Seeking Clarifications for Problematic Questions: Effects of Interview Language and Respondent Acculturation

Evgenia Kapousouz, Timothy P. Johnson, and Allyson L. Holbrook

Introduction

For scholars in various fields of inquiry, surveys are a critical and widely used tool for the systematic collection of information from respondents, using standardized instruments. The cognitive process by which respondents answer survey questions is generally believed to be a four-stage process. A question is administered to a respondent, and the respondent has to (1) comprehend the question and understand what it is asking him or her to do, (2) retrieve relevant information from memory, (3) form a judgment based on the information retrieved, and then (4) provide an answer using the response format provided (Tourangeau, Rips, & Rasinski, 2009). Error can be introduced into survey estimates through various mechanisms at each stage of this process. Ongena and Dijkstra (2006) suggested that the interview, its structure, and administration may represent underlying determinants of both good and poor survey responses. Consequently, errors associated with survey design, survey questions, interviewers, and respondents must all be considered when evaluating and addressing the quality of survey data. In the following, we review how error can be introduced by each of these sources and examine how survey respondents behave when error is introduced by asking deliberately bad survey questions.

Survey Design as a Source of Error

Face-to-face interviews collect information through direct communication between an interviewer and a respondent. Although face-to-face interviews are costlier compared with self-administered surveys, they continue to have a vital role in data collection in the United States (Garbarski, Schaeffer, & Dykema, 2016). The interview is intended to be an interpersonal event, where the interviewer and respondent in effect participate in scripted ways

(Fowler & Cannell, 1996; Lepkowski, Siu, & Fisher, 2000). However, because they involve people, survey interviews also operate as social and conversational interactions (Tourangeau et al., 2009). The form and quality of the interaction between respondent and interviewer can significantly affect the quality of answers provided by respondents during the interview process. In particular, the relationship of interviewers and respondents must be established very quickly and in an environment that is welcoming and nonthreatening for respondents to feel at ease. Interviewers also need to inspire respondent trust, especially when the interview includes sensitive questions (Fowler & Mangione, 1990).

Standardized interviewing is considered one of the keys to minimizing the measurement error that may be attributed to interviewers (Fowler & Mangione, 1990; Groves, 2004). All respondents should have as similar an experience as possible during an interview to ensure that any differences in the data collected are not due to the interview process, but rather to true differences in survey responses (Fowler & Mangione, 1990). Standardized interviews are expected to follow an established script. Accordingly, when respondents are unsure of the meaning of a question, interviewers can only provide standardized, nondirectional information, such as "Please, answer according to your understanding of the term" (Bell, Fahmy, & Gordon, 2016). Although standardization is preferred, it may at times limit the interaction between the interviewer and respondent in ways that can actually damage data quality (Bell et al., 2016; Fowler & Mangione, 1990).

Supporters of conversational interviewing underscore the belief that unscripted conversation during the interview process may significantly improve data quality. Specifically, Conrad and Schober (1999) reported that conversational interviewing—in which interviewers are trained to deviate from scripted question wording when necessary to achieve survey objectives—prevents possible comprehension problems by establishing shared meaning, particularly for questions concerned with objective phenomena (e.g., number of rooms in living quarters). On the other hand, Ong, Hu, West, and Kirlin (2018) reported that such interaction during the interview may increase interviewer error.

The Question as a Source of Error

The wording of questions may result in unnecessary measurement error if respondents misinterpret or misunderstand the true meaning of the questions (Tourangeau et al., 2009). Some questions may not have a fixed meaning (Groves, 2004), and some may have different meanings in a conversation

compared with a survey context (Schober, 1999). The questions may suffer from various issues such as poor grammar and syntax and semantic issues; being double barreled; using words with ambiguous or vague meaning; using terms or words that respondents are not familiar with; or being worded in such a way as to lead respondents to the "right" answer, while lengthy questions may cause respondents to forget the actual question (Tourangeau et al., 2009). As a result, respondents may in some cases be uncertain of the true meaning of the questions being posed to them.

One strategy researchers have used to examine question-based measurement error is to examine behaviors and responses to questions that deliberately introduce a source of error. For example, some surveys have used questions with unclear terms. Oksenberg, Cannell, and Kalton (1991) reported that this kind of problem in questions can significantly affect responses. Numerous unanticipated differences in question interpretation may also be introduced when respondents speak a different first language and are interviewed in their second language. In addition, respondents from different cultural backgrounds may vary in the likelihood that they will misinterpret any given question (Warnecke et al., 1997).

Differences in respondent culture, often operationalized using race or ethnicity, may lead to variability in understanding of survey questions or increased confusion about them. Researchers sometimes mistakenly assume that measurement is similar between different cultural groups. Perales, Baffour, and Mitrou (2015) asserted that "indigenous cultural imperatives may result in understanding of survey questions and response categories that can be different from other sectors" (p. 3). They suggested that, to improve the quality of survey data, the survey questions should adapt to the needs and culture of respondents. Research has shown ethnic background can affect survey quality because respondents' cultural background may affect the response patterns and interpretation of the questions (Dolnicar & Grün, 2007).

The Interviewer as a Source of Error

Interviewer error is defined as "variation in answers that can be associated with the people who did the interview" (Fowler & Mangione, 1990, p. 24). It can be associated with coverage and nonresponse errors when making initial contact with potential respondents and with measurement error when conducting the interview (West & Blom, 2017). Although interviewers are considered a key factor in promoting response accuracy (Bell et al., 2016),

empirical evidence suggests that slight deviations in question wording do not necessarily affect accuracy negatively (Dykema, Lepkowski, & Blixt, 1997). Interviewer error may vary depending on the question (West, Conrad, Kreuter, & Mittereder, 2017). Experienced interviewers in particular may be better able to manage the interaction with respondents and, in doing so, minimize measurement error (Garbarski et al., 2016).

