005 ***	0.006 ***
—	Asian/Pacific Islander	0.011 ***	0.007 ***	0.009 ***	0.007 ***	0.006 ***
—	Other	0.014 ***	0.012 ***	0.011 ***	0.014 ***	0.016 ***
CATI/CAPI	Spanish	−0.023 ***	−0.019 ***	−0.020 ***	−0.018 ***	−0.017 ***
Internet	Spanish	−0.010 ***	−0.005 ***	−0.005 ***	−0.005 ***	−0.002 *
CATI/CAPI	Indo-European	−0.010 ***	−0.007 ***	−0.008 ***	−0.007 ***	−0.005 ***
Internet	Indo-European	−0.003 **	0.000	0.000	−0.001	0.000
CATI/CAPI	Asian/Pacific Islander	−0.014 ***	−0.011 ***	−0.010 ***	−0.011 ***	−0.003 *
Internet	Asian/Pacific Islander	−0.008 ***	−0.004 **	−0.006 ***	−0.005 ***	−0.002
CATI/CAPI	Other	−0.018 ***	−0.016 ***	−0.016 ***	−0.018 ***	−0.017 ***
Internet	Other	−0.007 **	−0.005 *	−0.003	−0.008 **	−0.005

CATI = computer-assisted telephone interview; CAPI = computer-assisted personal interview.

Significance codes for *p* values: *** = 0; ** = .001; * = .01; ‘.’ = .05; ‘ ’ = .1.

Standard errors were computed with a jackknife procedure and repeated weights. Note the inclusion of the Internet mode.

Discussion and Limitations

The implications of these findings both support the work of the US Census Bureau Language Assistance Program and offer insight into areas on which to focus additional effort. The results of this study reinforce the findings of McGovern and Griffin (2003), while providing a perspective on spatial and temporal trends. Overall, resounding evidence suggests that there is a relationship between household language and mode of response to the ACS. Interactions between these effects then carry forward to influence allocation rates. Yet we should be cautious when communicating inference from these statistical models because of a lack of randomness in mode assignment. As stated in the design and methodology, survey modes are offered to respondents in succession from the Internet mode, to mail, then to CATI, and, finally, CAPI. Brochures are offered in multiple languages that provide non-English-speaking respondents with additional resources, such as a toll-free TQA phone number for assistance in their language or directions for how to obtain language guides or a printed Spanish questionnaire. However, this approach makes the strong assumption that those selected for contact are willing to perform additional steps to receive help. The validity of this assumption is challenged by the clear difference in mode of response distribution across languages. Therefore, the effect of household language and English proficiency may be confounded with mode of response. In addition to the lack of mode assignment, there may be a bias in the estimation of mode effect for the mail group because respondents who answer the questionnaire over the phone by calling the TQA number are included in the mail group and not separated into the CATI group or a separate telephone group. This mixing of interview and self-administer type modes may create distinctly different responses for subgroups within the mail group. Furthermore, care should be given when including additional socioeconomic variables in the model that may be correlated with language groups that commonly represent distinct demographic groups.

Conclusion

The work presented in this chapter provides a quantitative perspective on sociolinguistics in cross-cultural survey studies. As a result of increases and shifts in language diversity in the United States, the work of the Census Bureau has followed suit to provide increased accessibility for minority language speakers in the decennial census and the ACS. However, providing translations is not as simple as a lexical change but rather requires consideration of cultural communication norms and experience. It is these cultural differences that may adversely affect data quality, which is particularly evident across different modes of survey administration resulting in mode effect. The social engagement with an interviewer has been found to both positively and negatively affect data quality measures such as nonresponse and measurement error (Lavrakas, 2008). In the ACS, allocation rates represent the proportion of missing answers for an individual and are used as the metric for item nonresponse.

Using the publicly available microdata for the ACS, we have shown that the sequential aspect of survey mode phases and varying degrees of translation aid across modes led to significant differences in mode distribution across language groups throughout the course of the study from 2006 through 2017. The self-selection of mode and lack of random assignment may cause confounding of language and cultural subgroups with mode. Assuming identifiability, statistical models show that allocation rates are significantly lower for English speakers overall, but the interaction between whether a household speaks English and interview modes tends to improve their allocation rates compared with their language counterparts who chose to respond to the ACS by mail. However, allocation rates for the mail group after 2012 also appear to trend upward unexpectedly.

