LANGUAGE ABILITY AND MOTIVATION AMONG FOREIGNERS IN SURVEY RESPONDING

BRIAN KLEINER*
OLIVER LIPPS
ELIANE FERREZ

With increasing migration and linguistic diversification in many countries, survey researchers and methodologists should consider whether data provided by individuals with variable levels of command of the survey language are of the same quality. This paper examines the question of whether answers from resident foreign respondents who do not master available survey languages may suffer from problems of comprehension of survey items, especially items that are more complicated in terms of content and/or form. In addition, it addresses the extent to which motivation may affect the response quality of resident foreigners. We analyzed data from two large-scale surveys conducted in Switzerland, a country with three national languages and a burgeoning foreign population, employing a set of dependent measures of response quality, including don't know responses, extreme responding, mid-5 responding, recency effects, and straight-lining. Results show overall poorer response quality among foreigners, and indicate that both reduced language mastery and motivation among foreigners are relevant factors. This is especially true for foreign groups from countries that do not share a common language with those spoken in Switzerland. A general conclusion is that the more distant respondents are culturally and linguistically from the majority mainstream within a country, the more their data may be negatively affected. We found that more complex types of questions do generally lead to poorer response quality, but to a much lesser extent than respondent characteristics, such as nationality, command of the survey language, level of education, and age.

BRIAN KLEINER is head of Data and Research Information Services at FORS, the Swiss Centre of Expertise in the Social Sciences. OLIVER LIPPS is senior researcher at FORS, the Swiss Centre of Expertise in the Social Sciences. ELIANE FERREZ is a researcher at FORS, the Swiss Centre of Expertise in the Social Sciences.
*Address correspondence to Brian Kleiner; E-mail: brian.kleiner@fors.unil.ch.

1. RESPONDENT LANGUAGE ABILITY AND RESPONSE QUALITY IN SURVEYS

This paper examines the question of whether answers from resident foreign respondents who do not master available survey languages may suffer from problems of comprehension of survey items, especially for items that are more complicated in terms of content and/or form. By "content" we mean the concepts of a question or the cognitive tasks required to answer them (e.g., calculating, searching in memory, projecting into the future), while we take "form" to be the linguistic structure of a question (e.g., number of words, clauses, response categories). Given increasing migration and linguistic diversification in many countries, it is important to examine whether the responses provided by individuals who master a survey language to different degrees are of the same quality. In addition, there is the question of the quality of responses collected from resident foreigners who might be less motivated than non-foreigners to respond to surveys. These questions have implications especially for research that aims to assess the living conditions and/or opinions of national minority populations in relation to majorities.

To address these issues, we examined data from two large-scale surveys conducted in Switzerland, comparing the responses of Swiss respondents with two resident foreign groups: (1) those who come from neighboring countries where one of the three Swiss national languages (German, French, Italian) is spoken, and (2) those who come from other countries where none of the three Swiss national languages is spoken. Switzerland is unusually well suited for this purpose, since it is directly surrounded by countries where its own national languages are used (i.e., Germany, Austria, Liechtenstein, France, and Italy). For one of the two surveys, we make use of a variable that indicates respondents' self-assessed best mastered language. A set of measures of response quality allows us to disentangle to some extent the effects of language ability and motivation among the two resident foreign respondent populations.

1.1 Related Research

There is ample evidence in the literature that elderly people, very young people, and people with lower education levels tend to have more problems with question comprehension and/or answering (e.g., Narayan and Krosnick 1996; Fricker, Galesic, Tourangeau, and Yan 2005; Yan and Tourangeau 2008; Couper and Kreuter 2013), as measured by a variety of dependent variables, such as item nonresponse, extreme responding, and response times. In addition, some work has shown that comprehension of survey items is correlated with cultural or

racial/ethnic differences (Warnecke, Johnson, Chavez, Sudman, O'Rourke, et al. 1997; Harkness, Van de Vijver, and Mohler 2003; Holbrook, Cho, and Johnson 2006). Still other research shows that there are cultural styles of responding to survey questions that may create bias in data across groups (Johnson, O'Rourke, Chavez, Sudman, Warnecke, et al. 1997; Johnson, Cho, Holbrook, O'Rourke, Warnecke, et al. 2006b; Johnson, Holbrook, and Cho 2006a), while others have shown that the language of survey responding itself may create a sort of cultural lens for interpreting and responding to questions (Johnson et al. 2006a; Peytcheva 2008).

Moreover, there has been little research on the effects of *language ability* on question comprehension and consequently on response quality for respondents who do not have a strong command of a survey language. By language ability, we mean loosely the set of linguistic skills needed to understand a language and use it appropriately. A distinction should be made between oral skills needed for telephone or face-to-face surveys, and literacy skills needed for mail or web-based surveys. Our focus is on the former. The lack of research in this area may be because it is generally assumed by survey researchers that such people—once they opt to participate in surveys—are sufficiently competent in the available language of survey administration and therefore understand survey questions in the same way as native speakers. One exception is the work of Gray, D'Ardenne, Balarajan, and Uhrig (2011), which showed by way of cognitive interviews that non-native speakers of English in the UK had serious problems in understanding certain questions.

Respondent motivation may also play an important role in response quality. Cannell, Miller, and Oksenberg (1981) showed that the more respondents were motivated to exert the effort needed for response tasks, the better their response quality was. Similarly, Wenemark, Persson, Brage, Svensson, and Kristenson (2011) found that increasing the intrinsic motivation of respondents led to better data quality. Motivation, along with ability, is also a key concept in the satisficing framework (Krosnick 1991), where higher motivation correlates with more optimal responding behavior. Edwards and Cantor (2004) note that respondent motivation is necessary for investing the cognitive effort needed for survey tasks and adequately answering more complex questions.

There is a large body of literature that focuses on question complexity and how this may influence comprehension and responding to questions. In general, research suggests that questions that are complex in content and in form carry a higher cognitive load for respondents, which may result in misinterpretation of questions, satisficing, or breaking off in a survey. With respect to question content, some work indicates that questions that are subjective (Bassili 1996; Yan and Tourangeau 2008); that are abstract or that have vague concepts (Holbrook et al. 2006; Johnson et al. 2006b); that have infrequent words (Lenzner, Kaczmirek, and Lenzner 2009); that have memory demands (Tourangeau, Rips, and Rasinski 2000); that are hypothetical (Lenzner et al. 2009); or that require

calculation or numeric values (Johnson et al. 2006b; Lenzner et al. 2009) are generally more difficult.

