

Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues

Behr, Dorothée (Ed.)

Veröffentlichungsversion / Published Version
Konferenzband / collection

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Behr, D. (Ed.). (2018). *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (GESIS-Schriftenreihe, 19). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.58074>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nc/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see:
<https://creativecommons.org/licenses/by-nc/4.0>

gesis

Leibniz-Institut
für Sozialwissenschaften

Schriftenreihe

Band 19

Dorothee Behr (Ed.)

Surveying the Migrant Population

Consideration of Linguistic and
Cultural Issues

Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues

GESIS Series

published by GESIS – Leibniz Institute for the Social Sciences

Volume 19

edited by Dorothée Behr

Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues

Dorothee Behr (Ed.)

Surveying the Migrant Population

Consideration of Linguistic and Cultural Issues

Bibliographical information of the German National Library (DNB)

The German National Library lists this publication in the German National Bibliography; detailed bibliographical data are available via <http://dnb.ddb.de>.

ISBN 978-3-86819-030-4 (eBook)

ISSN 1869-2869

Publisher, printing

and distribution: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln, Tel.: 0221 / 476 94 - 0
info@gesis.org
Printed in Germany

Preface

The papers in this volume were almost all – with the exception of the paper by Comanaru and d’Ardenne – presented at the GESIS Symposium on “Surveying the Migrant Population: Consideration of Linguistic and Cultural Aspects”, which took place in Mannheim March 14-15, 2017. The editor of this volume would like to thank Prof. Dr. Michael Braun and Klara Raiber, both GESIS – Leibniz Institute for the Social Sciences, for their great support in producing and finalizing this volume.

Table of Content

Preface	1
Linguistic and Cultural Aspects in Migrant Surveys Introduction and Overview <i>Dorothee Behr, Patrick Brzoska, & Alisú Schoua-Glusberg</i>	5
The Development of a Research Programme to Translate and Test the Personal Well-being Questions in Sylheti and Urdu <i>Ruxandra Comanaru & Jo d'Ardenne</i>	13
The Place of Expert Review in Translation and Questionnaire Evaluation for Hard-to-Count Populations in National Surveys <i>Patricia Goerman, Mikelyn Meyers, & Yazmín García Trejo</i>	29
Measuring Migrants' Educational Attainment: The CAMCES Tool in the IAB-SOEP Migration Samples <i>Silke L. Schneider, Roberto Briceno-Rosas, Verena Ortmanns, & Jessica M. E. Herzing</i> .	43
Language Barriers during the Fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany <i>Jannes Jacobsen</i>	75
Surveying Migrants in the Context of the Low-income Panel PASS <i>Birgit Jesske</i>	85
Surveying Immigrants. The Role of Language Attrition and Language Change in the Application of Questionnaires <i>Patrick Brzoska</i>	95
Contributing Authors	107

Linguistic and Cultural Aspects in Migrant Surveys

Introduction and Overview

Dorothee Behr, Patrick Brzoska, & Alisú Schoua-Glusberg

1 Migration Research

Migration is a global phenomenon; the numbers of migrants¹ across the globe have risen rapidly in recent decades (Font & Méndez, 2013). The current situation in the world, in particular in war-torn countries such as Syria, is only pushing this trend up even further. In Germany, as a case in point, the European “refugee crisis” translated into about 476,000 applications for asylum in 2015 and about 745,000 applications in 2016 (Bundesamt für Migration und Flüchtlinge, 2017). With the increasing number of people immigrating into host countries and the rising numbers of those being granted asylum, societies need to better understand the processes of acculturation in order to steer and monitor integration with the ultimate aim of ensuring a peaceful and sustainable cohabitation between migrants and the host population. In addition, governments and public institutions need to understand migrants’ needs in order to adjust their services adequately. Survey data play an important role in this regard. Such data provide the basis for understanding migrants’ realities of life and for devising, evaluating, and potentially revising public (integration) policies and services. The larger the cultural distance between migrants and the majority population of the host countries, the more efforts may be needed in order to understand the respective needs and concerns and to respond adequately (Font & Méndez, 2013).

2 Methodological Challenges

Migration research has to overcome many methodological challenges; these have recently been discussed in quite some detail (e.g., *Hard-to-survey populations*, Tourangeau, Edwards, Johnson, Wolter, & Bates, 2014; *Surveying ethnic minorities and immigrant populations: Methodological challenges and research strategies*, Font & Méndez, 2013; *Methoden der Migrationsforschung: ein interdisziplinärer Forschungsleitfaden*, Maehler & Brinkmann, 2015; *Handbuch Diagnostische Verfahren für Migrantinnen und Migranten*, Maehler, Shajek, & Brinkmann, 2018). Particular attention in these books is directed towards the sam-

1 The term migrants shall be used in an encompassing sense, including *refugees and asylum seekers* (forced displacement) as well as *voluntary migrants*, whether moving to another country for a short time (sojourner) or for good. Furthermore, it includes *established migrants*, who possibly even grew up in the host country, and *new arriving migrants*.

pling of migrants. However, also questionnaire translation draws increased attention. This obviously results from the fact that those not mastering the language of a country² need to be surveyed in a language they are proficient in. Translation and, as an extension, linguistic aspects during fieldwork will be the main focus of these proceedings.