These data provide a wealth of fruitful opportunities for continued research in this area, for instance, joining the PUMS population and housing data sets to yield an additional depth of information. Comparing allocation rates for housing and personal items across modes and languages would be interesting. Beyond allocation rates, understanding the effects of language and modes on response distributions would shed light on possible sources of measurement error for personal and housing questionnaire items. In addition, a variable could be created to classify the different question types, such as check box, radio button, or fill in, for each item to compare how allocation rates and response distributions vary in these settings. Furthermore, all of these topics can incorporate spatial and temporal features to assess trends.

References

Akresh, I. R., Massey, D. S., & Frank, R. (2014). Beyond English proficiency: Rethinking immigrant integration. *Social Science Research*, *45*, 200–210.

Alba, R., Logan, J., Lutz, A., & Stults, B. (2002). Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demography*, *39*(3), 467–484.

Asiala, M. (2005). American Community Survey research report: Differential sub-sampling in the computer assisted personal interview sample selection in areas of low cooperation rates. DSSD 2005 American Community Survey Documentation Memorandum Series #ACS05-DOC2. Washington, DC: US Census Bureau.

Asparouhov, T., & Muthén, B. (2010). Resampling methods in Mplus for complex survey data. *Structural Equation Modeling*, *14*(4), 535–569.

Bates, L. (2013). Editing the MAF extracts and creating the unit frame universe for the American Community Survey. DSSD 2013 American Community Survey Universe Creation Memorandum Series #ACS13-UC-1. Washington, DC: US Census Bureau.

Bean, F. D., & Stevens, G. (2003). *America's newcomers and the dynamics of diversity*. New York, NY: Russell Sage Foundation.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, *51*(3), 279–292.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, *27*(3), 281–291.

Carrasco, L. (2003). The American Community Survey (ACS) en Español: Using cognitive interviews to test the functional equivalency of questionnaire translations. *Survey Methodology*, *2003*, 17.

Chan, A. Y., & Pan, Y. (2011). The use of cognitive interviewing to explore the effectiveness of advance supplemental materials among five language groups. *Field Methods*, *23*(4), 342–361.

Chen, J., & Shao, J. (2000). Nearest neighbor imputation for survey data. *Journal of Official Statistics*, *16*(2), 113.

De Leeuw, E. D. (1992). *Data quality in mail, telephone and face to face surveys*. Plantage Doklaan, Amsterdam: TT Publikaties.

De Leeuw, E. D., Hox, J. J., Dillman, D. A., & European Association of Methodology. (2008). *International handbook of survey methodology*. New York, NY: Lawrence Erlbaum Associates.

Dillman, D. (1978). *Mail and telephone surveys: The total design method*. New York, NY: John Wiley and Sons.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. New York, NY: John Wiley & Sons.

Dirmyer, R. (2017, April). The truth is out there: Leveraging census data using PROC SURVEYLOGISTIC. Paper presented at SAS Global Forum 2017, Orlando, FL. Retrieved from http://support.sas.com/resources/papers/proceedings17/0802-2017.pdf

Fay, R. E., & Train, G. F. (1995). Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. In *Proceedings of the Section on Government Statistics* (pp. 154–159). Alexandria, VA: American Statistical Association.

Fish, S. (2013, December 17). Percent of Spanish questionnaire requests out of mailout sample. 2013 American Community Survey Office, Special Studies Staff Memorandum Series #SSS13-3. Washington, DC: US Census Bureau.

Friedl, J. E. (2006). *Mastering regular expressions*. Sebastopol, CA: O'Reilly Media.

Genkova, P. (2015). Methodical problems in cross cultural studies: Equivalence—An overview. *Psychology*, *5*(5), 338–346.

Goerman, P., Caspar, R., Sha, M., McAvinchey, G., & Quiroz, R. (2007). *Census bilingual questionnaire research final round 2 report*. Statistical Research Division Report Series No. SSM2007/27. Suitland, MD: US Census Bureau.