With respect to question form, research suggests that questions that are longer (Holbrook et al. 2006; Johnson et al. 2006b); that contain more clauses (Yan and Tourangeau 2008); that are syntactically more complex (Lenzner et al. 2009); or that include negative particles (Saris, Revilla, Krosnick, and Shaeffer 2010) are generally harder for cognitive processing and responding. In addition, number and types of response categories may also affect the cognitive burden of a question (Tourangeau et al. 2000; Yan and Tourangeau 2008), and the placement of a question in a questionnaire may be a factor in question comprehension (Yan and Tourangeau 2008).

Studies have also demonstrated that lower-educated people and older people may have more difficulties in handling the increased processing demands of more complex questions (Krosnick, Narayan, and Smith 1996; Yan and Tourangeau 2008). Along these lines, it may be the case as well that non-native speakers of a survey language have more or less difficulty depending on the relative complexity of survey questions. Again, there is little in the literature that addresses this.

1.2 Assessing Response Quality

To assess whether the responses of respondents who do not master a survey language might be of poorer quality, especially for more complex or difficult types of questions, we refer to Tourangeau et al.'s cognitive model for survey responding (2000), which identifies four major components of survey responding. First, respondents must understand and interpret a question's meaning and pragmatic intent. Second, a respondent must retrieve the relevant information from memory. Third, he/she must form a judgment or estimate based on the understanding of the question and the information retrieved from memory. Finally, he/she must provide an acceptable answer that meets the needs of the survey item by mapping judgments or estimations to available response options.

Each of these processes in answering survey questions will carry a certain cognitive load, and questions will vary in terms of which of these components are the most taxing for respondents. Thus, for example, long, complicated, or abstract questions may create additional burden for comprehension, while questions that require digging into long-term memory will burden retrieval processes. Further, questions that demand calculation, evaluation, or hypothetical thinking will place a burden on the third stage. Comprehension and reporting are especially affected by question properties, whereas retrieval and judgment are based less on the question and more on characteristics of respondents (Yan and Tourangeau 2008).

Following the framework of Tourangeau et al. (2000), one would generally expect more difficulties for questions with higher task difficulty that place a

greater cognitive burden on able and willing respondents with respect to comprehension, retrieval, judgment, and reporting. Further, there should be poorer response quality among respondents who are less able to cope with the processing demands of more difficult questions. If a significant number of respondents who have not mastered a survey language have trouble in processing and responding to questions, then this should be reflected in respondent data for various selected dependent measures of response quality. Using data from several large-scale Swiss representative telephone surveys, we hypothesized that respondents who have not mastered the language of survey administration would exhibit poorer response quality, as measured by don't know responses, extreme responding, mid-5 responding, recency effects, and straight-lining (defined below). Each of these dependent variables has been used in other research as a measure of response quality (e.g., Krosnick 1991; Krosnick 2002; Kaminska, McCutcheon, and Billiet 2010). We also hypothesized that the effects would be most significant for more complex types of questions that place greater cognitive burden on respondents. With respect to motivation, one might expect resident foreign respondents to generally feel less concerned by national surveys on attitudes and living conditions and thus to put less effort into responding. Thus, our third hypothesis was that, beyond questions of language ability, foreign respondents would be less motivated in responding and would therefore exhibit poorer response quality than Swiss respondents.

2. DATA AND ANALYSES

2.1 Approach

Our approach permitted comparisons on selected response quality measures between Swiss respondents, resident respondents from neighboring countries where the Swiss national languages French, German, and Italian are spoken natively, and all other resident foreigners. Foreigners make up more than 20 percent of the population of Switzerland. Besides those from neighboring countries, the largest foreign populations are from ex-Yugoslavian countries, Portugal, Spain, and Turkey. Our division of foreigners into two groups was modeled on previous work that focused on unit nonresponse and representation of national minorities in Swiss general population surveys (Laganà, Elcheroth, Penic, Kleiner, and Fasel 2011; Lipps, Lagana, Pollien, and Gianettoni 2013). Such comparisons, in combination with two variables concerning respondent language mastery and motivation from one of the two surveys, would presumably allow us to disentangle the influence of language ability and the motivational effects of being a foreigner. To assess the effects of question characteristics and difficulty, we coded survey questions along many of the different dimensions that have been documented in the literature (e.g., question length, number of clauses, need for memory and calculation, and so on—see below).

2.2 The Data

In order to test our hypotheses, we employed the 2004 Swiss Household Panel (SHP) and the 2008 Swiss Labor Force Survey (SLFS). These specific surveys were chosen for a variety of reasons. First, both surveys include the necessary variables for adequately distinguishing (1) Swiss respondents from (2) resident non-Swiss respondents from the neighboring countries of Germany, France, Austria, Liechtenstein, and Italy (where one of the three Swiss national languages is spoken), and (3) resident non-Swiss respondents from other non-neighboring countries. Second, the relatively large sample sizes of these surveys allow for sufficient statistical power in comparing these groups and controlling for other respondent characteristics. To examine the effects of language, the SHP includes a variable where respondents indicated the language they "relate to and master best." For bilingual respondents, they were also able to indicate a "second-best" mastered language. This variable was not available for the SLFS. For motivation, the SHP also includes a variable that indicates the likelihood that respondents would participate in a next wave, as evaluated by interviewers (see section 2.4).

Third, both surveys include a wide range of question types that allow for examination of aspects of question complexity in relation to respondent ability (including language ability) and motivation. Also, the diversity of question types in the two surveys allows for analysis of multiple measures of response quality, such as items treated as "don't know," recency effects, and extreme responding. Fourth, in order to control better for different possible mode effects on response quality, both surveys were conducted using the same mode—by telephone.