Researchers have tried to reduce measurement error by matching interviewers with respondents. Webster (1996) reported that matching interviewer and respondent ethnicity increases the response rate and the item response effort. Similarly, Davis and Silver (2003) reported that in telephone surveys, matching the respondents' race with the interviewer's leads to better reporting results and less item nonresponse to sensitive questions; however, answers may not necessarily be improved. Firstly, some respondents may prefer interviewers from a different cultural background (Davis et al., 2013), and ethnicity matching may have a negative effect because respondents are more likely to produce answers acceptable to that cultural group (Fowler & Mangione, 1990). Similarly, Groves (2004) and Weisberg (2005) noted that matching may increase error because respondents tend to report more extreme answers to questions about culture. Research also has shown conflicting results regarding interviewer gender: Fowler and Mangione (1990) reported that gender matching leads to better data quality, whereas Groves and Magilavy (1986) underscored that there is no effect of interviewer gender. Matching respondents and interviewers on gender may have a significant effect only in some countries (Sahgal & Horowitz, 2011) and with certain questions where social desirability may be an issue (Lipps & Lutz, 2017). Although research suggests gender matching may lead to more accurate responses, men appear to be affected more by interviewer gender for some question topics (Catania et al., 1996). Early research has shown that poorly educated respondents are also more likely to respond differently based on interviewer gender (Cannell, Oksenberg, & Converse, 1977; Schuman & Presser, 1977).

The Respondent as a Source of Error

Measurement error can in addition be attributed to respondents, as they are burdened with the responsibility of understanding the intent of each survey question, recalling relevant memories, combining the information to produce a summary judgment, and accurately reporting their answer using the response format provided by the question (Tourangeau et al., 2009). Ideally,

all respondents would attentively go through the steps of comprehension, retrieval, judgment, and response selection and provide high-quality data. In reality, however, factors such as cognitive sophistication and motivation can discourage engagement in optimal behavior, induce compromise, and result in provision of a "merely satisfactory" answer (Krosnick, 1991, p. 215).

Using Respondent Behaviors as an Indicator of Error

Behavior coding—coding behaviors that take place during a survey interview—is one method used to assess respondent cognitive processing of survey questions (Holbrook, Cho, & Johnson, 2006). These can include both respondent and interviewer behaviors; the coding can be done by humans from observation of the interview (although this is rarely done in practice), audio recordings of the interview, or written transcripts of audio recordings, or by computers (e.g., using automated text analysis tools). Behavior coding can also be employed by researchers to identify problematic questions when respondents ask for clarification, as well as problematic interviewer behaviors (Fowler & Cannell, 1996). A clarification may be needed when respondents do not feel they understand precisely what a question is asking and request additional information in an effort to resolve the confusion (Schaeffer & Maynard, 1996). Respondents are more likely to require clarification when they do not comprehend a question (Fowler, 1992) or the question does not relate to their past experiences (Lepkowski et al., 2000). However, behavior coding may not be useful if respondents do not exhibit certain behaviors, such as asking for clarification.

One common respondent behavior captured by behavior coding is requests for clarification. If a question is not clear (e.g., asks respondent to report an opinion about a nonexistent policy or whether they had been diagnosed with a nonexistent illness), respondents should ask for clarification. However, respondents may be motivated to report an opinion without asking for further clarification, even if they had not given any thought to it previously (Schwarz, 1996). Cahalan, Mitchell, Gray, Westat, and Tsapogas (1994) reported that only 2 percent of respondents asked for clarification on well-written questions when participating in the National Survey of Recent College Graduates in 1993. In contrast, Schwarz (1996) asserted that about 30 percent of respondents will answer questions on nonexisting issues. Respondents do not always interrupt the survey process for querying clarification; they may decide not to interrupt or require only one clarification and then answer an unclear question without showing any confusion or hesitation. In some instances, they may simply

engage in satisficing behavior by providing an answer that is acceptable, even if not correct (Krosnick, 1991).

Another respondent behavior that is often captured by behavior coding is providing a qualified response (e.g., I'm not sure, but….) or a "don't know" response. People who provide a qualified response or do not know how to answer are more likely to have misinterpreted the question compared with respondents asking for clarification (Dykema et al., 1997). However, Lepkowski et al. (2000) underscored that respondents answering "don't know" or who request clarification tend to provide less accurate responses. Interruptions seem to be significantly correlated with inaccuracy of the question, such that respondents tend to interrupt the interview process when the question is unclear (Dykema et al., 1997), and respondents usually have comprehension difficulties when the questions are abstract and lengthy (Johnson et al., 2006). Research has shown that less educated respondents are more likely to interrupt the interview process and ask for clarification because they may need more help (Groves, 2004). Respondents also tend to express comprehension problems more often with conversational versus standardized interviewing (Conrad & Schober, 1999), perhaps because they feel less constrained to follow the interview script and freer to express difficulties, similar to the respondent–interview interaction during cognitive interviewing.

Consequently, there seem to be two reasons why respondents might not ask for clarification of a problematic question: (1) respondents may be under the impression that they understand the question adequately, even though that may not be true (Tourangeau et al., 2009), and (2) respondents are aware they have not understood the question but nonetheless provide an answer for self-presentation purposes (i.e., to avoid seeming clueless; Schwarz, 1996).

Social desirability pressures can have a profound impact on some respondents (Tourangeau & Yan, 2007). As mentioned earlier, the fact that some respondents answer questions instead of asking for additional information suggests that their goal may be to avoid appearing uninformed. Hence, they may express an opinion even if they have never thought about or do not have an opinion about a topic (Schuman & Presser, 1980). Although social desirability has been associated mostly with personal characteristics, there is strong evidence that culture can influence perceptions of social desirability (Johnson & Van de Vijver, 2003). In the United States, social desirability may have a larger effect on some minorities because they may in some circumstances regard the interview as a test and fear providing "wrong" answers (Davis & Silver, 2003).