Griffin, D. (2006). *Requests for alternative language questionnaires*. American Community Survey Discussion Paper. Washington, DC: US Census Bureau.

Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly*, *51*, S156–S172.

Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). New York, NY: John Wiley & Sons.

Groves, R. M., & Couper, M. P. (2012). *Nonresponse in household interview surveys*. New York, NY: John Wiley & Sons.

Harkness, J., Pennell, B. E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 453–473). Hoboken, NJ: John Wiley & Sons.

Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. *ZUMA-Nachrichten Spezial, 3*, 87–127.

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial, 3*, 2–40. Retrieved from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf

Johnson, T. P., & Van de Vijver, F. J. (2003). Social desirability in cross-cultural research. *Cross-Cultural Survey Methods, 325*, 195–204.

Joshipura, M. (2010, August). Evaluating the effects of a multilingual brochure in the American Community Survey. Paper presented at the *65th Annual American Association for Public Opinion Research Conference*. Chicago, IL.

Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, *6*(3), 223–239.

Keathley, D., Navarro, A., & Asiala, M. E. (2010). An analysis of alternate variance estimation methods for the American Community Survey group quarters sample. In *JSM Proceedings, Survey Research Methods Section* (pp. 1448–1461). Alexandria, VA: American Statistical Association.

Kinney, S. K., & Karr, A. (2017). Public-use vs. restricted-use: An analysis using the American Community Survey. US Census Bureau Center for Economic Studies Paper No. CES-WP-17-12. Washington, DC: Bureau of the Census.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: SAGE.

Lohr, S. L. (2019). *Sampling: Design and analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Lumley, T. (2011). *Complex surveys: A guide to analysis using R* (Vol. 565). New York, NY: John Wiley & Sons.

McGovern, P. D., & Griffin, D. H. (2003). *Quality assessment of data collected from non- English speaking households in the American Community Survey.* Washington, DC: Bureau of the Census.

Morel, J. G. (1989). Logistic regression under complex survey designs. *Survey Methodology*, *15*(2), 203–223.

Mouw, T., & Xie, Y. (1999). Bilingualism and the academic achievement of first-and second- generation Asian Americans: Accommodation with or without assimilation? *American Sociological Review, 64*(2), 232–252.

Ortman, J. M., & Shin, H. B. (2011, August). Language projections: 2010 to 2020. In *Annual meetings of the American Sociological Association* (Vol. 20). Las Vegas, NV: American Sociological Association.

Ortman, J. M., & Stevens, G. (2008, April). Shift happens, but when? Inter- and intragenerational language shift among Hispanic Americans. In *Annual Meetings of the Population Association of America* (pp. 17–19). New Orleans, LA: Population Association of America.

Pan, Y. (2003, November). The role of sociolinguistics in the development and conduct of federal surveys. *Proceedings of the Federal Committee on Statistical Methodology Research Conference* (pp. 1–13). Arlington, VA: Federal Committee on Statistical Methodology.

Pan, Y. (2004). Cognitive interviews in languages other than English: Methodological and research issues. In *JSM Proceedings, Section on Survey Research Methods* (pp. 4859–4865). Alexandria, VA: American Statistical Association.

Pan, Y., & de La Puente, M. (2005). Census Bureau guideline for the translation of data collection instruments and supporting materials: Documentation on how the guideline was developed. *Survey Methodology*. Retrieved from https://www.census.gov/srd/papers/pdf/rsm2005-06.pdf

Pan, Y., & Fond, M. (2014). Evaluating multilingual questionnaires: A sociolinguistic perspective. *Survey Research Methods*, *8*(3), 181–194.

Pan, Y., Landreth, A., Park, H., Hinsdale-Shouse, M., & Schoua-Glusberg, A. (2010). Cognitive interviewing in non-English languages: A cross-cultural perspective. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 91–113). Hoboken, NJ: John Wiley & Sons.

Pan, Y., Leeman, J., Fond, M., & Goerman, P. (2014). Multilingual survey design and fielding: Research perspectives from the US Census Bureau. Survey Methodology, Center for Statistical Research & Methodology Research Report Series (Survey Methodology #2014-01). Washington, DC: US Census Bureau. Retrieved from https://www.census.gov/srd/papers/pdf/RSM2014-01.pdf

Pan, Y., Sha, M. M., Park, H., & Schoua-Glusberg, A. (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: US Census Bureau.