There are several differences between the two surveys that create some interesting potential for analysis as well. Most notably, the SLFS is a highly factual survey, while the SHP includes many subjective and evaluative kinds of questions. Related to this, because the SHP includes many opinion questions concerning life in Switzerland (e.g., questions about politics), some of these may be more likely to be considered as of little relevance for foreigners, which could influence response quality by reducing motivation. It should be noted that the data used for both the SHP and the SLFS were from a refreshment sample of 3,389 and 14,532 cases, respectively, and so all of these respondents were participating in the surveys for the first time.

Finally, in addition to German, French, and Italian, languages available in both the SHP and SLFS, in the 2008 SLFS respondents had the option of doing the survey in English, Serbo-Croatian, Albanian, Turkish, and Portuguese. We can thus compare those foreigners who participated in their own language with those who did so in a non-native language with regard to the dependent measures of response quality. For the 2004 SHP, the language of the interview was determined by means of a variable "communication language," included as part of the sampling frame from the Swiss Federal Statistical Office, along with the

Table 1. Features of the SHP (2004) and SLFS (2008)

	Swiss Household Panel (2004): Refreshment sample	Swiss Labor Force Survey (2008): Refreshment sample
Composition	3,389 cases: 2,956 Swiss; 221 from neighboring countries; 212 from other countries	14,532 cases: 8,620 Swiss; 2,586 from neighboring countries; 3,326 from other countries
Scope	Swiss residential adult population	Swiss residential adult population
Topics	Health, well-being and attitudes, politics, social networks, economics, education, and labor	Structure of the labor force and employment behavior patterns
Types of questions	Evaluative, factual, demographic	Factual, demographic
Survey type	Longitudinal	Rotating panel
Available survey languages	German, French, Italian	German, French, Italian, Serbo-Croatian, Albanian, English, Portuguese, Turkish
Survey mode	Telephone (CATI)	Telephone (CATI)

addresses of the sample members. Thus, for example, if a respondent in the French-speaking region was known to be a German speaker, then the interview was conducted in German.

For the SLFS, language was determined on the basis of the address and the linguistic region to which it belongs. Sample members were first contacted in one of the three Swiss official languages. If interviewers detected a potential language problem, then they could propose another non-national language to the respondent. For both the SLFS and the SHP, if a switch of language was desired by the respondent, the interview was then transferred to another interviewer who had mastered the requested language, and the new language was recorded. Without exception, all interviews across the two surveys were conducted by interviewers who had mastered the survey language used. It was very rare that there was a change of language *during* an interview for either survey. If this happened, then the survey was redone in the second language from the beginning. Neither in the SLFS nor in the SHP was it possible to have a proxy for language reasons. Table 1 lays out some key elements of the two surveys.

2.3 Question Coding

Following findings from the literature that point to aspects of question complexity, we selected a set of question codings, to the extent that it was feasible

given the questions available in the surveys. Questions were divided into three types: *Demographic* ones referred to stable social characteristics of respondents or related individuals, such as age, sex, level of education, income, and family size. *Factual/behavioral* questions referred to facts, conditions, or behaviors of or around respondents (e.g., "How many people in total do you have under your orders, either directly or indirectly?"). *Evaluative* questions (attitudes/beliefs/expectations) involved some appeal to the judgment of respondents (e.g., "Would you like to change the number of hours that you work each week?").

The questions were also coded according to their response format. A first type was those that required a yes or no answer. Multiple response questions were those with two or more response options, each labeled (not including yes/no questions). Scale questions included a sequence of ordered numbers (at least four) in their response format, with the extreme points labeled and the points in between not labeled. It should be noted that this kind of question, frequent in the SHP, was not used in the SLFS. Open questions had no specific options available to choose from. These could be questions where respondents were asked for a figure or a date, or else where they could give an open-ended response that was then coded by the interviewer (e.g., "Why are you looking for a new job?").

Related to this, the number of response categories (spoken or explicitly made available to respondents by the interviewers) was also coded for each question. The length and the structure of the questions were coded by counting the number of sentences, clauses, and words in them (including those within response options that were spoken out loud by interviewers). This was done for German, French, and Italian separately. We defined sentences as having at least one subject and one verb, and clauses as including at least and at maximum one subject and one conjugated verb. We did not code questions in the non-national languages for number of words, clauses, or sentences, and so analyses for these did not include cases where the survey was not conducted in German, French, or Italian.

Questions were coded for whether they required *counting* or some degree of *calculation* on the part of respondents (e.g., "For how many hours or days per week does a close relative normally look after the children?"), as well as for *hypothetical thinking* about a possible state of affairs (e.g., "Let's assume that you can't find a part-time job. In that case would you be prepared to accept a full-time job?"). Questions were also coded for whether they required searching in long-term *memory* for something that occurred in the past (e.g., "How many times in the last 10 years were you unemployed?"). We distinguished between intensive searching, moderate searching, and little or no searching in long-term memory.

The *placement* of questions in the survey (i.e., the first third, second third, or last third of the questions) was coded, since whether a question is earlier or later in a survey interview could influence response quality, given the effects

of fatigue or declining motivation. In addition, we coded questions as to whether they were eligible for examining recency effects, as well as extreme and mid-5 responding. Questions that could involve *recency effects* were defined as those with four or more labeled response options. *Extreme responding* and *mid-5 responding* were examined for 11-point-scale questions, only possible for the SHP, where only the extreme points were labeled (the 2008 SLFS did not contain any questions with scales). All questions in the SHP with 11-point scales were thus coded as eligible for these analyses. Finally, questions were coded for their likely relevance for foreigners, since this might mediate any resulting differences between the nationality/language groups for the dependent measures of response quality. Foreigners might indeed feel that certain questions do not really apply to them and so may not be motivated to answer (e.g., "Overall, how satisfied are you with the way in which democracy works in our country, if 0 means 'not at all satisfied' and 10 'completely satisfied' "?).

Variables that were constructed or not orally administered during the survey were excluded from analyses. In the end, we coded 395 variables for the SHP and 399 variables for the SLFS. The questions were coded separately by two people. Their codings were then compared, discussed, and resolved if necessary. For the most part, the coding of the questions was a mechanical exercise (e.g., number of sentences, words), but even for the more subjective question characteristics (e.g., memory, question type) there was a high degree of agreement, and all disputed cases were resolved.