Producing a questionnaire for migrants shares many features with producing questionnaires in *cross-national* survey research (which usually also comprises a cross-cultural perspective). The particularities of cross-national research and its impact on data comparability have been summarized by leading scholars in the field – see in particular the chapters on questionnaire design and translation in *Survey methods in multicultural, multinational, and multiregional contexts* (Harkness et al., 2010), *Cross-cultural survey methods* (Harkness, van de Vijver, & Mohler, 2003) or the *Cross-cultural survey guidelines* (Survey Research Center, 2016). The precondition for producing comparable cross-national survey instruments is to thoroughly take into account the respective country-specific cultures, both the “hard” culture, for instance in terms of institutions, and the “soft” (van de Vijver & Leung, 2011) culture, for instance in terms of behavior, norms or values. Migration research adds to the complexity of cross-national research by blurring the cultural boundaries: We no longer juxtapose respondents from different countries and cultures, but request survey information from respondents possibly raised in one cultural setting but now being embedded in another cultural setting. Of course, the latter comes in many shades: Migrants may have been born in the host country or may have come later, and thus they may be more or less familiar with the cultural setting of the host country. Added to the two interdependent cultural frames are the two languages that are typically brought to a survey targeted at migrants, that is, the first language of migrants (usually their ‘mother tongue’) and, to some extent at least, the language of the host country. The blurring of linguistic and cultural boundaries makes migration research different from cross-national research.³ Hence, approaches for migration research may partly deviate from what is known in cross-national survey methodology.

In the following, key issues will be addressed which need to be tackled in order to survey migrants in the appropriate way – both culturally and linguistically. In the following, the term “host country” will be used for the country that has received the migrants, the term “country of origin” will be used to refer to the country the migrants originally came from (at least 1st generation migrants).

2 To simply matters, we only speak of ‘the language of a country’ throughout the chapter, even though there may be several official or main languages used in a country.

3 Other types of cross-cultural or multilingual research not further discussed here are within-country research that includes well established and acculturated linguistic minorities or multilingual research in the different official of main languages of a country.

2.1 The Target Population

As with every survey, the target group needs to be defined clearly. This is even more crucial with migrants because the exact definition of the target group determines the language version and any adaptations to content – and possibly format – of the questionnaire. Questions to be asked are: How long has the target population been living in the host country? Do researchers plan to survey “more established migrants”, that is, those who have lived in the host countries for several decades and often also for many generations, or more recent migrants, such as refugees and asylum seekers? These questions determine whether the survey can take place in the language of the host country or whether it should take place in a language the respondent is proficient in (e.g., the language of the country of origin or the language of an ethnic group). As the chapters by Jacobsen and Jesske in this volume illustrate, pragmatic fieldwork decisions may even pave the way for a bi-lingual survey administration (where respondents can decide for themselves which language version of a questionnaire they prefer). The length of stay in a country can indicate familiarity with the socio-cultural context of that country and hence with topics that can be addressed meaningfully and without offense (e.g., without breaking taboos). In addition, the origin of the target population will have an impact on whether surveys and their sometimes peculiar ‘survey speak’ is known to respondents. Formea et al. (2014), for instance, report on problems related to Likert scales and their possible replacement or extension with graphics or percentages for the Somali population in Rochester, Minnesota (US). Last but not least, knowledge of the target population’s literacy level is needed in order to decide on the mode of the survey and whether interviewers will be needed to conduct the interviews.

2.2 Language Particularities

If researchers opt for using the language of the host country, care needs to be taken to make it suitable in style and formality for the level of language proficiency of the target group: While the wording of general population surveys often cannot be adapted, surveys particularly targeted at the migrant population can be improved, for instance, by taking on board criteria of plain or simple language⁴ – even though surveys in general should, of course, always be written in as simple a language as possible.

If researchers decide on providing a translation (often in the migrant’s first language), it is far from sufficient to choose a language without further specifications on the target population. There are, for instance, many different varieties of the Arabic language. *Diglossia* – that is, the existence of markedly different language variants for different social situations (e.g. informal and formal) – and dialects depending on the country or region where respondents come from will play a role as to which “Arabic” should be used in a given survey (see also Al-Sobh, Abu-Melhim, & Bani-Hani, 2015, for more information on the Arabic language). Surveys for migrants in Spanish face similar, even though possibly not as pronounced challenges. In the US, as a case in point, official surveys in Spanish strive

4 For German, please see: <http://research.uni-leipzig.de/leisa/de/> (16 October, 2017); for English, <http://www.plainlanguage.gov/> (19 January, 2018)

to match the diverse Hispanic backgrounds of US immigrants, which may be achieved with a Spanish-language common version that allows for country-specific terminologies where differences cannot be surmounted (Martinez, Marin, & Schoua-Glusberg, 2006).