Park, H., Sha, M. M., & Pan, Y. (2014). Investigating validity and effectiveness of cognitive interviewing as a pretesting method for non-English questionnaires: Findings from Korean cognitive interviews. *International Journal of Social Research Methodology*, *17*(6), 643–658.

Prior, R. (2019, April 3). US Census forms to be online in 7 new languages, from Arabic to Tagalog. CNN. Retrieved from www.cnn.com/2019/04/03/us/us-census-languages-trnd/index.html

Rao, J. N., & Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, *76*(374), 221–230.

Rao, J. N., & Scott, A. J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics*, *12*(1), 46–60.

Rumbaut, R. G. (1997). Paradoxes (and orthodoxies) of assimilation. *Sociological Perspectives*, *40*(3), 483–511.

Rumbaut, R. G., Massey, D. S., & Bean, F. D. (2006). Linguistic life expectancies: Immigrant language retention in Southern California. *Population and Development Review*, *32*(3), 447–460.

Scott, A. (2007). Rao-Scott corrections and their impact. In *JSM Proceedings, Section on Survey Research Methods* (pp.3514–3518). Alexandria, VA: American Statistical Association.

Sha, M., Pan, Y., & Lazirko, B. (2013). Adapting and improving methods to manage cognitive pretesting of multilingual survey instruments. *Survey Practice*, *6*(4), 1–8.

Shin, H. B., & Kominski, R. (2010). *Language use in the United States, 2007.* Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Singer, P., & Ennis, S. (2002). *Census 2000 content reinterview survey: Accuracy of data for selected population and housing characteristics as measured by reinterview.* Washington, DC: US Census Bureau, Demographic Statistical Methods Division.

Stevens, G. (1999). A century of US censuses and the language characteristics of immigrants. *Demography*, *36*(3), 387–397.

Sudman, S., Bradburn, N., Schwarz, N., & Gullickson, T. (1997). Thinking about answers: The application of cognitive processes to survey methodology. *PsycCRITIQUES*, *42*(7): 652.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*(2), 275–304.

Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics Press.

US Census Bureau. (2002). *Meeting 21st century demographic data needs—Implementing the American Community Survey: Report 2: Demonstrating survey quality*. Washington, DC: US Department of Commerce.

US Census Bureau. (2004). *Meeting 21st century demographic data needs—Implementing the American Community Survey: Report 7: Comparing quality measures: The American Community Survey's three-year averages and Census 2000's long form sample estimates*. Washington, DC: US Department of Commerce.

US Census Bureau. (2012). *Accuracy of the data*. Retrieved from https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/ACS_Accuracy_of_Data_2012.pdf?#

US Census Bureau. (2014). *American Community Survey design and methodology report*. Retrieved from https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

US Census Bureau. (2015). *Sample size and data quality, American Community Survey*. Retrieved from www.census.gov/acs/www/methodology/sample_size_and_data_quality/

US Census Bureau. (2018). *About PUMS*. Retrieved from www.census.gov/programs-surveys/acs/technical-documentation/pums/about.html

US Census Bureau. (2019). *The American Community Survey questionnaire*. Retrieved December 12, 2019, from https://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2019/quest19.pdf?

Walker, A. H., & Restuccia, J. D. (1984). Obtaining information on patient satisfaction with hospital care: Mail versus telephone. *Health Services Research*, *19*(3), 291–306.

Walker, K. (2019). tigris: Load Census TIGER/Line shapefiles. R package version 0.8.2. Retrieved from https://CRAN.R-project.org/package=tigris

Wang, H. L. (2019). For the first time, US census to collect responses in Arabic among 13 languages. NPR. Retrieved from www.npr.org/2019/03/31/629409884/for-the-first-time-u-s-census-to-collect-responses-in-arabic-among-13-languages

Weisberg, H. F. (2009). *The total survey error approach: A guide to the new science of survey research*. Chicago, IL: University of Chicago Press.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29.