2.4 Analyses

The dependent measures of response quality included responses coded as "don't know" (hereafter referred to as "don't knows"), extreme responding, mid-5 responding, recency effects, and straight-lining. The definitions for these measures are shown in figure 1.

For each measure, the level of response quality was defined as the proportion of relevant questions affected by one of the response quality measures for each individual. For don't knows, there were 380 question items from the SHP and 176 from the SLFS used in the analyses. For extreme responding and mid-5 responding, there were 62 question items each used for the SHP analyses. Thirteen questions were used from the SHP in the three straight-lining blocks of items, while 24 questions were used from the SLFS for recency effects. We employed general linear models with the assumption of a binomially distributed dependent variable and a logit link function (Papke and Wooldridge 1996). We estimated robust unbiased standard errors (Cameron and Trivedi 2009).

While both surveys allowed for assessment of don't knows as an indicator of response quality, the other measures adopted differed across the two surveys according to the distribution of question types. For example, in the

Don't know	Proportion of responses coded as "don't know" for individuals across all asked questions. Neither the SHP nor the SLFS had any questions that *explicitly* offered a "don't know" or "no opinion" response option. For both studies, interviewers were trained to probe at least once when a respondent indicated not having or knowing an answer.
Extreme responding	Number of responses that were at the extreme beginning or end points of 11-point scales divided by the total possible number of such cases for applicable questions.
Mid-5 responding	Proportion of responses at the center point of 11-point scales, i.e., coded as 5 (on scales from 0 to 10).
Recency effects	Based on the proportion of responses that selected the last mentioned item from a list of four or more items for all applicable questions.
Straight-lining	*Straight-lining* was defined as the proportion of blocks where responses were identical (i.e., same selected response category) across blocks of questions with the same stem and response format. There were three such relevant blocks from the SHP selected and employed for analyses.

Figure 1. Definitions of Dependent Variables for Response Quality.

SHP there were too few variables that allowed for a reliable examination of recency effects, the risk being that the contents of the response options for a small number of questions could be a strong determinant of selecting a last-spoken option. Also, given that the SHP included many sequences of Likert-type scale questions (unlike the SLFS), two of the dependent variables (extreme and mid-5 responding) were applicable only for the SHP.

The "nationality" variable was constructed as follows. Respondents claiming Swiss nationality or who had lived in Switzerland since at least the age of two were categorized as "Swiss." Otherwise, those resident respondents who claimed nationality from any of the countries of Germany, Austria, Liechtenstein, France, or Italy were treated as "Neighboring foreigners," and all remaining resident respondents were deemed to be "Non-neighbor foreigners." For the SLFS, we further divided the "Non-neighbor foreigner" group into those who conducted the survey in one of the three national languages ($n = 2,025$) versus those who conducted the survey in some other language (presumably their own native language, $n = 1,301$). We expected those who conducted the survey in their own native language to have fewer difficulties of comprehension than those who did so in a non-native language. Other independent variables included level of education, age, and sex. For the analyses themselves, in addition to descriptive statistics (means and standard errors distinguished by nationality), we built regression models using the above-mentioned dependent and independent variables, including variables related to question complexity and relevance.

Finally, two key variables were available in the SHP as potential mediators of the dependent response quality measures. The first was interviewers' evaluation of the level of respondent understanding of survey questions: "good," "fair," or "poor." This serves as a rough indication of the level of mastery of respondents of the survey language. While certainly subjective, we contend that such

assessments by interviewers are fairly accurate and reliable, based on clues given by respondents during the survey interaction, for example, through hesitations, requests for repetition or clarification, and answers that do not correspond to the questions (see Suchman and Jordan 1990; Moore and Maynard 2002).

The second variable was interviewers' assessment of the likelihood of respondents to repeat the panel survey in the next wave: "absolutely," "probably, yes," "maybe," and "no." This variable serves as a proxy for the level of respondent motivation. Again, we believe that the survey interaction produces enough signs of respondent motivation level to allow for dependable judgments by interviewers (e.g., signs of impatience, anxiety, discomfort, interest, enthusiasm). Indeed, the assessments of interviewers were significantly correlated with actual participation in the next wave in 2005. Both of these mediating variables are included in the regression models described below for the SHP for the dependent measures of response quality.

3. RESULTS

Table 2 presents the means and standard errors for the five dependent measures of response quality, by nationality. The results indicate consistently reduced response quality across the different measures for the non-neighbor foreign groups, including significantly higher levels of don't knows, recency effects (SLFS), extreme responses (SHP), mid-5 responses (SHP), and straight-lining (SHP) than the Swiss groups. In the SLFS, the two non-neighbor foreigner groups behaved in much the same way for don't knows and recency effects, with higher levels than both the Swiss and resident neighbors. Notably, the resident neighbors had equivalent or even slightly better response quality than the Swiss for these measures.

3.1 Response Quality, Language Ability, and Motivation

The results shown in table 2 did not control for other respondent characteristics, nor did they give any indication of possible differential effects of language ability and respondent motivation on the dependent measures of response quality. In order to assess the influence of nationality, language ability, and motivation on the various measures, we ran regression models using the SHP 2004 data. The model presented in table 3 controlled for gender, education level, age, whether the survey language had been mastered by the respondent, as well as "respondent understanding" and "likelihood to participate in the next wave."

