

Mother Tongue or Non-Native Language? – The Influence of Language on Response Behavior in Surveys

Florian Heinritz

*Leibniz Institute for Educational Trajectories &
Universität Hamburg*

Abstract

Today, an increasing number of surveys offer respondents the choice of which language they want to answer the questionnaire. In later data analysis, however, the language in which the respondent answers the questions is often ignored, and no distinction is made regarding whether that language is the respondent's mother tongue. Several psychological theoretical considerations and empirical observations indicate that respondents' answering behaviors are influenced by whether the questions are presented in their mother tongue or a non-native language. Therefore, the extent to which these mechanisms and effects of language used are also applicable and relevant in social science studies remains unclear. Based on models of cognitive load, satisficing, and language-dependent memory, the influence of language nativeness on response behavior is explained from a theoretical point of view. The research question will be answered by analyzing the data from the refugee study ReGES (Refugees in the German Educational System). The results of the analyses show that there is a difference in response behavior depending on whether a question is answered in a mother tongue or a non-native language. The implications, both from a survey methodological point of view and for further research, will be discussed.

Keywords: non-native language effect; language; multilingual surveys; response behavior; refugees



© The Author(s) 2025. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Due to the increase in labor migration (International Labour Office [ILO], 2010, 2018) and the number of refugees (International Organization for Migration [IOM], 2019), multilingual interviews recently became more relevant. Since multilingual interviews have already been conducted commonly in many countries with multiple national languages, the methodological challenges of conducting identical questionnaires in different languages (e.g., Hunt & Bhopal, 2004; McKay et al., 1996; Pan et al., 2014) and other methodological aspects of multilingual surveys have already been investigated in detail (e.g., Blohm & Diehl, 2001; Dotinga et al., 2005; Schoua-Glusberg, 2004).

However, even if these methodological challenges are considered, there can be differences in answering questions depending on the language used (e.g. Peytcheva, 2018). Nevertheless, research regarding language differences has often focused on bilingual respondents. But the following three developments raise the relevance for shifting the focus: First, due to the increase in migrants, there is an increasing group of people whose mother tongue is not one of the national languages of a country. Second, the number of forced migration is increasing at the same time, which typically means that people cannot properly prepare their migration by, for example, learning the national language of the host country. And third, relatively new survey technologies, such as multilingual computer- or web-based questionnaires, make it possible that the number of languages offered in a survey from which the respondent can choose no longer depends on the interviewer's language skills, as is usually the case in interviews with interviewers.

The interaction of these three points—the greater diversity of different mother tongues within countries, the rising number of people without knowledge of the national languages and the technical possibility for respondents to select their preferred language for responding to questionnaires or individual questions from a range of languages—opens a new question: Does it make a difference whether respondents answer a question in their mother tongue rather than in a non-native language? Analyses of numerous psychological studies and experiments suggest that there is a difference between answering a question in a non-native language as opposed to the mother tongue (for a summary see Hadjichristidis et al., 2019). However, whether and how these effects are also relevant in surveys has rarely been analyzed (e.g., Kappelhof, 2017). Therefore, the aim of this paper is, on the one hand, to investigate whether differences in response

Notes

The project ReGES is funded by the German Federal Ministry of Education and Research under grant number FLUCHT03. However, the authors have sole responsibility for the content of this publication.

Direct correspondence to

Florian Heinritz, Leibniz Institute for Educational Trajectories & Universität Hamburg,
E-Mail: florian.heinritz@lifbi.de

behavior depend on whether questions are posed and answered in a mother tongue instead of a non-native language. On the other hand, the influence of language nativeness will also be analyzed for other practical aspects relevant for surveys, such as the duration of an interview or the accuracy of statements.

For this purpose, data from a German refugee survey ReGES (Will et al., 2021) are used to analyze the extent to which the language used influences the length of the survey, the accuracy of the information provided, and the actual response. Analyzing the data from this refugee study enables an investigation of the impact of language on response behavior in actual surveys based on computer-assisted self-interviews (CASI) in eight languages and—since the refugees are all newly arrived immigrants—allows a clear distinction between mother tongue and non-native language. Conversely, the data is not based on an experimental study.

Having this in mind, theoretical models and explanations are presented, and hypotheses are formulated in a first step. Subsequently, I briefly describe the data and the operationalization of the subsequent analyses before the results of the multivariate analyses are presented. The results show that items are answered more quickly in a mother tongue than in a non-native language. Likewise, by looking at items on gender roles and religiosity, it turns out that items about social norms are answered more in accordance with the norms associated with the mother tongue if these items are answered in a mother tongue. Finally, the results are discussed in a concluding section.

Theoretical Background and Previous Research

The following theoretical considerations focus on why people answer questions in a survey differently in their mother tongue than in a non-native language. As mentioned in the introduction, this paper is about people who have one or more mother tongues and have later learned a non-native language.

From a survey methodological point of view, the influence of language on response behavior can be explained by the model of satisficing (Krosnick, 1991). In addition, I will focus on two psychological approaches to explain why response behavior may change depending on the language used to answer the question: cognitive load theory and language-dependent memory. As shown in Figure 1, I will apply these models to the four survey methodology-relevant steps of answering a question, which are the comprehension of the question, retrieval of relevant information, judgment, and response (Tourangeau, 1984).

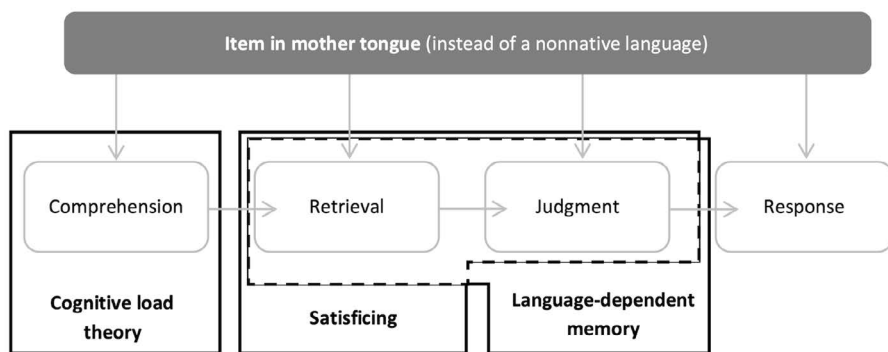


Figure 1 Overview of theories and their link to the respondent's task in the response process

While each theoretical model is used to derive a hypothesis in the following sections, it should be noted that the three theoretical models are not based on contradictory assumptions but rather complement each other.

Cognitive Load Theory

The initial task of the respondent in the process of answering a question is to comprehend the question or, more precisely, to understand the text (Tourangeau, 1984). Whereas it can be assumed that—despite different language skills (and depending on the survey mode used also despite different reading skills)—respondents have sufficient language skills to understand questions posed in their mother tongue, language skills of a respondent in a non-native language usually vary greatly.