If respondents do not fully comprehend survey questions, we can have no confidence in the quality of the information being reported. Researchers must ensure that all participants have a common understanding of the questions being asked and the response options being provided. In addition to racial and ethnic identification, respondents' first language is an important element of culture and, in addition to interviewer performance, may be a significant indicator of whether respondents ask for clarification on problematic questions. The remainder of this chapter examines how the language in which an interview is conducted and respondent acculturation influence respondent reactions to deliberately problematic questions.

Vygotsky (1962) underscores the importance of language in cognitive development. Understanding respondents' cognitive processing during the interview may be complex but necessary if we want to assess survey quality. People from various countries tend to think differently due to differences in their languages. Language is a significant indicator of question perception because each language has different syntactic properties, grammatical structures, and semantic categories, so language determines what information is retrieved (Peytcheva, 2018). Park and Goerman (2018) reported that respondents in the United States not speaking in English are more likely to face difficulties answering questions adequately, and Perales et al. (2015) asserted more generally that the use of second language in the interview process may hinder the understanding of the survey questions. Researchers have additionally identified issues with applying cognitive interview techniques to certain linguistic groups and in cognitive interviews with non-English speakers; it is hard for non-English speakers to paraphrase, ask for clarification, and think aloud (Park & Goerman, 2018). Hence, the first hypothesis was the following:

> **H_1: Respondents from recent immigrant groups (e.g., Mexican and Korean Americans) who prefer to be interviewed in English will be less likely to ask for clarification when confronted with problematic survey questions than respondents from these groups who prefer to be interviewed in their ethnic native language (e.g., Spanish or Korean).**

Scholars have conceived of and described culture in different ways. For example, Hofstede (1980) defines culture as "the collective programming of the mind which distinguishes the members of one human group from another" (p. 21), whereas others conceive of culture based on how groups interact with or adapt to their physical and social environments (Triandis, 2007). Acculturation is defined as "the process by which immigrants adopt

the attitudes, values, customs, beliefs, and behavior of a new culture" (Abraído-Lanza, White, & Vásquez, 2004, p. 535). Thornton et al. (2010) discussed differences in perceptions of question meaning based on cultural context such that respondents should first comprehend the question to evaluate what information to provide and what the researcher needs to study—in other words, the pragmatic meaning. The pragmatic meaning of a question cannot be reached only by words because the context in which the question is asked is also very important (Uskul, Oyserman, & Schwarz, 2010).

Several studies have examined the effect of culture on response style; however, only a few studies have investigated variance within the same cultural group. One approach is to focus on levels of acculturation to a host culture among immigrants (Davis, Resnicow, & Couper, 2011). Measuring acculturation in bicultural respondents is very complex because the adoption of the second culture influences cognitive development (Tadmor & Tetlock, 2006). In everyday life, we can see how people from different cultural backgrounds may have different understandings in conversation based on the expressions, nuances, and colloquialisms used. For example, in English we say, "My name is," while in Spanish the exact translation of *Me llamo es* is "they call me." Therefore, it is essential that researchers employ terminology that is adequate for all cultures and do not assume that effective communications can be constructed in a similar manner for all populations (Marin, Gamba & Marin, 1992). Johnson (1998) discussed in detail the concept of equivalence in cross-cultural research, concluding there are two main dimensions for equivalence: (1) interpretive equivalence, which refers to "subjective cross-cultural comparability of meaning," and (2) procedural equivalence, which is concerned with "the objective development of comparable survey measures across cultural groups" (p. 38). Furthermore, Bailey and Marsden (1999) found that the interpretation of survey questions depends on the context, regardless of respondent cultural background. However, Qiufen (2014) concluded that even though there are differences in interpretation between and within groups, researchers can find significant similarities in question interpretation from people with similar cultural backgrounds, such that respondents make the same assumptions and follow similar trains of thought.

Acculturated Latino respondents in the United States experience more comprehension issues compared with native-born whites and African Americans, as do less educated respondents (Cho, Holbrook, & Johnson, 2013). Similarly, acculturated Asians tend to provide responses similar to Canadian Caucasian respondents, while less acculturated Asians provided

less emotionally expressive answers when reporting their symptoms (Lai & Linden, 1993). In line with these findings, Johnson, Shavitt, and Holbrook (2010) reported that nonwhite respondents tend to agree and provide more acquiescent responses. We assume respondents who are more likely to agree are less likely to request clarification for questions that are specifically designed to be problematic. In theory, all respondents should request clarification when confronted with a poorly designed question. As discussed earlier, however, some respondents will not request clarification and instead will answer an ambiguous question for several possible reasons, including comprehension errors (e.g., they believe they clearly understood the question, even if that is not possible), social desirability pressures (e.g., they wish to avoid appearing uninformed), or satisficing (e.g., providing an acceptable response rather than an optimal one). Less acculturated respondents may avoid asking for clarification for any of the aforementioned reasons. Therefore, the second hypothesis was the following:

> **H_2: Among respondents from recent immigrant groups (e.g., Mexican and Korean Americans), those who are more acculturated to American culture will be more likely to request clarification when confronted with problematic survey questions.**

Deliberately Problematic Questions

Difficult or problematic questions can be a very useful tool in survey methodology because they can be used to measure question reliability. Respondents often answer questions even when they are not familiar with or know nothing about the policy, event, or object about which the question is asking. Researchers may include problematic questions in the survey instrument that feature nonexisting words or topics to test whether respondents will have an opinion (Bishop, Oldendick, Tuchfarber, & Bennett, 1980). Intuitively, one might expect that all respondents would request clarifications when confronted with deliberately problematic questions. In fact, previous research has demonstrated that some respondents are more likely to provide an opinion for problematic questions than others, with race being a significant indicator, such that African Americans are less likely to query (Bishop et al., 1980; Bishop, Tuchfarber, & Oldendick, 1986). If respondents provide an opinion to problematic questions that they cannot be expected to understand, they may also answer legitimate, nonproblematic questions that they do not fully understand. Therefore, we employed

problematic questions to understand which groups of respondents ask for clarification when needed.