Findings suggest that resident non-neighbor foreigners had consistently poorer response quality than Swiss respondents across the different measures, controlling for other socio-demographic factors. Results by education and age were consistent with those reported in the literature—more educated respondents had better data quality than less educated respondents, while older

Table 2. Means and Standard Errors for Measures of Response Quality, by Nationality

	SHP 2004	SLFS 2008
Don't know		
Native Swiss	0.010 (0.0001)	0.007 (0.0001)
Neighbors	0.014 (0.001)	0.006 (0.0002)
Non-neighbor foreigners 1 (de_fr_it)	0.023 (0.002)	0.008 (0.0003)
Non-neighbor foreigners 2 (sc_al_en_tu_po)	(n/a)	0.008 (0.0004)
Recency effects	(n/a)	
Native Swiss		0.171 (0.0010)
Neighbors		0.170 (0.0017)
Non-neighbor foreigners 1 (de_fr_it)		0.175 (0.0020)
Non-neighbor foreigners 2 (sc_al_en_tu_po)		0.176 (0.0025)
Extreme responding		(n/a)
Native Swiss	0.345 (0.003)	
Neighbors	0.337 (0.012)	
Non-neighbor foreigners 1 (de_fr_it)	0.391 (0.013)	
Non-neighbor foreigners 2 (sc_al_en_tu_po)	(n/a)	
Mid-5 responding		(n/a)
Native Swiss	0.157 (0.001)	
Neighbors	0.152 (0.005)	
Non-neighbor foreigners 1 (de_fr_it)	0.168 (0.006)	
Non-neighbor foreigners 2 (sc_al_en_tu_po)	(n/a)	
Straight-lining		(n/a)
Native Swiss	0.192 (0.004)	
Neighbors	0.204 (0.015)	
Non-neighbor foreigners 1 (de_fr_it)	0.241 (0.018)	
Non-neighbor foreigners 2 (sc_al_en_tu_po)	(n/a)	
N	3,389	14,532

NOTE.—The first group of non-neighbor foreigners answered the survey in: de = German; fr = French; or it = Italian. The second group answered in: sc = Serbo-Croatian; al = Albanian; en = English; tu = Turkish; or po = Portuguese.

respondents had poorer data quality than younger respondents. For don't knows, the resident neighbor foreigners also had poorer data quality than the Swiss. Notably, non-mastery of the survey language (as first or second-best mastered language) correlated with poorer response quality only for don't knows. The respondent understanding variable was significantly related to poorer response quality for all of the measures except straight-lining. Finally, we find that the motivation variable "likelihood to participate in next wave" had a significant influence for don't knows and mid-5 responding—those less likely to be expected to participate in the next wave exhibited poorer response

Table 3. Measures of Response Quality in the SHP (2004), Controlling for Socio-Demographic Characteristics, Respondent Understanding, and Likelihood to Participate in Next Wave

	Don't knows	Extreme responding	Mid-5 responding	Straight-lining
Nationality (ref: native Swiss)				
Resident neighbor foreigners	0.373**	−0.035	−0.010	0.152
Resident non-neighbor foreigners	0.522**	0.200**	0.082	0.363**
Sex (ref: female)				
Male	−0.435**	−0.070**	−0.033	−0.266**
Education (ref: low)				
Medium	−0.052	−0.114**	0.044	−0.126
High	−0.362**	−0.256**	−0.114**	−0.315**
Age (ref: 18–59)				
Age 60–69	0.130	0.285**	0.074*	0.470**
Age 70+	0.441**	0.420**	0.080*	0.906**
Best mastered language (ref: survey language best mastered language)				
Survey language is 2nd-best mastered language	0.121	0.00786	0.009	0.026
Survey language neither 1st nor 2nd-best mastered language	0.482**	0.0482	−0.087	0.076
Respondent understanding[a]	−0.448**	−0.0734*	−0.068*	−0.102
Likelihood to participate in next wave[a]	−0.150**	−0.00135	−0.056**	−0.057
Constant	−3.238**	−0.400**	−1.39**	−1.029**
Observations	3,270	3,270	3,270	3,268

*0.05 significance level.
**0.01 significance level.
[a]Variable treated as linear in model.

quality for these measures. However, no effects were detected for extreme responding or straight-lining.

Turning to the Swiss Labor Force Survey, table 4 gives results for don't knows and recency effects, controlling for gender, education, and age. It should be noted that no variables were available in the SLFS 2008 that relate to language, understanding, or motivation of respondents. For these data, non-neighbor resident foreigners who completed the survey in a Swiss national language had poorer data quality than the Swiss for the dependent variables, controlling for other factors. Resident neighboring foreigners actually had significantly

Table 4. Dependent Measures of Response Quality in the SLFS (2008), Controlling for Socio-Demographic Characteristics

	Don't knows	Recency effects
Nationality (ref: native Swiss)		
Resident neighbor foreigners	−0.108**	0.043**
Resident non-neighbor foreigners	0.094*	0.125**
Resident non-neighbor foreigners (surveyed in non-national language)	0.037	0.141**
Sex (ref: female)		
Male	−0.287**	−0.100
Education (ref: low)		
Education, medium	−0.234**	0.024
Education, high	−0.184**	−0.018
Age (ref: 18–59)		
Age 60–69	0.071	0.299
Age 70+	−0.157**	0.043**
Constant	−4.714**	−1.655**
Observations	14,532	14,532

*0.05 significance level.
**0.01 significance level.

lower levels of don't knows than the Swiss. Most interestingly, the non-neighbor foreigners who completed the survey in a non-national language (presumably in a mastered language) had rates of don't knows that were more equivalent to the Swiss than to the other foreigners.

3.2 Don't Knows and Question Characteristics

The measure of don't knows is well suited for examining effects of language ability, motivation, and question complexity among foreign groups, since answering "don't know" is an ideal way to minimize effort and save face for questions that are difficult to understand or simply not relevant for respondents. Indeed, don't knows can have both cognitive and motivational causes and meanings (Loosveldt, Pickery, and Billiet 2002), and so this measure, in relation to other specific variables, can help disentangle these different factors. Also, not responding to questions is a direct reflection of comprehension problems (de Leeuw, Hox, and Huisman 2003), and more so than the other measures of response quality. To take an example, while responding to the extremes of scales may indicate a reduction of effort in providing answers, it is not more likely to result from poor question understanding than non-extreme responding. As will be shown in the following progression of results, the characteristics of questions and several special features of the two surveys allow for distinguishing these different drivers of don't knows.

We note here that unlike for analysis of "don't knows," which include all variables from the two surveys coded for question characteristics, the variables included for analysis of the other dependent measures were small subsets of the total available. For example, the variables used to examine extreme responding for the SHP accounted for 16 percent of the total available from that survey. Thus, examination of the effects of question characteristics was limited for these other measures.