Especially in the case of the group of refugees focused on in this paper, it can be assumed that the non-native language skills of the host country's language are lower than in the mother tongue, since refugees often have to leave the country unexpectedly and sometimes do not yet know in which country they will be placed. This makes texts in a non-native language comparatively harder to comprehend. Additionally, it is necessary to consider that items in surveys are even more difficult to understand because respondents usually cannot understand the meaning of the text from the context but have to understand each item individually (Calderón et al., 2006, pp. 50–51).

However, even if people understand a non-native language almost as well as their mother tongue, it is still the case that the cognitive load is higher for understanding a non-native language than a mother tongue (e.g., Hasegawa et al., 2002). Since cognitive load is defined as the amount of working memory

resources used to complete a mental activity such as comprehending a text (e.g., Paas & Van Merriënboer, 1994; Sweller, 1994) and these human memory resources are limited, a high cognitive load also means a higher cognitive effort (e.g., Paas et al., 2003).

As a result, comprehension of the questions is more challenging in a non-native language, either due to a lack of language proficiency or the higher cognitive load required to understand questions in a non-native language. The time required to answer items should vary depending on the language used. Therefore, respondents should need less time to respond to questions posed in their mother tongue than they would to questions posed in a non-native language:

H1: *Answering questions in a mother tongue takes less time than answering questions in a non-native language.*

Satisficing

As a model based on a combination of cognitive load theory and general rational choice theory (see Esser, 1990), satisficing can also be used to explain differences in response behavior based on language in the following two steps: retrieval and judgment (Tourangeau, 1984). According to rational choice theory (RCT), respondents evaluate in the answering process of each question how the highest possible subjective expected utility can be achieved by answering the question. However, since neither cost nor negative sanctions nor particularly high gains are expected in a scientific, voluntary survey, the subjective expected utility is evaluated as relatively low. Therefore, the use of cognitive load is often reduced to a minimum, meaning that respondents put less effort into answering the question when the task is difficult. This could, of course, be done by responding to the question in the mother tongue. However, this paper does not delve further into the reasons for answering a question in a non-native language when the mother tongue is available.

This minimal cognitive effort, which is also known as satisficing (Krosnick, 1991), can influence response behavior by reducing the cognitive load. One possible consequence of satisficing is reducing the cognitive load through heaping (Gideon et al., 2017). This means that the respondents try to minimize their cognitive effort by rounding open numerical answers instead of choosing the more cognitively demanding process of intensively retrieving the exact number and judging whether the number given is actually correct. Therefore, rounded answers are less accurate than unrounded answers (Battisin et al., 2003). A problematic consequence of this on data report and analysis is that rounded or estimated (and thus less accurate) answers given by the respondent can lead to a loss of validity. Therefore, it is also important to minimize the measurement error of heaping.

According to satisficing theory, the difficulty of the task is a significant factor influencing satisficing and, consequently, heaping (Krosnick, 1991). As previously stated in the considerations for Hypothesis 1, it can be assumed that answering a question in a non-native language increases the task difficulty and thus fosters satisficing. Therefore, the following relationship is predicted in Hypothesis 2:

H2: *Items presented in a mother tongue tend to be answered more accurately.*

Language-Dependent Memory

A second model that often serves as a possible explanation in this context is language-dependent memory (e.g., Marian & Neisser, 2000), which is applied here to the two steps of retrieval and judgment. This model relies on different circumstances of language learning. For example, the emotional context of the learning hypothesis states that a language is associated with emotions when that language is learned and used in an emotional context (Harris et al., 2006). It can therefore be assumed that the mother tongue, which is learned in childhood under the influence of many emotions, is much more strongly linked to emotions than a non-native language, which is usually learned in a (less emotional) educational context. The relationship of social norms and language is similar (see Nichols et al., 2016): Social norms are mainly internalized in the mother tongue and are thus more activated by the mother tongue than in a non-native language. Therefore, such experiences and learned norms are stored in long-term memory in the language in which the experiences were made or norms were acquired (e.g., Marian & Fausey, 2006; Marian & Kaushanskaya, 2004). These theoretical assumptions refer not only to differences between the mother tongue learned in childhood and a later learned non-native language but also, in the case of bilingual persons, to memories, norms or emotions associated with different languages (e.g., Danziger & Ward, 2010; Dewaele & Nakano, 2012; Marian & Kaushanskaya, 2004).

The theory of language-dependent memory implies that emotions and social norms play a much greater role in the mother tongue than in a non-native language, in which “cool-headed responses toward certain moral dilemmas” and “less condemnation of moral and social violations” can be expected (Hadjichristidis et al., 2019, p. 264). Therefore, depending on the emotions and social norms associated with a language, questions can be answered differently according to the language in which they are presented, and norms are less activated in a non-native language (Geipel et al., 2015).

For these reasons, it can be assumed that those questions, where norms or emotions linked to a language have to be retrieved and judged in the answering process, will be answered differently in a mother tongue or a non-native

language. This means that respondents would answer such questions in their mother tongue more emotionally and with the knowledge of social norms linked to their mother tongue. In a more formally learned non-native language, the associated social norms would be less activated, and emotionality would also decrease. It can be assumed that questions about social norms or emotions in a non-native language are less likely to be answered in accordance with the social norms and emotions incorporated in the mother tongue but rather answered more rationally and therefore more in conformity with the social desirability resulting from the interview context (e.g., the culture of the country where the interview takes place). Hypothesis 3 states therefore:

H3: *If items about social norms or emotions are presented in a mother tongue, the answers will be answered more in accordance with the norms associated with the mother tongue.*

Current State of Research

The question of how language influences response behavior has been addressed by many studies in different disciplines, each with a different focus. However, almost no study has analyzed the actual influence of the use of mother tongue or non-native language on response behavior in surveys. To obtain a sense of the empirical evidence supporting the hypotheses, some studies that have dealt with assumptions similar to the abovementioned hypotheses are briefly presented below.

In the United States, where a large proportion of the population is bilingual, studies have more frequently investigated the effect of language on response behavior in surveys (e.g., Diaz-Morales et al., 2006; Guarnaccia et al., 1989; Pérez, 2009; Welch, 1973). However, they have made no distinction between mother tongue and non-native language. Most of these studies show that the response behavior depends in different ways on whether the questionnaire was completed in English or another language.

As one of the few studies that focused on differences between mother tongue and non-native language, Harzing and Maznevski (2002) showed in an experiment with students that questions are answered in accordance with the cultural values linked to a language. This finding corresponds to other studies (e.g., Lee, 2001; Marin et al., 1983), although they did not distinguish between mother tongue or non-native language, and is in line with the theoretical assumptions of language-dependent memory theory. Similarly, the results from Kappelhof (2017), using data from a Dutch study on ethnic minorities, show that individuals answer items on family ties more traditionally in their mother tongue.