For migration research, there may be instances where a *strict* application of just one language is not advisable. A particularity of migration research exists in the fact that institutions, services, and other elements bound to the host country often cannot be translated, since there may often not be a corresponding equivalent available. Moreover, migrants are likely to be more familiar with their original name than with a more or less equivalent term in translation. Consequently, some terms (e.g., *Gesundheitsamt* – a local public health authority in Germany) may stay in the original language – and possibly be supplemented with a translation or a paraphrase in the migrant language (e.g., Carrasco, 2003; Sauer, 2008; Schenk, Ellert, & Neuhauser, 2007).

When using well-established instruments for which an official translation (and validation) already exists, researchers should carefully consider whether this official translation, as used in the country of origin, is linguistically and culturally suitable also for the migrant population in the host country. Brzoska, for instance, calls for carefully vetting a Turkish translation produced for Turks in Turkey before using it for Turkish migrants in Germany, especially if those have been living in Germany for a long time (chapter in this volume; see also Brzoska, Yilmaz-Aslan, & Razum, 2013, and Brzoska, 2014).

Finally, some languages may not exist in a written form – a use of such languages may then pose challenges for standardizing the instrument. In this volume, Comanaru and d'Ardenne illustrate based on Sylheti, a spoken language in Bangladesh, how researchers may approach the translation task in such cases.

2.3 Questionnaire Particularities

Often, a questionnaire is developed in the language of the researchers before it is translated into the language(s) of the targeted migrant population(s). It is paramount that the different cultural contexts of migrants are taken into account already during questionnaire design, that is, when selecting concepts, indicators, and wordings for individual questions. For instance, some topics may be perceived as taboo topics by migrants and may likely not lead to an honest answer, if an answer is provided at all. As a case in point, questions about whether a child was adopted proved problematic (even offensive) among Korean respondents in the US, since these questions run counter to blood ties being regarded as essential to family organization in the Korean culture (Pan & Lubkemann, 2014). Other topics may assume previous knowledge on a topic that cannot be taken for granted or they may assume, especially in health settings, symptoms that may not be of importance for migrant groups (Glaesmer, Brähler, & von Lersener, 2012). And other topics, again, may require clear instructions and clarity on whether the country of origin or the host country should be referred to. Education questions may be a typical example of questions taking into consideration educational paths in the system of the country of origin, while the rest of the questionnaire may refer to the host country. The particularities involved in measuring education of migrants are discussed by Schneider, Briceno-Rosas, Ortmanns, and Herzing in this volume. Several checklists may help to raise awareness of cultural or

linguistic traps that should be avoided. Schoua-Glusberg’s Cultural Sensitivity Checklist for Instruments (2017) may serve as a good starting point for developers (see Table 1). It should ideally be worked through in close cooperation with native co-researchers, consultants, translators or interpreters.

Table 1 Schoua-Glusberg’s Cultural Sensitivity Checklist for Instruments (slightly adapted from Schoua-Glusberg, 2017, p. 14)

<ul style="list-style-type: none"> ▪ Is there anything in the questions or the style of asking that might be poorly received or even offensive? ▪ Do any questions sound strange for the study population context? ▪ Would any questions be difficult for the study population to answer? Why? ▪ Are any questions NOT likely to elicit a full answer? If so, why? ▪ Are there questions that may not yield usable information or collect the intended information? ▪ If data will be collected in a group setting, such as a focus group, are any questions too sensitive to be answered in front of others? ▪ Are any questions uncomfortable for interviewers to ask? ▪ Is the introduction/consent/explanation of the purpose of the interview presented in the best order for the local discourse style?

Other attempts to streamline the problem-spotting process for cross-cultural questionnaires include the revised Questionnaire Appraisal System (QAS-04) that has been enriched by codes for cross-cultural consideration and translation problems (Dean, Caspar, McAvinchey, Reed, & Quiroz, 2007). The earlier problems can be identified (best if before or during the design of the questionnaire), the easier it becomes to remedy those by reconceptualising, rewording or providing background information for teams on adaptation possibilities.