In order to evaluate the relative importance of question and respondent characteristics in relation to this binary dependent variable (don't know vs. substantive answer given), we generated a series of regression models with the SHP data taking its "cross-classified" data structure into account (see Fielding and Goldstein 2006), including all relevant variables that were shown to have significant effects on don't knows in independent regression analyses that controlled for age, level of education, and sex of respondent. Question characteristics that did not have significant effects in independent regression analyses included sentences, clauses, and words for the SHP, and sentences, clauses, and hypothetical questions for the SLFS. The data set for this model was extended such that each individual had as many records as relevant question variables, resulting in 652,235 substantive (i.e., excluding non-applicable questions) respondent-question combinations. The data structure was then such that all respondents were clustered within interviewers.

As shown in table 5, higher levels of don't knows among resident non-neighbor foreigners persist when controlling for sex, level of education, and age. It appears that resident neighbor foreigners also had significantly higher rates of don't knows than the Swiss for this survey, controlling for these other characteristics. The model included the additional variable of survey language mastery, with results indicating that language mastery is indeed an explanatory factor for level of don't knows, while foreigner status remains statistically significant.

Finally, with respect to question characteristics, the results are in the expected direction: generally more complex and difficult questions led to higher levels of don't knows. For example, questions requiring heavy retrieval from memory and those placed in the middle or late in interviews induced more don't knows. Non-subjective questions (i.e., demographic, factual) had lower levels compared to subjective questions, and yes/no and three-response-category questions had lower levels than scale questions. On the other hand, questions requiring some (non-heavy) retrieval from long-term memory and calculation had no effect. Notably, questions that were less relevant for foreigners significantly increased the likelihood of don't knows.

There were few significant interactions between question characteristics and the nationality and best-mastered-language variables. This indicates that the different national and linguistic subgroups had similar levels of difficulty with more complicated questions. One exception to this is for questions that were coded as potentially non-relevant for foreigners—not surprisingly, all foreigners had higher levels of don't knows than the Swiss for these questions.

Table 5. Cross-Classified Binary MCMC Estimated Regression Model (Logit Link) Coefficients for Don't Knows in the SHP (2004)

Nationality (ref: native Swiss)	
Resident neighbor foreigners	0.418**
Resident non-neighbor foreigners	0.707**
Sex (ref: female)	
Male	−0.433**
Education (ref: low)	
Education, medium	−0.187**
Education, high	−0.529**
Age (ref: 18–59)	
Age 60–69	0.193**
Age 70+	0.686**
Best mastered language (ref: survey language best mastered language)	
Survey language is 2nd-best mastered language	0.271**
Survey language is neither 1st nor 2nd-best mastered language	0.583**
Relevance of questions for foreigners (ref: relevant questions)	
Non-relevant questions (for foreigners)	1.544**
Retrieval from memory (ref: no/little retrieval from long-term memory)	
Questions requiring some retrieval from long-term memory	−0.236
Questions requiring heavy retrieval from long-term memory	1.863**
Calculation (ref: questions not requiring calculation)	
Questions requiring calculation	0.121
Placement (ref: questions with early placement in survey)	
Questions with middle placement in survey	0.686**
Questions with late placement in survey	0.831**
Response format (ref: scale questions with 5 or more points)	
Yes/no questions	−1.385**
Questions with three response categories	−1.109**
Question type (ref: evaluative questions)	
Demographic questions	−0.451*
Factual questions	−0.763**
Constant	−5.58**
Bayesian DIC statistics	61,462
Variance (interviewer level)	0.25**
Variance (question level)	2.07**

NOTE.—Data are from the SHP 2004 refreshment sample.
*0.05 significance level.
**0.01 significance level.

It should be noted that there were interviewer effects with respect to don't knows, controlling for respondents and questions—interviewers in the SHP accounted for about 4 percent of the variance, whereas respondents and variables accounted for 47 percent each. In logistic models, the variance at the lowest

level is constrained to the area under the logistic curve ($\pi^2/3 \approx 3.29$); see Snijders and Bosker (1999).

Table 6 shows results for the SLFS, again with a regression applying a cross-classified data structure, including question characteristics. Since the interviewers of the SHP 2004 contributed very little to the variance compared to the other levels, we did not include the interviewer level here for the SLFS.

Table 6. Cross-Classified Binary MCMC Estimated Regression Model (Logit Link) Coefficients for Don't Knows in the SLFS (2008)

Nationality (ref: native Swiss)	
Resident neighbor foreigners	0.01
Resident non-neighbor foreigners (survey in G, F, I)	0.26**
Resident non-neighbor foreigners (surveyed in non-national language)	0.25**
Sex (ref: female)	
Male	−0.38**
Education (ref: low)	
Education, medium	−0.57**
Education, high	−0.62**
Age (ref: 18–59)	
Age 60–69	−0.08**
Age 70+	0.23**
Number of words in question (ref: less than 20 words)	
Words (20 or more)	−0.03**
Retrieval from memory (ref: no/little retrieval from long-term memory)	
Questions requiring some retrieval from long-term memory	−0.01
Questions requiring heavy retrieval from long-term memory	−0.40*
Calculation (ref: questions not requiring calculation)	
Questions requiring calculation	1.81**
Placement (ref: questions with early placement in survey)	
Questions with middle placement in survey	1.31**
Questions with late placement in survey	1.52**
Response format (ref: scale questions with 5 or more points)	
Yes/no questions	−0.60**
Questions with three response categories	0.01
Question type (ref: evaluative questions)	
Demographic questions	0.68
Factual questions	−0.05
Constant	−6.26
Bayesian DIC statistics	90,914
Variance (question level)	4.07

NOTE.—Data are from the SLFS 2008 refreshment sample.
*0.05 significance level.
**0.01 significance level.

The results are generally consistent with our expectations, with a few notable exceptions. Most importantly, the non-neighbor foreigners who completed the survey in a language other than German, French, or Italian had higher don't know levels than the Swiss, at about the same level as those who completed the survey in one of the three national languages. We take up this result in the discussion section. Also, results for respondents 60–69 years of age, memory, and number of words were not in the expected direction. Again, as with the SHP, for the SLFS data there were no significant interactions between question characteristics and the nationality variable.