In addition, various psychological experiments have focused explicitly on the differences in response behaviors between respondents using their mother

tongue or a non-native language (for an overview, see Hadjichristidis et al., 2019). Some studies have shown that people perceive non-native languages less emotionally (e.g., Caldwell-Harris & Ayçiçeği-Dinn, 2009; Harris et al., 2003). Other studies have also shown that decisions in non-native languages were therefore made more rationally and less emotionally (e.g., Cipolletti et al., 2015; Costa et al., 2014; Geipel et al., 2015; Hadjichristidis et al., 2019; Hayakawa & Keysar, 2018; Shin & Kim, 2017).

In sum, these studies show that parts of the assumptions have already been empirically proven, so the theoretical explanations are a useful basis for the assumptions in the individual hypotheses. To test the hypotheses empirically, the data for the later analyses will be described in the following section.

Data and Methods

Unlike many recent studies on differences due to mother tongue and non-native language, the hypotheses are tested using a large-scale study. Specifically, data from the first wave from 2018 of the German refugee study ReGES “Refugees in the German Educational System” are used. Since a considerable number of children and young refugees came to Germany in the context of asylum immigration in the mid-2010s, this study focuses on the educational trajectories of young refugees by interviewing adolescents and parents, even though the sampling units were young refugees children (at least four years old but not attending school at that time) and refugees adolescents (between 14 to 16 years old) (Will et al., 2021). The target population was sampled via a complex, multi-stage sampling process from the German registration office across five federal states in Germany. For this purpose, a random sample of the nationalities of the most common refugee nationalities in Germany at that time was taken (for details, see Steinhauer et al., 2018). Respondents were contacted personally by interviewers after receiving an invitation letter, resulting in 5,711 completed interviews in the first wave (for further details regarding the sample over the waves, see Heinritz & Will, 2021 and von Maurice & Will, 2023).

To prevent panel conditioning, which may also influence response behavior, only the first wave of the study ReGES is considered. The advantage of this study is that the questionnaires of the first wave were offered in eight different languages: English, German, Arabic, Kurmanji, Pashto, Tigrinja, Farsi, and French. The languages—in addition to German as the original language of the study—were chosen based on the most common official languages of the respondents’ countries of origin (Gentile et al., 2019), knowing that they are not always the mother tongues of the respondents. In all these languages, native speakers of the respective languages were employed as interviewers. In order to contact the respondent in the correct language, the nationality of the respondents was

used as an indicator for the language in which the interviewer should contact the person (e.g., respondents from Syria should be contacted by Arabic-speaking interviewers).

Sample

The main part of the interviews was a CASI. As it could be assumed that there were many illiterate people in the sample, the CASI was offered with audio files and the interviewers were also allowed to read the question aloud to the respondents. In order to minimize the possible influences and interactions of the interviewer¹, only those CASI interviews are analyzed in which the interviewers did not read out questions ($n = 2,031$). In addition, there were some cases ($n = 84$) where the respondents stated that they could not read, but neither used audio files nor asked the interviewer for help. These implausible cases are excluded as well as cases with implausible data on the mother tongues. This automatically excluded people with poorer reading skills in the analyses sample. Therefore, a total of 1,865 persons are considered in the following analyses. Furthermore, since it can be expected that social desirability differs depending on the country of origin or culture (Tourangeau & Yan, 2007, p. 860), the test of Hypothesis 3 includes only persons from Syria as the largest group of the sample with the same country of origin and therefore similar cultural background.

Table 1 provides a first description of the sample that served as the basis for the multivariate analyses. When looking at the sample in the second column in Table 1, which reports the mother tongues (multiple answer) of the respondents, it can be observed that the languages used correspond to the official languages of the countries of origin of the sample: Arabic (as the official language in Syria and Iraq) is the most common mother tongue, with 81.29% of respondents listing this language as one of their mother tongues, followed by Kurmanji (as an official language in Iraq and spoken in parts of northern Syria) and Farsi (as an official language in Iran), with 8.26%. Although Kurmanji was the mother tongue of many respondents, items views in this language cannot be included in the following analyses, as the complete translation of Kurmanji had to be revised during the fieldwork and quality problems remained due to the complexity of the language (see Gentile et al., 2019).

In ReGES, respondents were free to choose the language used for answering. On the one hand, the respondents could choose the interview language at the beginning; on the other hand, the language could be changed individually for

1 Interviewers can influence response behavior in many ways. These includes characteristics of the interviewer such as ethnicity, gender or age (e.g., Glantz and Michael, 2014; Groves et al., 2009; Loosveldt, 2008).

Table 1 Distribution and use of language in the sample in percent

Units	Mother tongues of respondents	Languages used for answering	Language matches in each language
	Respondents in analyses sample	Screens of analyses sample	Screens of analyses sample displayed in the language
Arabic	81.29	73.78	90.99
German	1.39	19.38	3.75
Farsi	8.26	6.03	98.94
English	7.56	0.71	24.86
Tigrinya	0.16	0.06	100
Pashto	0.70	0.08	100
Kurmanji	19.46	*	*
French	0.59	-	-
N	1,865	219,325	219,325

Source: ReGES data, own calculations, Wave 1.

*Language was not considered in the analysis sample.

each question.² The third column in Table 1 illustrates the proportion of screens displayed in each language. For example, 73.78% of the screens were last displayed in Arabic and 6.03% in Farsi, which roughly reflects the proportion of people with these languages as mother tongues. This distribution of mother tongues seems to correlate in most cases with the languages actually used in the survey (see the second column in Table 1). In fact, in 90.99% of the screens that were answered in Arabic, Arabic was also the mother tongue, and in Farsi, it was also the mother tongue in 98.94% of the cases (see the fourth column in Table 1). The greatest difference between the number of native speakers and number of users of this language in the interviews can be seen for German. Although only 1.39% of the respondents stated that German was their mother tongue, 19.38% of the screens were answered in German. Therefore, it is not surprising that only 3.75% of the items answered in German were answered by native speakers.

2 As mentioned in the beginning, there are more studies that offer several languages and where the respondent can choose the language. However, one problem of the data analysis of other studies that should not be underestimated is that sometimes the language used is not logged at all, or at least not listed in scientific use files, or the respondent's mother tongue is not surveyed. This once again illustrates the lack of attention given to the possible influence of language on the data.

In total, 74.10% of the screens analyzed were answered by the respondents in their mother tongue. However, these screens with language matches are distributed differently among the respondents. A total of 64.34% of respondents conducted the entire survey in their mother tongue and thus had a language match for each question, whereas 21.66% did not answer a single question in their mother tongue. This already indicates that only a small proportion of respondents took advantage of the opportunity to change languages during the survey. In fact, looking at the number of times respondents changed languages, 78.18% of respondents did not change language at all (regardless of whether the language used was their mother tongue or not) and less than 4 % changed the language more than 10 times.