Data and Methods

The survey employed in this study was completed in June 2010 by the Survey Research Laboratory at the University of Illinois at Chicago. The primary goal of the survey was to measure racial and ethnic variability in survey question processing and response behaviors. All respondents were Chicago residents between 18 and 70 years old. Stratified sampling was used for this study; each stratum represented a targeted race and ethnic group. Participants were first contacted via telephone. After being screened for eligibility, individuals were invited to visit the Survey Research Laboratory to participate in face-to-face interviews. All interviews were audio and video recorded. Respondents were given $40 and a parking voucher for participation. The survey included 151 Mexican American and 150 Korean American respondents from whom the data reported here were obtained. Due to the complex procedures used to obtain a sufficient sample of each race and ethnic group, it is not possible to estimate a response rate based on American Association for Public Opinion Research guidelines.

Respondents could choose whether they were interviewed in English or Spanish or Korean, depending on their ethnic group. The goal was to conduct half of the interviews of Mexican Americans in Spanish and half of the interviews of Korean Americans in Korean, with the rest conducted in English. All respondents were matched with an interviewer of the same race because there is evidence that respondents are more likely to provide more accurate information when their race matches that of their interviewer (Davis, Couper, Janz, Caldwell, & Resnicow, 2009). From our sample, 75 Mexican respondents were interviewed in English and 75 in Spanish. Similarly, half of all Korean respondents chose English as their preferred language, and the other half selected Korean. The questionnaire in English was constructed by the principal investigators and reviewed by the Survey Research Laboratory's Questionnaire Review Committee. Special attention was given to Spanish and Korean translations. For each language, one translation expert conducted an initial translation, and then a team of experts reviewed the translation to identify problematic words or phrases and come to a resolution on the final translation.

The analysis focuses on the respondents' reactions to four problematic questions that were purposely included at different points throughout the

questionnaire, as a method of validation for respondents' behaviors and answers. Specifically, we measured whether respondents asked for clarification when they were asked a question about a fictitious topic. The four deliberately problematic questions were (1) "Has a doctor ever told you that you have a hyperactive emissarium?"; (2) "Have you ever tried to cut down on the amount of tracines in your diet?"; (3) "How worried are you about your ordinal health?"; and (4) "Do you favor the Health Opportunity Act of 2006?" Questions 1, 2, and 4 each had two response options ("yes" or "no" or "favor" or "oppose"), whereas Question 3 had a 4-point scale: "very worried," "somewhat worried," "only a little worried," and "not at all worried." All four problematic questions deliberately mentioned nonexistent topics to examine whether respondents would ask for clarification. The respondents could not provide an informed answer if they did not ask for clarification, so we measured whether language and acculturation significantly affected their reactions to those questions.

Subsequent to field work, all interviews were behavior coded. Table 2-1 shows the subset of verbal behavior codes that involved respondent requests for clarification at the question level. These values were summed into a single measure indicating whether respondents asked for any type of clarification after being asked each question. We examined as dependent variables whether respondents requested clarification for each of the four problematic survey questions of interest, and we also created a summed index that represents the total number of questions for which respondents asked for clarification.

Logistic hierarchical models and hierarchical linear modeling[1] were used for the analysis in recognition that the variables of interest were measured at multiple levels, including the respondent level and the interviewer level. The independent variables were grouped based on both interviewer and respondent characteristics. Logistic hierarchical models were used to analyze the dichotomous dependent variables, and hierarchical linear modeling was used for the index of all problematic questions. On the interviewer level, there were three covariates: (1) whether the interviewer is the same gender as the respondent, (2) whether the absolute difference in age between the interviewer and respondent is 5

[1] We used hierarchical models so we could capture effects on two levels: (1) interviewer level and (2) respondent level. The main advantage of hierarchical models is that they are highly accurate because they can isolate the interviewer effect and the respondent, thus we can investigate both within group and between group relationships in a single analysis.

Table 2-1. Explanation of verbal behavior codes

Verbal Behavior Code List	Explanation
Interruption with question	Respondent interrupts initial question reading with a question.
Clarification (Unspecified)	Respondent indicates uncertainty about the question, but it is unclear whether the problem is related to the construct or the context (e.g., "What is the question asking?" or "What?").
Clarification (Construct/ statement)	Respondent makes a statement indicating uncertainty about question meaning (e.g., "I'm not sure what 'depressed' means.").
Clarification (Construct/ question)	Respondent asks for clarification of question meaning (e.g., "What do you mean by 'depressed'?" or "Depressed?").
Clarification (Context)	Respondent indicates an understanding of the meaning of the construct but indicates uncertainty about the question meaning within the context of the question as stated (e.g., "What do you want to know about being depressed?"; "How often do you pay with cash at restaurants?" Response: "Does that include debit cards?").
Clarification (Not enough information)	Respondent indicates that there is not enough information given in the question to answer. (Key phrases include "It depends on the situation."; "It is case by case."; and "I don't have enough information.").
Clarification (Response format)	Respondent indicates uncertainty about the format for responding (e.g., "I'm not sure how to answer that."; "What else, is that all you are offering me?"; or "Are you asking for a percentage?").
Clarification (Response option meaning)	Respondent asks for clarification of a response option meaning (e.g., "What is the meaning of 'sometimes'?").

years or less, and (3) whether the interviewer has previous experience working with the Survey Research Laboratory. The covariates measured at the respondent level were gender, age, ethnicity, education, language, and acculturation. Education was used as a factor variable, and the reference group was high school graduates. The respondents could choose the language in which they would be interviewed. We found that most of the respondents born in the United States chose to be interviewed in English, while those born in Mexico or Korea preferred Spanish or Korean, respectively ($r = .71$). Hence, language preference is strongly associated with country of birth and, consequently, culture.