4. DISCUSSION

Our study aimed to examine whether respondent command of the survey language can have a measurable and significant effect on response quality, especially among resident foreigners who participate in surveys in a language that they have not mastered, and for more complex types of questions. It addressed as well the extent to which motivation may also be a factor with respect to the response quality of foreigners. The results indicate that resident foreigners from non-neighboring countries who participated in the two Swiss surveys overall had consistently poorer response quality than Swiss respondents and resident foreign respondents from neighboring countries where German, French, or Italian are spoken. For example, they were more likely than the Swiss in the Swiss Household Panel to have an item coded as don't know, to answer at the extremes of scales, to choose a mid-5 response, and to straight-line. In the Swiss Labor Force Survey, the non-neighbor foreigners had higher levels of don't knows and recency effects than the Swiss and neighbors. Also, in the SHP there were higher levels of don't knows among respondents who did not claim to have a mastery of the survey language, another indication that command of the survey language may affect response quality.

More complex questions did generally lead to poorer response quality, at least as measured by the level of don't knows. Presumably, the increased cognitive burden and level of effort required to answer such questions had a negative effect. However, the lack of significant interactions in the SHP between question characteristics on the one hand and respondent nationality and best mastered language on the other suggest that language ability may not be so much at play.

Moreover, while these findings did show a systematic difference between resident foreigners from neighboring countries (where German, French, and Italian are spoken) and resident foreigners from non-neighboring countries, they do not unambiguously provide an answer to the question of the effects of language ability on responding. First, in the cross-classified regression model for the SLFS (table 6), non-neighbor respondents who completed the survey in their native language (e.g., Serbo-Croatian, Albanian, Portuguese) also had

higher levels of don't knows and recency effects than the Swiss and neighbors. It is conceivable that selection effects could be at play here, in that those who chose to complete the survey in their native language may have been different in their characteristics from those who completed the survey in one of the three national languages. Indeed, non-neighboring foreigners who completed the survey in a non-national language had overall lower levels of education than those who did the survey in one of the three national languages, and this may account to some extent for their poorer response quality.

Second, after controlling for respondent characteristics (see table 3), results from the SHP showed that the foreigners from neighboring countries had significantly higher levels of don't knows than the Swiss. In this case, we do not have language as a recourse for explaining the difference, since these individuals presumably had a mastery of the language of their country of origin. Also, as shown in table 3, mastery of the survey language was relevant for don't knows, but not for the other measures of response quality, controlling for other factors.

Rather, it might be argued that respondent motivation may also have played a role, since this can influence the amount of effort invested in providing answers to questions (Krosnick 1991). These specific findings could be explained in terms of reduced motivation on the part of foreigners. First, as shown in table 3, the variable "likelihood to participate in the next wave," as a proxy for respondent motivation, was significant in accounting for poorer response quality for two of the dependent measures (don't knows and mid-5 responding).

Second, it should be noted that the SHP includes quite a few opinion questions that could be viewed as less relevant for non-Swiss respondents, including those from neighboring countries, which thus might have led to more don't knows for these groups. Indeed, levels of don't knows for these questions among foreigners were relatively high. Further, the SLFS, with no questions coded as irrelevant for foreigners, showed *lower* levels of don't knows among neighbor respondents compared to the Swiss, and there were no significant differences between these groups for recency effects. While the SLFS did not include any "irrelevant" questions, it is possible that a generally reduced level of motivation could explain to some extent the higher levels of don't knows and recency effects for the non-neighbor foreigners.

In sum, our data and analyses allowed us to identify both language ability and motivation as explanatory factors with respect to poorer response quality among different foreign populations, at least for certain measures. It appears that language ability may be relevant in explaining elevated levels of don't knows among resident foreign groups who are less likely to have mastered the survey language. Motivation may contribute to some extent to poorer response quality in terms of don't knows and mid-5 responding.

While we were unable to determine clearly the distinctive effects of these two factors for the different foreign groups, we can confirm from our study at least that not all foreigners are alike in how they respond to surveys—for some

foreign groups there may be reduced response quality for certain measures, due to language ability or lower motivation, while for others there may be fewer problems or none at all. Most likely, as a general rule, those groups that are more distant culturally and linguistically will be more likely to provide responses of poorer quality.

To put things more into perspective, our findings did not indicate any severe problems among the responses of resident foreign respondents for the two surveys that we examined—the differences were rather small. Nonetheless, they do indicate in a consistent way the potential for measurement error among foreign populations. We believe that more study is needed to assess the relative effects of language ability and motivation on response quality among resident foreign survey populations, as well as the extent to which response quality may be reduced among foreign populations for different types of surveys and survey topics.

It should be noted that motivational factors may also account for observed differences in response quality across education and age groups, in that survey questions are generally geared, unconsciously or not, toward a "normal" respondent, that is, one most similar to survey designers themselves (highly educated, middle-aged, etc.). With respect to question complexity, there may be finer methods to identify specific question characteristics that tend to increase cognitive and task burden and that reveal empirical effects on response quality for different groups.

With respect to the limitations of this work, we can note several. First, the measures of language ability that we employed were not ideal—the two variables used were based on respondents' own self-assessments and the assessments of interviewers. Second, the variable "likelihood to participate in the next wave" is only an indirect proxy for respondent motivation, although it should be noted that there was a highly significant correlation between the assessments of interviewers and the actual participation of respondents in the next wave. In any event, respondent behaviors like high don't know rates or frequent straight-lining may have influenced interviewers' judgments about whether respondents were likely to take part in the next wave. Future work that examines the role of language ability and motivation on response quality should rely on more direct measures. Third, our results are based on two national telephone surveys from Switzerland, and so may not be representative. Further study could determine whether these findings hold for surveys done with other modes, and in other countries.

References

Bassili, J. N. (1996), "The How and the Why of Response Latency Measurement in Telephone Surveys," in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, eds. N. Schwarz and S. Sudman, pp. 319–346, San Francisco: Jossey-Bass.

Cameron, C., and P. Trivedi (2009), *Microeconometrics Using Stata*, College Station, TX: Stata Corp.