A parent interview in this first wave contained just over 300 questions and an adolescent interview contained at least 250 questions. However, not all items were considered. Items for which translation quality problems were identified through the translation process of follow-up waves and for which the translation was therefore modified in the follow-up waves are excluded in the corresponding language. With all these limitations, the analysis sample contained 1,865 persons who together answered a total of 227,448 items. Due to the restrictions made by operationalization (see below), for 219,325 items, it was possible to clearly identify the language in which the item was answered (although this number and the number of respondents will be lower in the multivariate analyses due to missing values in the data).³ This is also the basis for Table 1.

Operationalization

The independent variable of whether an item was viewed in a native or non-native language is no longer clearly identifiable in the case of items where the respondent changed the language several times. Therefore, only items that were either viewed in only one language or for which the language was switched only one time by the respondents were considered. In the latter cases, it is assumed that the item was answered in the language that was switched to. In Hypothesis 3, in which the theoretically assumed explanation for the respondent's behavior is language-dependent memory, only items that were viewed in one language without changing language were considered to ensure that the memories were associated with only one language.

The time for answering an item in Hypothesis 1 is measured as the respondent's cumulative time spent viewing an item. All items viewed by a respondent

3 For detailed descriptive analyses of the complete sample of the ReGES study see Will et al. (2018) and Appendix 1.

for more than ten minutes⁴ were considered interview interruptions and were not included in the analyses.

To measure the accuracy of open answers in Hypothesis 2, all open numerical answers except dates are considered. Dates are not considered because, on the one hand, dates such as month and year of birth or month and year of schooling are easier to remember and therefore less cognitively demanding (Burton & Blair, 1991); on the other hand, cultural differences⁵ have to be considered. For these reasons, only open numerical answers that refer to frequencies are considered.⁶ In each questionnaire, there was a maximum of 7 open numerical answers. As done in comparable research (e.g., Holbrook et al., 2014; Schober et al., 2015), it is expected that all answers divisible by 5 will tend to be rounded so that the accuracy of open answers is operationalized in binary form: If the answer is divisible by 5, it is assumed to be rounded and therefore less accurate ($1 = \text{rounded}$).

To test Hypothesis 3, items on religiosity and gender role attitudes are analyzed because it can be assumed here that social norms differ between Syria, the country of origin, and Germany, the host country. It can be assumed that gender roles are more traditional in Syria than in Germany and that these gender roles are anchored in the mother tongue of the respondents. If, however, items about gender roles (using a sum score from 4 “egalitarian” to 16 “traditional”) or religiosity (using a four-point ordinal scale from “not at all religious” to “very religious”) are answered in a non-native language, it can be assumed that the question will be answered more rationally and will probably be answered in a more socially desirable way, e.g., in accordance with the norms of the host country. In the case of Germany as the host country, therefore, the answer will be more liberal or secular.

Methods of Analysis

To test the hypotheses, different regression analyses are performed for each hypothesis. It is important to remember that the objective of the ReGES study was to describe educational trajectories of refugees. Consequently, the data

-
- 4 It would also be reasonable to evaluate an interruption at 5 minutes or at 15 minutes; however, the results presented would not differ in the core of the conclusions.
 - 5 This is less about different calendars than, for example, the phenomenon that the target group of the ReGES study is born in January more often than average. One potential explanation for this phenomenon is that in some cultures, birthdays are not a significant event and are therefore not celebrated. Therefore, some refugees may be unaware of their birthday, resulting in the mention of 01.01. as the birthday in official documents to avoid leaving the date and month empty.
 - 6 An overview of these 13 items can be found in Appendix 2. However, since these 13 items also include many items that were asked separately for each child in the parent questionnaire, the number of items actually asked varies greatly.

were not collected through an experimental design.⁷ Therefore, possible confounding variables are included in the analyses: Whether an item is answered in a mother tongue or not can be influenced by the country of origin (e.g., the national language of the country of origin is not offered as the language of the survey), by the length of stay in Germany (e.g., longer stay in Germany improves German language skills and thus the probability that items are answered in German), by education (e.g., higher education usually means better non-native language skills and thus higher chances to answer items in a non-native language) and age (e.g., cohort effect: today, non-native languages are taught more often at school, so that young people are more likely to be able to answer items in a non-native language). All these variables could also affect the dependent variable of the respective hypothesis in different ways.⁸

Additionally, a translation issues may result in respondents preferring to answer items in a language other than their mother tongue. Although all items for which translation problems were identified (when using them in later waves) were excluded from the analyses, regional discrepancies in item comprehension may persist, particularly in Arabic and Kurmanji (Gentile et al., 2019). Furthermore, it is also possible that the language was changed for one of the items identified as having been translated inaccurately and that the language was not changed back for the next items (which are included in the analyses). Additionally, unidentified translation issues can increase the cognitive load and make the understanding of an item more difficult. Therefore, the translation is additionally included as a control variable in Hypotheses 1 and 2, which are based on cognitive load and item understanding.

In Hypothesis 1, the length of an item, as measured by the number of characters, may affect respondents' preference to read longer texts in their native language. At the same time, it can be assumed that the length of a text also influences the time taken to answer an item. Therefore, this variable is also included in Hypothesis 1 as a control variable. When examining all possible screens that could be displayed in the CASI questionnaire, a quantitative analysis reveals that the average character lengths of the question in its original German version is

7 Obviously, languages were not randomly assigned. However, it can be assumed that the selection of language depends less on the characteristics of the respondents and more on the individual interview situation. An analysis of language changes within the analyzed CASI and across the follow-up interviews (that were not completely self-administered) shows, for example, that the selected language remained constant for less than 25% of the parents analyzed here. Examples of these situational factors are the availability of interviewers in the respondent's mother tongue or comprehension problems due to translation issues.

8 Furthermore, the proficiency in different languages may also influence the language in which an individual prefers to respond, as well as linguistic comprehension and the cognitive load associated with language processing. Thus, language competence may be an additional confounding variable. Unfortunately, this variable is not included in the ReGES study for all languages, which must be considered when interpreting the results.

239. The mean length of the texts of the questions in English was slightly shorter (223 characters) and considerably longer in Arabic with 312 characters.

Except for the two last-mentioned variables, all other control variables are characteristics of the respondents and not of the item, which are the actual units of analysis. Thus, the items are nested in a two-level structure by respondents, and standard errors clustered by respondents are estimated in the regression analyses. Furthermore, since the proportion of men and women in the sample is clearly biased⁹, gender is also included as a control variable to avoid sample bias.

Empirical Results

As discussed in the previous section, different regression models are used in Table 2 for the multivariate analyses. In Table 2, Model 1.1, using a linear regression model (OLS model, robust standard errors clustered 1,490 respondents), the significant coefficient shows that items answered in a mother tongue took an average of 1.17 seconds longer. A possible explanation for this could be that, on the one hand, Arabic was the native language in 90.99% of the items that were answered in Arabic. Arabic items have more characters on average than German or English, which more often represent non-native languages. Keeping the control variables constant, the highly significant coefficient shows that respondents need an average of 2.44 seconds less time to answer an item in their mother tongue than an item in a non-native language. Therefore, the results in Model 1.2 support H1.