As for acculturation, the index used consisted of the 17-item Stephenson Multigroup Acculturation Scale (SMAS), which includes questions about

friends, acquaintances, food, current affairs, and history (Stephenson, 2000). Only Mexican and Korean respondents answered these questions because there was no reason for native-born white and African Americans to respond to acculturation questions. The SMAS Cronbach's alpha internal reliability coefficients for Mexicans and Koreans were 0.81 and 0.88, respectively. Table 2-2 shows the independent variables used in the analysis. There are four models that represent each question and a final model in which the dependent variable is the index. All analyses were conducted using the R programming language, using the libraries tidyverse and lme4.

Results

Tables 2-3 through 2-5 provide descriptive statistics for the study variables. We looked at the correlations between the independent variables and determined that none of the models suffered from multicollinearity.

Figure 2-1 shows the percentage of respondents asking for clarification, which varied for each question. A relatively small percentage asked for clarification for each of these items. Specifically, for Question 1 ("Has a doctor ever told you that you have a hyperactive emissarium?"), 17 percent of the respondents asked for clarification. For Question 2 ("Have you ever

Table 2-2. Explanation of independent variables

Variable Name	Explanation
Interviewer Level	
I_worked	Previous interviewer work experience with the Survey Research Laboratory. (0 = no; 1 = yes)
Ad_age	The absolute age difference between the interviewer and the respondent. (range = 0–47)
Samesex	Same sex as respondent. (0 = no; 1 = yes)
Respondent Level	
Female	Respondents' gender. (0 = male; 1 = female)
Age	Respondents' age. (range = 18–70)
Mexican	Respondents' ethnicity. It is a dummy variable. (0 = Koreans; 1 = Mexicans)
Educ	Respondents' education. It is a factor variable, where the base category is high school degree. The other groups are (1) less than high school, (2) some college, (3) four-year college degree, and (4) graduate degree.
Language	The language of the questionnaire (0 = English; 1 = Korean or Mexican)
Acculturation	Higher values indicate greater adjustment to American culture. (range = 47–96)

Table 2-3. Descriptive statistics of continuous independent variables

Variable	Minimum	Mean	Median	Max	*SD*	*n*
Ad_age	0.10	16.38	14.16	47.27	12.21	301
Age	18	39.82	38	69	15.49	301
Acculturation	47	72.43	73	96	9.30	228

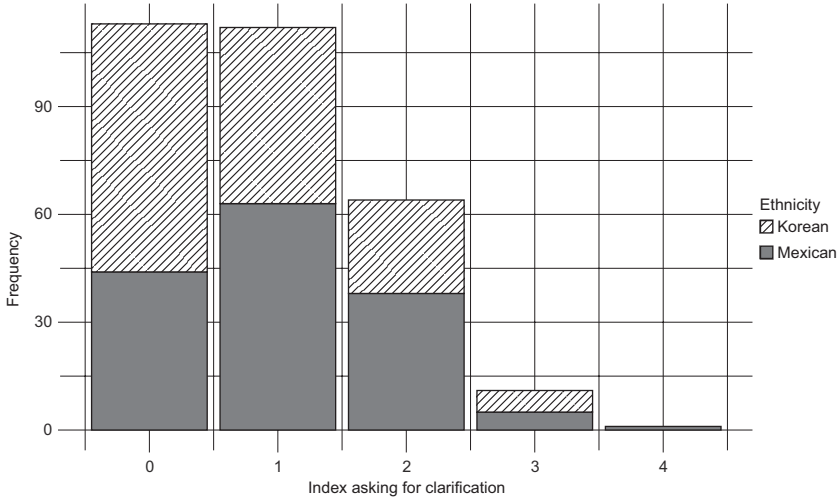
Table 2-4. Distribution of dichotomous independent variables

	No		Yes	
Variable	***n***	**%**	***n***	**%**
I_worked	273	91	28	9
Samesex	149	50	152	50
Female	127	45	157	55
Language	134	47	150	53

Table 2-5. Distribution of the factor variable "education"

	Less than High School		High School Degree		Some College		4-Year College Degree		Graduate Degree	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Education	65	22	55	18	70	23	85	28	26	9

Figure 2-1. Distribution of the index of all problematic questions, by race



tried to cut down on the amount of tracines in your diet?"), 39 percent asked for clarification. For Question 3 ("How worried are you about your ordinal health?"), 25 percent asked for clarification. For Question 4 ("Do you favor the Health Opportunity Act of 2006?"), 12 percent asked for clarification. These findings are similar to those reported earlier by Schwarz (1996) but higher compared with Bishop et al. (1980, 1986). Because of the nature of the questions, it is possible respondents did not consider that a clarification was required. For example, for Question 1, respondents may have answered "no" without asking for clarification because they know that their doctor never told them they had a hyperactive emissarium. This may explain the relatively small percentage that asked for clarification. However, we notice that fewer respondents posed a query for Question 4, for which a clarification was needed to provide an opinion. By looking at the distribution of the dependent variable, it is apparent that a large number of respondents did not ask any clarifications, while very few people asked for clarification on all four questions.