Cannell, C. F., P. V. Miller, and L. Oksenberg (1981), "Research on Interviewing Techniques," in *Sociological Methodology*, ed. S. Leinhardt, pp. 389–437, San Francisco: Jossey-Bass.

Couper, M. P., and F. Kreuter (2013), "Using Paradata to Explore Item-Level Response Times in Surveys," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 271–286.

de Leeuw, E. D., J. Hox, and M. Huisman (2003), "Prevention and Treatment of Item Nonresponse," *Journal of Official Statistics*, 19, 153–176.

Edwards, W. S., and D. Cantor (2004), "Towards a Response Model in Establishment Surveys," in *Measurement Errors in Surveys*, eds. P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, pp. 211–236, Hoboken, NJ: John Wiley & Sons.

Fielding, A., and H. Goldstein (2006), "Cross-Classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review," Research Report RR791, Department for Education and Skills, University of Birmingham.

Fricker, S., M. Galesic, R. Tourangeau, and T. Yan (2005), "An Experimental Comparison of Web and Telephone Surveys," *Public Opinion Quarterly*, 69, 370–392.

Gray, M., J. D'Ardenne, M. Balarajan, and N. Uhrig (2011), "Cognitive Testing of Wave 3 Understanding Society Questions," Understanding Society Working Paper Series, No. 03, Colchester.

Harkness, J., F. J. R. Van de Vijver, and P. Ph. Mohler (2003), *Cross-Cultural Survey Methods*, Hoboken, NJ: Wiley.

Holbrook, A., Y. I. Cho, and T. Johnson (2006), "The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties," *Public Opinion Quarterly*, 70(4), 565–595.

Johnson, T. P., A. L. Holbrook, and Y. I. Cho (2006a), "The Effects of Acculturation on Survey Question Comprehension among Latino Respondents in the U.S.," ASA Section on Survey Research Methods, 3186–3192, available at http://www.amstat.org/sections/srms/proceedings/y2006/Files/JSM2006-000500.pdf, last accessed May 28, 2015.

Johnson, T. P., Y. I. Cho, A. L. Holbrook, D. P. O'Rourke, R. B. Warnecke, and N. Chavez (2006b), "Cultural Variability in the Effects of Question Design Features on Respondent Comprehension of Health Surveys," *Annals of Epidemiology*, 16, 661–668.

Johnson, T. P., D. P. O'Rourke, N. Chavez, S. Sudman, R. B. Warnecke, and L. Lacey (1997), "Social Cognition and Responses to Survey Questions among Culturally Diverse Populations," in *Survey Measurement and Process Quality*, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, pp. 87–113, New York: Wiley.

Kaminska, O., A. L. McCutcheon, and J. Billiet (2010), "Satisficing among Reluctant Respondents in a Cross-National Context," *Public Opinion Quarterly*, 74(5), 956–984.

Krosnick, J. A. (1991), "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology*, 5, 213–236, doi:10.1002/acp.2350050305.

Krosnick, J. A. (2002), "The Causes of No-Opinion Responses to Attitude Measures in Surveys: They Are Rarely What They Appear to Be," in *Survey Nonresponse*, eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, pp. 87–100, New York: Wiley.

Krosnick, J. A., S. Narayan, and W. R. Smith (1996), "Satisficing in Surveys: Initial Evidence," *New Directions for Evaluation*, 70, 29–44.

Laganà, F., G. Elcheroth, S. Penic, B. Kleiner, and N. Fasel (2011), "National Minorities and Their Representation in Social Surveys: Which Practices Make a Difference?" *Quality & Quantity*, 47(3), 1287–1314.

Lenzner, T., L. Kaczmirek, and A. Lenzner (2009), "Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment," *Applied Cognitive Psychology*, 24, 1003–1020.

Lipps, O., F. Lagana, A. Pollien, and L. Gianettoni (2013), "Under-Representation of Foreign Minorities in Cross-Sectional and Longitudinal Surveys in Switzerland," in *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies*, eds. Font, J., and M. Méndez, pp. 241–267, Amsterdam University Press.

Loosveldt, G., J. Pickery, and J. Billiet (2002), "Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey," *Journal of Official Statistics*, 18, 545–557.

Moore, R. J., and D. W. Maynard (2002), "Achieving Understanding in the Standardized Survey Interview: Repair Sequences," in *Standardization and Tacit Knowledge*, eds. D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaffer, and J. van der Zouwen, pp. 281–312, New York: Wiley.

Narayan, S., and J. A. Krosnick (1996), "Education Moderates Some Response Effects in Attitude Measurement," *Public Opinion Quarterly*, 60, 58–88.

Papke, L. E., and J. Wooldridge (1996), "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 619–632.

Peytcheva, E. (2008), "Language of Administration as a Source of Measurement Error: Implications for Surveys of Immigrants and Cross-Cultural Survey Research," Ann Arbor, MI: ProQuest, UMI Dissertation Publishing.

Saris, W., M. Revilla, J. A. Krosnick, and E. M. Shaeffer (2010), "Comparing Questions with Agree/Disagree Response Options to Questions with Construct-Specific Response Options," *Survey Research Methods*, 4, 61–79.

Snijders, T., and R. Bosker (1999), *Multilevel Analysis*, Newbury Park, CA: Sage.

Suchman, L., and B. Jordan (1990), "Interactional Troubles in Face-to-Face Survey Interviews," *Journal of the American Statistical Association*, 85(409), 232–253.

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

Warnecke, R. B., T. P. Johnson, N. Chavez, S. Sudman, D. P. O'Rourke, L. Lacey, and J. Horm (1997), "Improving Question Wording in Surveys of Culturally Diverse Populations," *Annual Epidemiological Psychology*, 7, 334–342.

Wenemark, M., A. Persson, H. N. Brage, T. Svensson, and M. Kristenson (2011), "Applying Motivation Theory to Achieve Increased Response Rates, Respondent Satisfaction, and Data Quality," *Journal of Official Statistics*, 27(2), 393–414.

Yan, T., and R. Tourangeau (2008), "Fast Times and Easy Questions: The Effects of Age, Experience, and Question Complexity on Web Survey Response Times," *Applied Cognitive Psychology*, 22(1), 51–68.