In contrast, H2 cannot be confirmed based on the analyses. The coefficient of the binary logistic regression (average marginal effects, robust standard errors clustered for 1,461 respondents) in Model 2.2 is not significant after including the control variables, even at a significance level of 10%.

The linear regression (OLS model) in Model 3.1 in Table 2 confirms the assumption of H3 and shows that items are answered more traditionally in a mother tongue than in a non-native language. The magnitude and significance of this coefficient increases when the control variables (in Model 3.2) are included so that H3 can be confirmed by analyzing gender roles. A separate analysis of the models for adolescents and parents (see Appendix 3) shows that, including the control variables, the effect is greater for parents with a coefficient of 1.72, while

⁹ Although this imbalance corresponds to the gender distribution of the refugees in Germany (e.g., Neske & Rich, 2016; Rich, 2016), given that families are interviewed in the ReGES study, it can be assumed here (and the feedback from the interviewers has also shown) that fathers as “classical heads of household” are more likely to answer the CASI interview than women so that the sample seems to be slightly self-selective.

Table 2 Multivariate analyses

Dependent variable	Hypothesis 1		Hypothesis 2		Hypothesis 3 (gender roles)	
	Duration in seconds		Accuracy (rounded = 1)		Gender roles, 4 (egalitarian) to 16 (traditional)	
	Model 1.1	Model 1.2	Model 2.1	Model 2.2	Model 3.1	Model 3.2
Item in mother tongue	1.17* (0.47)	-2.44*** (0.67)	-0.06* (0.02)	0.02 (0.04)	0.76** (0.25)	0.87*** (0.25)
<i>Controls:</i>						
Country of origin		✓		✓		✓
Age		✓		✓		✓
Length of stay in Germany		✓		✓		✓
Education		✓		✓		✓
Gender		✓		✓		✓
Translation		✓		✓		
Length of text		✓				
Pseudo/ Adjusted R^2	.000	.011	.003	.021	.007	.033
N	168,459	168,459	2,667	2,667	1,143	1,143

Source: ReGES data, own calculations, Wave 1.

Notes: Estimates with standard errors in parentheses.

* $p < .05$,

** $p < .01$,

*** $p < .001$.

the effect is not significant for adolescents. One possible explanation for this is that gender roles are not yet as pronounced in adolescents and are therefore less linked to the mother tongue.

Table 3 shows that the results of the ordinal logistic regression (average marginal effects) with religiosity as the dependent variable correspond to H3. It can be assumed that most people tend to experience religion in their mother tongue, so that the level of religiosity is reported to be higher in the mother tongue than in a non-native language. Even when controlling for variables such as age, education or length of stay in Germany, the significant average marginal effects show that, for example, the probability that a person states that he or she is very religious is almost 3 percentage points higher in the mother tongue than if the

Table 3 Ordinal logistic regression for religiosity (average marginal effects)

	Model 1	Model 2
Item only in mother tongue		
<i>Not at all religious</i>	-0.03** (0.01)	-0.03** (0.01)
<i>Not very religious</i>	-0.07** (0.02)	-0.06** (0.02)
<i>Quite religious</i>	0.08** (0.03)	0.07** (0.03)
<i>Very religious</i>	0.03*** (0.01)	0.03** (0.01)
Controls:		
Age, length of stay in Germany, gender, education		✓
Adjusted R^2	.004	.018
N	1,377	1,377

Source: ReGES data, own calculations, Wave 1.

Notes: Estimates with standard errors in parentheses.

* $p < .05$,

** $p < .01$,

*** $p < .001$.

item was answered in a non-native language. In line with the analysis about gender roles, the effect in Model 2 is also stronger for the parents and no longer significant for the adolescents in separate analyses (see Appendix 4).

Discussion

Due to the increasing number of migrants, there are more and more people who have a mother tongue other than the national language. At the same time, technological innovation has made it possible for respondents in many surveys to choose whether to answer questions in their native language or in a non-native language. The present analyses based on data from the German refugee study ReGES have shown that there is a difference in response behavior when a question is answered in a native language instead of a non-native language, even if not all hypotheses could be confirmed. The data have shown that when considering the time for answering an item, the cognitive load seems to be higher and the understanding of a question is more difficult when a question is presented and answered in a non-native language (H1). Depending on how long a survey is,

this effect can be clearly noticeable in the overall time and thus possibly cause exhaustion or reduce the respondents' willingness to cooperate in subsequent follow-up surveys of a panel study.

A relationship between more precise information and language nativeness (H2) could not be demonstrated. However, since the first model for Hypothesis 2 without control variables shows a significant effect, it can be assumed that both the willingness for higher cognitive effort and to give accurate unrounded information and the willingness to answer an item in a non-native language that is difficult to understand depend strongly on the control variables included (such as age, gender or on the cognitive willingness or ability to perform in general, for which the educational level can be regarded as an indicator).

The significant coefficients in Model 2.2 and 3.2 (Table 2) show a clear correlation between the language used and responses to sensitive questions. However, regarding causality, a few limitations of the data and the study design of the ReGES study must be considered here. It is possible that an individual's religiosity or attitudes toward gender roles may also influence the choice of language. For example, more liberal respondents might be more willing to answer a survey in a non-native language. However, field experience indicates that the selection of language is more dependent on the interview situation. For instance, an analysis of the language used by parents within the CASI of the first survey wave combined with the starting language used in the follow-up interviews that were not fully self-administered reveals—without considering language changes within the interviews of the follow-up waves—that the language used remained consistent for less than 25% of the parents.¹⁰ Furthermore, the fact that respondents have no rational reasons for voluntarily conducting the interview in a non-native language instead of their native language (according to the assumptions of RCT and cognitive load theory) strengthens the hypothesis that external, situational factors and peculiarities of the ReGES study are responsible for the fact that respondents did not complete the survey in their mother tongue, although this was offered in almost all cases. The data showed—similar to other studies (e.g., Kinnunen et al., 2015)—that even if the mother tongue is offered, the mother tongue is not automatically chosen. An investigation of these factors would provide more evidence (and would also help to evaluate whether the costs and efforts for a multilingual survey are truly worth it). In order to conduct a

10 It is clear that the refugees' German language abilities will continue to develop over time, allowing for an increasing number of respondents to be interviewed in German in subsequent survey waves. However, if the language selected is dependent on the characteristics of the respondents (e.g., religiosity), it can be reasonably assumed that the language used by the respondents will remain stable throughout the interview and across survey waves. This indicates that situational factors (e.g., availability of native speaker interviewers, comprehension issues, etc.) may be more influential than respondents' self-selection in determining language choice.

more detailed analysis of the effects of the language used, future survey experiments can be used to confirm and understand the aforementioned effects.