Four logistic hierarchical models are presented in Table 2-6 to examine the variables associated with providing appropriate responses to each problematic survey question. A final model examined associations between the same set of variables and the index measure of the number of problematic questions that respondents answered appropriately. None of the variables at the interviewer level proved to be significant. At the respondent level, gender, age, and acculturation were also not significantly associated with requests for clarification, while education was only significant ($p < .01$) for respondents with some college education (compared to high school graduates) for Question 1. The direction of the relationship is positive such that respondents with some college education tended to ask for clarification more often. Mexican respondents were more likely to ask for clarification ($p < .01$) compared with Koreans only for Question 1: "Has a doctor ever told you that you have a hyperactive emissarium?" Language was the only significant covariate ($p < .01$) for the index of all four problematic questions because respondents interviewed in Korean or Spanish were more likely to ask for clarification.

Given this set of findings, we partially confirmed the first hypothesis (i.e., people interviewed in Spanish or Korean are more likely to provide higher data quality when confronted with problematic questions), and we rejected the second hypothesis because respondents adjusting to American culture were not more likely to ask for clarification.

Table 2-6. Logistic hierarchical and hierarchical linear models examining requests for clarifications to problematic questions

	Question 1	Question 2	Question 3	Question 4	All Questions
(Intercept)	−0.66 (0.35)*	0.28 (0.47)	−0.23 (0.42)	−0.23 (0.42)	−0.62 (0.80)
l_worked	−0.21 (0.10)**	0.16 (0.17)	−0.03 (0.13)	−0.03 (0.13)	−0.08 (0.29)
ad_age	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.02 (0.01)
samesex	−0.02 (0.05)	0.01 (0.07)	−0.03 (0.06)	−0.03 (0.06)	−0.06 (0.12)
female	−0.01 (0.05)	−0.05 (0.07)	−0.09 (0.06)	−0.09 (0.06)	−0.18 (0.12)
age	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.01 (0.01)
educless than high school	0.01 (0.08)	−0.15 (0.11)	−0.16 (0.10)	−0.16 (0.10)	−0.21 (0.19)
educsome college	0.14 (0.08)*	−0.12 (0.11)	−0.07 (0.09)	−0.07 (0.09)	0.00 (0.18)
educfour year college degree	0.06 (0.08)	−0.10 (0.11)	−0.01 (0.09)	−0.01 (0.09)	−0.08 (0.18)
educgraduate degree	0.07 (0.10)	−0.14 (0.14)	0.20 (0.12)	0.20 (0.12)	0.05 (0.23)
Mexican	0.22 (0.06)***	0.13 (0.11)	−0.11 (0.08)	−0.11 (0.08)	0.19 (0.19)
language	0.25 (0.08)***	0.09 (0.11)	0.14 (0.10)	0.14 (0.10)	0.53 (0.19)***
acculturation	0.00 (0.00)	0.00 (0.01)	0.01 (0.00)	0.01 (0.00)	0.01 (0.01)
AIC	265.24	398.24	343.83	343.83	624.99
BIC	316.62	449.61	395.21	395.21	676.36
Log Likelihood	−117.62	−184.12	−156.92	−156.92	−297.50
Num. obs.	227	227	227	227	227
Num. groups: interviewer id	16	16	16	16	16
Var: interviewer id (Intercept)	0.00	0.01	0.00	0.00	0.04
Var: Residual	0.13	0.23	0.19	0.19	0.67

***$p < .01$, **$p < .05$, *$p < .1$

AIC = Akaike information criterion; BIC = Bayesian information criterion.

Note: Models Question 1 through Question 4 employed logistic regression, and All Questions employed hierarchical linear modeling.

Discussion

The purpose of this chapter was to examine interviewer- and respondent-level variables that can predict whether respondents require a query on deliberately problematic questions in a cross-cultural study and to test two hypotheses regarding the effects of language and acculturation. Asking for clarification is considered necessary before providing an opinion on problematic questions

because it shows that respondents carefully listen to the question, try to understand the meaning of it, and are not only trying to satisfy the interviewer. In general, we did not find many significant effects, maybe because relatively few people asked for clarification in the first place, and we cannot easily predict the respondents who will do so. Although there was variability among interviewers, we found that interviewer experience and matching interviewers and respondents in terms of ethnicity were not significant.

After controlling for interviewer-level variables, few characteristics also proved to be significant at the respondent level. When examining the index of all four problematic questions, only language was found to be associated with requests for clarification: those who were interviewed in Korean or Spanish were more likely to ask for question clarifications. This evidence partially supports Hypothesis 1, but language was not significant for all models. However, it has a positive direction consistent with our hypothesis in all but one model. The findings are in line with previous research; Peytcheva (2018) found that the language in which the instrument was administered affected responses. Previous research has shown contradictory results of the effect of education. For the current research, education was significant only for two questions, such that more educated respondents were more likely to ask for clarification. Our findings are in line with some of the previous research (Bishop et al., 1980, 1986; Olson, Smyth, & Ganshert, 2019). Nevertheless, Johnson et al. (2018), contrary to previous research, found that more educated respondents tend to face more comprehensive issues.

Contrary to Hypothesis 2, acculturation did not appear to be associated with the likelihood that respondents would request clarification of problematic questions. Although none were significant, we found negative coefficients in each model, such that less acculturated respondents were more likely to require a query, which is contrary to previous research indicating that nonacculturated Spanish-speaking respondents in the United States are more likely to produce item nonresponse by answering "don't know" (Lee, Keusch, Schwarz, Liu, & Suzer-Gurtekin, 2018). This difference may be explained by the different scales used in each study. Usually, acculturation scales target specific groups (Celenk & Van de Vijver, 2011); however, there are significant differences in acculturation measures even within the same cultural groups (Unger, Ritt-Olson, Wagner, Soto, & Baezconde-Garbanati, 2007). Additionally, in the current study, we did not take into consideration

the birth country of respondents, and we assessed acculturation for all respondents with different cultural backgrounds using the same measure.