Furthermore, the individual relational explanatory power of the language match between the respondent's mother tongue and the language used with a goodness of fit R^2 of less than .05 in each analysis is quite low in all the models presented, so it can be assumed that there are many more aspects that contribute to explaining different response behavior. For example, cultural considerations as well as linguistic characteristics of a language (such as the possible influence of grammatical gender (e.g., Boroditsky et al., 2003; Garnham et al., 2016)) could also help to explain language-dependent response behaviors. Regarding linguistic aspects, another point that would be relevant for further research would be whether and to what extent dialects also cause a difference in the response to items. According to language-dependent memory theory, this should have an influence, which would be relevant for the language focused on in this paper, Arabic. Indeed, it can be assumed that not all respondents who stated that Arabic is their mother tongue actually learned the Modern Standard Arabic used in the questionnaire as their mother tongue but rather an Arabic dialect.

Conclusion

The results of the influence of language show that language is relevant to response behavior and needs to be taken into account from a survey methodological point of view. This not only means that more research is to be done on this topic but also that the complexity of non-native-language surveys requires more attention in the practical implementation of surveys.

Even though one hypothesis could not be confirmed, and thus only a few practice-relevant statements on the effect of the language used in surveys can be made, the present paper shows that it can make a difference whether an item was answered in a mother tongue or in a non-native language. Therefore, the language used as a possible factor in response behavior should not be neglected, especially since an increasing number of surveys are offered in multiple languages.

Both the empirical results and the theoretical considerations suggest that it makes sense to offer surveys for respondents in their mother tongue. It enables people to participate who otherwise would not have been able to due to language barriers (e.g., Feskens et al., 2006; Jacobsen, 2018). Especially in surveys where, according to rational choice theory, there are hardly any incentives to take part, offering their mother tongue enables the respondents to participate with less cognitive load, which might positively influence the motivation and thus the data quality. Furthermore, offering surveys in more mother tongues might—which should be investigated additionally (Watson & Wooden, 2009, p. 165)—influence the general willingness to cooperate by showing respect for the respondent.

Data

Refugees in the German Educational System (2021). Raw data. Leibniz Institute for Educational Trajectories. Scientific-Use-File available via <https://doi.org/10.5157/ReGES:RC1:SUF:1.0.0> and <https://doi.org/10.5157/ReGES:RC2:SUF:1.0.0>

(Note: the data on which language was used for which item is currently not published for reasons of data protection. Therefore, the raw data was used. In the raw data there is one variable of which the recoding is also explained in the do-files.)

References

- Battistini, E., Miniaci, R., & Weber, G. (2003). What do we learn from recall consumption data? *The Journal of Human Resources*, 38(2), 354–385. <https://doi.org/10.2307/1558748>
- Blohm, M., & Diehl, C. (2001). Wenn Migranten Migranten befragen: Zum Teilnahmeverhalten von Einwanderern bei Bevölkerungsbefragungen. *Zeitschrift für Soziologie*, 30(3), 223–242. <https://doi.org/10.1515/zfsoz-2001-0304>
- Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–78). Bradford Books. <https://doi.org/10.7551/mitpress/4117.003.0010>
- Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly*, 55(1), 50–79. <https://doi.org/10.1086/269241>
- Calderón, J. L., Morales, L. S., Liu, H., & Hays, R. D. (2006). Variation in the readability of items within surveys. *American Journal of Medical Quality*, 21(1), 49–56. <https://doi.org/10.1177/1062860605283572>
- Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, 71(3), 193–204. <https://doi.org/10.1016/j.ijpsycho.2008.09.006>
- Cipolletti, H., McFarlane, S., & Weissglass, C. (2016). The moral foreign-language effect. *Philosophical Psychology*, 29(1), 23–40. <https://doi.org/10.1080/09515089.2014.993063>
- Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PLoS ONE*, 9(4), e94842. <https://doi.org/10.1371/journal.pone.0094842>
- Danziger, S., & Ward, R. (2010). Language changes implicit associations between ethnic groups and evaluation in bilinguals. *Psychological Science*, 21(6), 799–800. <https://doi.org/10.1177/0956797610371344>
- Dewaele, J.-M., & Nakano, S. (2012). Multilinguals' perceptions of feeling different when switching languages. *Journal of Multilingual and Multicultural Development*, 34(2), 107–120. <https://doi.org/10.1080/01434632.2012.712133>
- Díaz-Morales, J. F., Ferrari, J. R., Díaz, K., & Argumedo, D. (2006). Factorial structure of three procrastination scales with a Spanish adult population. *European Journal of Psychological Assessment*, 22(2), 132–137. <https://doi.org/10.1027/1015-5759.22.2.132>
- Dotinga, A., van den Eijnden, R. J. J. M., Bosveld, W., & Garretsen, H. F. L. (2005). The effect of data collection mode and ethnicity of interviewer on response rates and

- self-reported alcohol use among Turks and Moroccans in the Netherlands: An experimental study. *Alcohol and Alcoholism*, 40(3), 242–248. <https://doi.org/10.1093/alcalc/agh144>
- Esser, H. (1990). „Habits“, „Frames“ und „Rational Choice“. Die Reichweite von Theorien der rationalen Wahl (am Beispiel der Erklärung des Befragtenverhaltens). *Zeitschrift für Soziologie*, 19(4), 231–247. <https://doi.org/10.1515/zfsoz-1990-0401>
- Feskens, R., Hox, J., Lensvelt-Mulders, G., & Schmeets, H. (2006). Collecting data among ethnic minorities in an international perspective. *Field Methods*, 18(3), 284–304. <https://doi.org/10.1177/1525822X06288756>
- Garnham, A., Oakhill, J., von Stockhausen, L., & Sczesny, S. (2016). Editorial: Language, cognition, and gender. *Frontiers in Psychology*, 7, 772. <https://doi.org/10.3389/fpsyg.2016.00772>
- Geipel, J., Hadjichristidis, C., & Surian, L. (2015). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS ONE*, 10(7), e0131529. <https://doi.org/10.1371/journal.pone.0131529>
- Gentile, R., Heinritz, F., & Will, G. (2019). *Übersetzung von Instrumenten für die Befragung von Neuzugewanderten und Implementation einer audiobasierten Interviewdurchführung* (LifBi Working Paper No. 86). Leibniz Institute for Educational Trajectories. https://www.lifbi.de/Portals/2/Working%20Papers/WP_LXXXVI.pdf
- Gideon, M., Helppie-McFall, B., & Hsu, J. W. (2017). Heaping at round numbers on financial questions: The role of satisficing. *Survey Research Methods*, 11(2), 189–214. <https://doi.org/10.18148/srm/2017.v11i2.6782>
- Glantz, A., & Michael, T. (2014). Interviewereffekte. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 313–322). Springer Fachmedien. https://doi.org/10.1007/978-3-531-18939-0_21
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). John Wiley & Sons.
- Guarnaccia, P. J., Angel, R., & Worobey, J. L. (1989). The factor structure of the CES-D in the Hispanic health and nutrition examination survey: The influences of ethnicity, gender and language. *Social Science & Medicine*, 29(1), 85–94. [https://doi.org/10.1016/0277-9536\(89\)90131-7](https://doi.org/10.1016/0277-9536(89)90131-7)
- Hadjichristidis, C., Geipel, J., & Boaz, K. (2019). The influence of native language in shaping judgement and choice. *Progress in Brain Research*, 247, 253–272. <https://doi.org/10.1016/bs.pbr.2019.02.003>
- Hadjichristidis, C., Geipel, J., & Surian, L. (2019). Breaking magic: Foreign language suppresses superstition. *Quarterly Journal of Experimental Psychology*, 72(1), 18–28. <https://doi.org/10.1080/17470218.2017.137178>
- Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24(4), 561–579. <https://doi.org/10.1017/S0142716403000286>
- Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 257–283). Multilingual Matters. <https://doi.org/10.21832/9781853598746-012>
- Harzing, A.-W., & Maznevski, M. (2002). The interaction between language and culture: A test of the cultural accommodation hypothesis in seven countries. *Language and Intercultural Communication*, 2(2), 120–139. <https://doi.org/10.1080/14708470208668081>

- Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage*, 15(3), 647–660. <https://doi.org/10.1006/nimg.2001.1001>
- Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition*, 173, 8–15. <https://doi.org/10.1016/j.cognition.2017.12.010>
- Heinritz, F., & Will, G. (2021). *Selektive Teilnahme von Geflüchteten an der Panelstudie ReGES* (LIlBi Working Paper No. 96). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/LIlBi:WP96:1.0>
- Holbrook, A. L., Anand, S., Johnson, T. P., Cho, Y. I., Shavit, S., Chávez, N., & Weiner, S. (2014). Response heaping in interviewer-administered surveys: Is it really a form of satisficing? *Public Opinion Quarterly*, 78(3), 591–633. <https://doi.org/10.1093/poq/nfu017>
- Hunt, S. M., & Bhopal, R. (2004). Self report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture. *Journal of Epidemiology & Community Health*, 58(7), 618–622. <https://doi.org/10.1136/jech.2003.010074>
- International Labour Office. (2010). *International labor migration: A rights-based approach*. International Labour Organization. https://webapps.ilo.org/public/libdoc/ilo/2010/110B09_59_engl.pdf
- International Labour Office. (2018). *Ilo global estimates on international migrant workers: Results and methodology*. International Labour Organization. <https://www.onlinelibrary.iihl.org/wp-content/uploads/2020/05/2018-I1-2.pdf>
- International Organization for Migration. (2019). *World migration report 2020*. International Organization for Migration. https://publications.iom.int/system/files/pdf/wmr_2020.pdf
- Jacobsen, J. (2018). Language barriers during the fieldwork of the IAB-BAMF-SOEP survey of refugees in Germany. In D. Behr (Ed.), *Surveying the migrant population: Consideration of linguistic and cultural issues* (pp. 75–84). GESIS – Leibniz-Institut für Sozialwissenschaften. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-58536-7>
- Kappelhof, J. (2017). Survey research and the quality of survey data among ethnic minorities. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, & B. T. West (Eds.), *Total survey error in practice* (pp. 235–252). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119041702.ch11>
- Kinnunen, J. M., Malin, M., Raisamo, S. U., Lindfors, P. L., Pere, L. A., & Rimpelä, A. H. (2015). Feasibility of using a multilingual web survey in studying the health of ethnic minority youth. *JMIR Research Protocols*, 4(2), e53. <https://doi.org/10.2196/resprot.3655>
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. <https://doi.org/10.1002/acp.2350050305>
- Lee, T. (2002). *Language-of-interview effects and Latino mass opinion*. SSRN. <https://doi.org/10.2139/ssrn.303165>
- Loosveldt, G. (2008). Face-to-face interviews. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 201–220). Taylor & Francis Group/Lawrence Erlbaum Associates.
- Marian, V., & Fausey, C. M. (2006). Language-dependent memory in bilingual learning. *Applied Cognitive Psychology*, 20(8), 1025–1047. <https://doi.org/10.1002/acp.1242>
- Marian, V., & Kaushanskaya, M. (2004). Self-construal and emotion in bicultural bilinguals. *Journal of Memory and Language*, 51(2), 190–201. <https://doi.org/10.1016/j.jml.2004.04.003>

- Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, 129(3), 361–368. <https://doi.org/10.1037/0096-3445.129.3.361>
- Marin, G., Triandis, H. C., Betancourt, H., & Kashima, Y. (1983). Ethnic affirmation versus social desirability. *Journal of Cross-Cultural Psychology*, 14(2), 173–186. <https://doi.org/10.1177/0022002183014002003>
- McKay, R. B., Breslow, M. J., Sangster, R. L., Gabbard, S. M., Reynolds, R. W., Nakamoto, J. M., & Tarnai, J. (1996). Translating survey questionnaires: Lessons learned. *New Directions for Evaluation*, 70, 93–104. <https://doi.org/10.1002/ev.1037>
- Neske, M., & Rich, A.-K. (2016). *Asylerstantragsteller in Deutschland im ersten Halbjahr 2016: Sozialstruktur, Qualifikationsniveau und Berufstätigkeit* (BAMF-Kurzanalyse 4-2016). Bundesamt für Migration und Flüchtlinge. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-67527-7>
- Nichols, S., Kumar, S., Lopez, T., Ayars, A., & Chan, H.-Y. (2016). Rational learners and moral rules. *Mind & Language*, 31(5), 530–554. <https://doi.org/10.1111/mila.12119>
- Paas, F., & Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351–371. <https://doi.org/10.1007/BF02213420>
- Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8
- Pan, Y., Leeman, J., Fond, M., & Goerman, P. (2014). *Multilingual survey design and fielding: Research perspectives from the U.S. census bureau* (Research Report Series 2014-01). U.S. Census Bureau. <https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/rsm2014-01.pdf>
- Pérez, E. O. (2009). Lost in translation? Item validity in bilingual political surveys. *The Journal of Politics*, 71(4), 1530–1548. <https://doi.org/10.1017/S0022381609990156>
- Peytcheva, E. (2018). Can the language of survey administration influence respondents' answers? In T. P. Johnson, B.-E. Pennell, I. A. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts* (pp. 325–340). John Wiley & Sons, Inc.
- Rich, A.-K. (2016). *Asylerstantragsteller in Deutschland im Jahr 2015: Sozialstruktur, Qualifikationsniveau und Berufstätigkeit* (BAMF-Kurzanalyse 3-2016). Bundesamt für Migration und Flüchtlinge. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-67504-2>
- Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., Johnston, M., Vickers, L., Yan, H. Y., & Zhang, C. (2015). Precision and disclosure in text and voice interviews on smartphones. *PLoS ONE*, 10(6), e0128337. <https://doi.org/10.1371/journal.pone.0128337>
- Schoua-Glusberg, A. (2004). *Assessing comprehension of translated questionnaires with qualitative methods* (Statistics Canada International Symposium Series – Proceedings No. 11-522-X20040018746). Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/11-522-x/2004001/8746-eng.pdf>
- Shin, H. I., & Kim, J. (2017). Foreign language effect and psychological distance. *Journal of Psycholinguistic Research*, 46(6), 1339–1352. <https://doi.org/10.1007/s10936-017-9498-7>
- Steinhauer, H.-W., Zinn, S., & Will, G. (2019). Sampling refugees for an educational longitudinal survey. *Survey Methods: Insights from the Field*. Advance online publication. <https://doi.org/10.13094/SMIF-2019-00007>

- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). The National Academic Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–83. <https://doi.org/10.1037/0033-2909.133.5.859>
- von Maurice, J., & Will, G. (2023). Data from the panel study ‘Refugees in the German Educational System (ReGES)’. *Journal of Open Psychology Data*, 11(1), 1–15 <https://doi.org/10.5334/jopd.77>
- Watson, N., & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 157–81). Wiley. <https://doi.org/10.1002/9780470743874.ch10>
- Welch, S., Comer, J., & Steinman, M. (1973). Interviewing in a Mexican-American community: An investigation of some potential sources of response bias. *Public Opinion Quarterly*, 37(1), 115–126. <https://doi.org/10.1086/268065>
- Will, G., Balaban, E., Dröschner, E., Homuth, C., & Welker, J. (2018). *Integration von Flüchtlingen in Deutschland: Erste Ergebnisse aus der ReGES-Studie* (LIIfBi Working Paper No 76). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/LIIfBi:WP76:2.0>
- Will, G., Homuth, C., von Maurice, J., & Roßbach, H.-G. (2021). Integration of recently arrived underage refugees: Research potential of the study ReGES—Refugees in the German Educational System. *European Sociological Review*, 37(6), 1027–1043. <https://doi.org/10.1093/esr/jcab033>

Appendix

Appendix 1 Characteristics of control variables on the level of respondents

	M/Freq.	SD	Min.	Max.
Country of origin				
<i>Afghanistan</i>	6.38 %			
<i>Iraq</i>	8.69 %			
<i>Iran</i>	1.82 %			
<i>Syria</i>	77.05 %			
<i>Other</i>	6.06 %			
Age	27.84	13.90	14	75
Length of stay in Germany	28.81	9.20	3	53
Gender				
<i>Male</i>	62.47 %			
<i>Female</i>	37.53 %			
Education				
<i>ISCED 0: Preprimary education</i>	4.74 %			
<i>ISCED 1: Primary Education</i>	20.48 %			
<i>ISCED 2: Lower secondary education</i>	11.26 %			
<i>ISCED 3: Upper secondary education</i>	17.05 %			
<i>ISCED 4: Postsecondary nontertiary education</i>	4.24 %			
<i>ISCED 5: Short-cycle tertiary education</i>	6.65 %			
<i>ISCED 6: Bachelor or equivalent</i>	7.77 %			
<i>ISCED 7: Master of equivalent</i>	8.10 %			
<i>ISCED 8: Doctoral or equivalent</i>	0.16 %			
<i>Missing values</i>	19.57 %			

Source: ReGES data, own calculations, Wave 1, *n* = 1,865 respondents.

Appendix 2 Items used for measuring accuracy

Name	Question
p3100000	The following questions are about your living situation in Germany now. In how many different accommodation facilities have you lived since your arrival in Germany? Please list all stations from the preliminary reception center to your current accommodation.
p3241140	On average, how many hours does your child spend at the childcare facility per week?
p6242120	How many hours of German language classes does your child attend at preschool per week?
p6242140	How many hours of German language classes does your child attend outside of his or her preschool per week?
p3140000	The following questions are about your child's living situation in Germany. In how many different accommodation facilities has your child lived since his or her arrival in Germany? Please list all stations from the preliminary reception center to your current accommodation.
p6242220	How many hours of German language classes does your child attend per week?
p3241180	On average, how many hours per week does your child spend with a childminder or nanny?
p6242410	On a normal weekday, how many hours does your child spend in situations where he or she hears or speaks German?
p3241250	On average, how many hours did your child spend at the childcare facility per week?
p3241300	On average, how many hours per week did your child spend with a childminder or nanny?
t6242220	For how many hours a week do you take German classes for refugees and migrants at school?
t6242240	For how many hours a week do you take German classes for refugees and migrants outside of school?
t6242420	On a normal weekday, how many hours do you spend in situations where you hear, speak, read or write German?

Source: ReGES, parents- and adolescent questionnaires, Wave 1.

Appendix 3 Multivariate analyses, separately for adolescents and parents

Dependent variable	Hypothesis 1		Hypothesis 2		Hypothesis 3 (gender roles)	
	Duration in seconds		Accuracy (rounded = 1)		Gender roles, 4 (egalitarian) to 16 (traditional)	
	Model 1.1	Model 1.2	Model 2.1	Model 2.2	Model 3.1	Model 3.2
	Parents	Adolescents	Parents	Adolescents	Parents	Adolescents
Item in mother tongue	-2.28** (0.80)	-2.87* (1.20)	0.90 (0.24)	2.12 (1.06)	1.72*** (0.32)	-0,56*** (0.41)
<i>Controls:</i>						
Country of origin	✓	✓	✓	✓	✓	✓
Age	✓	✓	✓	✓	✓	✓
Length of stay in Germany	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓
Translation	✓	✓	✓	✓		
Length of text	✓	✓				
Pseudo/ Adjusted R^2	.015	.008	.034	.036	.052	.037
<i>N</i>	99,069	69,390	1,903	764	729	414

Source: ReGES data, own calculations, Wave 1.

Notes: Estimates with standard errors in parentheses.

* $p < .05$,

** $p < .01$,

*** $p < .001$.

Appendix 4 Ordinal logistic regression for religiosity (average marginal effects, Model 2), separately for adolescents and parents

	Parents	Adolescents
Item only in mother tongue		
<i>Not at all religious</i>	-0.05** (0.02)	0.00 (0.02)
<i>Not very religious</i>	-0.11*** (0.03)	0.00 (0.03)
<i>Quite religious</i>	0.12** (0.04)	0.00 (0.03)
<i>Very religious</i>	0.04*** (0.01)	0.00 (0.02)
Controls:		
Age, length of stay in Germany, gender, education	✓	✓
Adjusted R^2	.025	.023
N	848	529

Source: ReGES data, own calculations, Wave 1.

Notes: Estimates with standard errors in parentheses.

* $p < .05$,

** $p < .01$,

*** $p < .001$.