In general, the problem was much broader than we initially thought. Few people asked for clarification of problematic questions, a trend that may affect data quality for legitimate questions as well. Specifically, researchers should be concerned about two issues: (1) respondents providing an opinion on issues, events, or policies that they are not familiar with and (2) respondents providing an answer to a question that they could not have completely understood. Researchers should be extremely conscious when designing their instruments because respondents will not always request assistance when confronted with problematic questions. Therefore, they should focus on careful design and pretesting of questionnaires. Well-designed and tested questionnaires are essential for multinational, multiregional, and multicultural respondents (Harkness, Edwards, Hansen, Miller, & Villar, 2010).

Researchers should pretest instruments with each of the cultural groups that will participate in the research. Once a survey is launched, it is very challenging to predict whether respondents will need assistance with some of the questions, and it is usually too late to make substantive changes. This applies to all ethnicities examined in the current study.

However, the limitations to this study require further consideration. One limitation is that there were only four deliberately problematic questions included here. Additional questions, samples, and strategies will be necessary to more thoroughly examine the predictors of respondent requests for clarification. In the current study, we were unable to examine the effect of question characteristics. Another limitation is that, because the research was conducted only with two ethnicities, Korean and Mexican, our findings are likely not generalizable to other cultural groups. Each culture differs from the others, and although Mexican and Korean cultures come from different continents, they have some similarities. Furthermore, the sample size in our study is relatively small because only 301 respondents were available for these analyses. In the future, we plan to expand our research and compare how likely Americans are to provide an opinion in response to problematic questions compared with people from different cultural backgrounds. We also plan to compare responses to these deliberately problematic questions with other survey questions included in this study to investigate the effects of poor question structure on response latencies and further explore cultural similarities and differences in the survey response process.

References

Abraído-Lanza, A. F., White, K., & Vásquez, E. (2004). Immigrant population and health. In N. Anderson (Ed.), *Encyclopedia of health and behavior* (pp. 533–537). Newbury Park, CA: SAGE.

Bailey, S., & Marsden, P. V. (1999). Interpretation and interview context: Examining the General Social Survey name generator using cognitive methods. *Social Networks*, *21*(3), 287–309.

Bell, K., Fahmy, E., & Gordon, D. (2016). Quantitative conversations: The importance of developing rapport in standardized interviewing. *Quality & Quantity*, *50*, 193–212.

Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudo-opinions on public affairs. *Public Opinion Quarterly*, *44*(2), 198–209.

Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, *50*(2), 240–250.

Cahalan, M., Mitchell, S., Gray, L., Westat, A. C., & Tsapogas, J. (1994, August). *Recorded interview behavior coding study: National Survey of Recent College Graduates*. Proceedings of the American Statistical Association, Section on Survey Research Methods, Toronto.

Cannell, C., Oksenberg, L., & Converse, J. (1977). *Experiments in interviewing techniques: Field experiments in health reporting 1971–1977.* Ann Arbor, MI: University of Michigan Survey Research Center, National Center for Health Services Research.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, *60*(3), 345–375.

Celenk, O., & Van de Vijver, F. J. (2011). Assessment of acculturation: Issues and overview of measures. *Online Readings in Psychology and Culture*, *8*(1), 10.

Cho, Y. I., Holbrook, A., & Johnson, T. P. (2013). Acculturation and health survey question comprehension among Latino respondents in the US. *Journal of Immigrant and Minority Health*, *15*(3), 525–532.

Conrad, F. G., & Schober, M. F. (1999). Conversational interviewing and data quality. *Proceedings of the Federal Committee on Statistical Methodology Research Conference, Tuesday B Sessions* (pp. 21–30). Arlington, VA: Federal Committee on Statistical Methodology.

Davis, D. W., & Silver, B. D. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science*, *47*(1), 33–45.

Davis, R. E., Caldwell, C. H., Couper, M. P., Janz, N. K., Alexander, G. L., Greene, S. M., … Resnicow, K. (2013). Ethnic identity, questionnaire content, and the dilemma of race matching in surveys of African Americans by African American interviewers. *Field Methods*, *25*(2), 142–161.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2009). Interviewer effects in public health surveys. *Health Education Research*, *25*(1), 14–26.

Davis, R. E., Resnicow, K., & Couper, M. P. (2011). Survey response styles, acculturation, and culture among a sample of Mexican American adults. *Journal of Cross-Cultural Psychology*, *42*(7), 1219–1236.

Dolnicar, S., & Grün, B. (2007). Cross-cultural differences in survey response patterns. *International Marketing Review*, *24*(2), 127–143.

Dykema, J., Lepkowski, J. M., & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 287–310). New York, NY: John Wiley & Sons.

Fowler, F. J., Jr. (1992). How unclear terms affect survey data. *Public Opinion Quarterly*, *56*, 218–231.

Fowler, F. J., Jr., & Cannell, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco, CA: Jossey-Bass.

Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: SAGE.

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2016). Interviewing practices, conversational practices, and rapport: Responsiveness and engagement in the standardized survey interview. *Sociological Methodology*, *46*(1), 1–38.

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

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, *50*(2), 251–266.

Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R., & Villar, A. (2010). Designing questionnaires for multipopulation research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 191–202). Hoboken, NJ: Wiley.

Hofstede, G. (1980). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: SAGE.

Holbrook, A. L., Cho, Y. I., & Johnson, T. P. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly*, *70*, 565–595.

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial*, *3*, 1–40.

Johnson, T. P., Cho, Y. I., Holbrook, A. L., O'Rourke, D., Warnecke, R. B., & Chavez, N. (2006). Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Annals of Epidemiology*, *16*, 661–668.

Johnson, T. P., Holbrook, A., Cho, Y. I., Shavitt, S., Chavez, N., & Weiner, S. (2018). Examining the comparability of behavior coding across cultures. In T. P. Johnson, B. E. Pennell, I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 271–292). Hoboken, NJ: Wiley & Sons.

Johnson, T. P., Shavitt, S., & Holbrook, A. L. (2010). Culture and response styles in survey research. In D. Matsumoto & F. Van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 130–178). New York, NY: Cambridge University Press.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236.

Lai, J., & Linden, W. (1993). The smile of Asia: Acculturation effects on symptom reporting. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, *25*(2), 303.

Lee, S., Keusch, F., Schwarz, N., Liu, M., & Suzer-Gurtekin, Z. T. (2018). Cross-cultural comparability of response patterns of subjective probability questions. In T. P. Johnson, B.-E. Pennell, I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 457–476). Hoboken, NJ: John Wiley & Sons.

Lepkowski, J. M., Siu, V., & Fisher, J. (2000). Event history analysis of interviewer and respondent survey behavior. In A. Ferligoj & A. Mrvar (Eds.), *Developments in survey methodology* (pp. 3–20). Metodološki Zvezki, 15. Ljubljana, Slovenia: FDV.

Lipps, O., & Lutz, G. (2017). Gender of interviewer effects in a multitopic centralized CATI panel survey. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology (MDA)*, *11*(1), 67–86.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, *23*(4), 498–509.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics*, *7*(3), 349–365.

Olson, K., Smyth, J. D., & Ganshert, A. (2019). The effects of respondent and question characteristics on respondent answering behaviors in telephone interviews. *Journal of Survey Statistics and Methodology*, *7*(2), 275–308.

Ong, A. R., Hu, M., West, B. T., & Kirlin, J. A. (2018). Interviewer effects in food acquisition surveys. *Public Health Nutrition*, *21*(10), 1781–1793.

Ongena, Y. P., & Dijkstra, W. (2006). Methods of behavior coding of survey interviews. *Journal of Official Statistics*, *22*(3), 419–451.

Park, H., & Goerman, P. L. (2018). Setting up the cognitive interview task for non-English speaking participants in the U.S. In T. P. Johnson, B.-E. Pennell, I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 457–476). Hoboken, NJ: John Wiley & Sons.

Perales, F., Baffour, B., & Mitrou, F. (2015). Ethnic differences in the quality of the interview process and implications for survey analysis: The case of indigenous Australians. *PLoS ONE*, *10*(6), 1–20.

Peytcheva, E. (2018). Can the language of survey administration influence respondent's answers? In T. P. Johnson, B.-E. Pennell, I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 457–476). Hoboken, NJ: John Wiley & Sons.

Qiufen, Y. U. (2014). Understanding the impact of culture on interpretation: A relevance theoretic perspective. *Intercultural Communication Studies*, *23*(3).

Sahgal, N., & Horrowitz, J. H. (2011, May 13). Interviewer gender effects in international surveys: An analysis of Muslim publics in six countries. *Comparative research on world suffering, extremism & evangelicalism*. Presented at the meeting of American Association for Public Opinion Research, Phoenix, AZ.

Schaeffer, N. C., & Maynard, D. W. (1996). From paradigm to prototype and back again: Interactive aspects of cognitive processing standardized survey interviews. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicating processes in survey research* (pp. 75–88). San Francisco, CA: Jossey-Bass.

Schober, M. F. (1999). Making sense of questions: An interactional approach. In M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 77–93). Hoboken, NJ: Wiley & Sons.

Schuman, H., & Presser, S. (1977). Attitude measurement and the gun control paradox. *Public Opinion Quarterly*, *41*(4), 427–438.

Schuman, H., & Presser, S. (1980). Public opinion and public ignorance: The fine line between attitudes and nonattitudes. *American Journal of Sociology*, *85*(5), 1214–1225.

Schwarz, N. (1996). Cognition and communication: Judgmental biases, research methods, and the logic of conversation. Mahwah, NJ: Lawrence Erlbaum Associates.

Stephenson, M. (2000). Development and validation of the Stephenson Multigroup Acculturation Scale (SMAS). *Psychological Assessment*, *12*(1), 77.

Tadmor, C. T., & Tetlock, P. E. (2006). Biculturalism: A model of the effects of second-culture exposure on acculturation and integrative complexity. *Journal of Cross-Cultural Psychology*, *37*(2), 173–190.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2009). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883.

Thornton, A., Achen, A., Barber, J. S., Binstock, G., Garrison, W. M., Ghimire, D. J., … Yount, K. (2010). Creating questions and protocols for an international study of ideas about development and family life. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 59–74). Hoboken, NJ: Wiley.

Triandis, H. C. (2007). Culture and psychology: A history of the study of their relationship. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 59–76). New York, NY: Guilford Press.

Unger, J. B., Ritt-Olson, A., Wagner, K., Soto, D., & Baezconde-Garbanati, L. (2007). A comparison of acculturation measures among Hispanic/Latino adolescents. *Journal of Youth and Adolescence*, *36*(4), 555–565.

Uskul, A. K., Oyserman, D., & Schwarz, N. (2010). Cultural emphasis on honor, modesty, or self-enhancement: Implications for the survey-response process. 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. 191–202). Hoboken, NJ: Wiley.

Vygotsky, L.S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Warnecke, R. B., Johnson, T. P., Chavez, N., Sudman, S., O'Rourke, D. P., Lacey, L., & Horm, J. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, *7*(5), 334–342.

Webster, C. (1996). Hispanic and Anglo interviewer and respondent ethnicity and gender: The impact on survey response quality. *Journal of Marketing Research*, *33*(1), 62–72.

Weisberg, H. F. (2005). *The total survey error approach*. Chicago, IL: University of Chicago Press.

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*(2), 175–211.

West, B. T., Conrad, F. G., Kreuter, F., & Mittereder, F. (2017). Nonresponse and measurement error variance among interviewers in standardized and conversational interviewing. *Journal of Survey Statistics and Methodology*, *6*(3), 335–359.