2.4 Translation and Pretesting

Best practice in instrument translation calls for team or committee approaches that bring together translators, researchers, social workers, or other experts – that is, those persons who possess in-depth linguistic and cultural knowledge of the target group, translational expertise, and knowledge of the measurement instrument. The team approach is described in Mohler, Dorer, de Jong, & Hu (2016) in detail. Experience shows that briefing and training on questionnaire particularities is particularly important when translators have not had prior experience with questionnaires (e.g., training on the importance of simple and accessible language, consistency of re-occurring elements or ‘scale speak’). Pretesting is an indispensable tool for testing the comprehensibility, the cultural applicability, and the general flow of the questionnaire, thereby finding answers to the following questions: Are concepts familiar to respondents? Are frames of reference clear (e.g., reference to host

country or country of origin)? Have an adequate dialect and formality level been chosen (which may be of utmost relevance for languages where different regional or formal/informal varieties exist)? Even though pretesting is a core element of the best practice translation model TRAPD – translation, review, adjudication, pretesting, and documentation (Harkness, 2003) – resource constraints in projects may sometimes prevent its implementation. Goerman, Meyers, and Garcia Trejo, in this volume, provide an alternative solution for ensuring a best possible translation when financial or budget constraints prevent the implementation of a pretest.

2.5 Fieldwork

Fieldwork for migrant groups cannot necessarily be handled in the same way as it is handled for the majority population who speaks the language of the host country. The challenge of finding bilingual interviewers is only one of many. Jacobsen and Jesske in this volume present examples of how to conduct interviews when bilingual professional interviewers for face-to-face interviews are not available. These authors further illustrate what it may mean to deviate from standardized interviewing procedures, for instance by allowing switching between language versions, having interviewers help respondents make their way through the questionnaire, or enlisting the help of an (informal) interpreter. In the same vein, that is, the matter of de-standardizing survey interviews for a migrant population, Pan and Lubkemann (2014) have listed numerous culture-specific examples of interviewers applying culture-specific conversational norms in order to gain access to respondents in the first place and keep up the interview until the end. These latter forms of cultural adaptations mirror the last point of Schoua-Glusberg's sensitivity list (2017), namely, whether local discourse norms have been respected when introducing and explaining the interview to the respondent – and this may also apply to written discourse. Ethnic or gender matching are additional concerns that have been discussed elsewhere (Font & Méndez, 2013).

3 Conclusions and Outlook

Planning and conducting research on migrant groups entails many additional steps beyond the monocultural survey life cycle. These steps do have a major impact on costs, timeline, and persons involved (e.g., Font & Méndez, 2013; Formea et al., 2014), which is why they should be thoroughly considered when planning a survey. Researchers need to ensure, amongst others, the cultural applicability and relevance of the measurement instrument, the language fit for the target population, and the cultural appropriateness of the interview (situation). Particularly in Germany since 2015, migration research has reached new levels with the increased need to understand, assess, and consider migrants' needs as well as to inform, revise and evaluate stakeholders' strategies, policies or programs. In this situation, it is paramount for researchers to be aware of and willing to deal with cultural and linguistic challenges. The present proceedings aim to contribute towards this awareness

and encourage researchers to a) critically reflect on their survey strategies, b) to consult the available methodological literature on cross-cultural survey implementation (literature from the US is a prime example), and c) to share their experiences by documenting and publishing own procedures and lessons learned. Subsequent researchers will then be able to draw on existing information. We stand to learn a lot with every survey, and the more open and transparent we are with our own processes that worked but also with those that failed, the more we can contribute to the quality of subsequent studies.

References

- Al-Sobh, M. A., Abu-Melhim, A. H., & Bani-Hani, N. A. (2015). Diglossia as a result of language variation in Arabic: Possible solutions in light of language planning. *Journal of Language Teaching and Research*, 6(2), 274–279.
- Brzoska, P., Yilmaz-Aslan, Y., & Razum, O. (2013). Can we survey migrants by questionnaires developed for the population of the respective source countries? Paper presented at ESRA 2013 (Ljubljana). Retrieved from: <http://www.europeansurveyresearch.org/conference/programme?sess=152>
- Brzoska, P. (2014). Psychometrically relevant differences between source and migrant populations. Frankfurt am Main: Peter Lang.
- Bundesamt für Migration und Flüchtlinge (2017). *Aktuelle Zahlen zu Asyl. Ausgabe Dezember 2017*. Nürnberg. Retrieved from <http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Statistik/Asyl/aktuelle-zahlen-zu-asyl-dezember-2017.html?nn=7952222>
- Carrasco, L. (2003). The American Community Survey (ACS) en español: Using cognitive interviews to test the functional equivalency of questionnaire translations. *STUDY SERIES (Survey Methodology #2003–17)*.
- Dean, E., Caspar, R., McAvinchey, G., Reed, L., & Quiroz, R. (2007). Developing a low-cost technique for parallel cross-cultural instrument development: The Question Appraisal System (QAS-04). *International Journal of Social Research Methodology*, 10(3), 227–241.
- Font, J., & Mendez, M. (Eds.) (2013). *Surveying ethnic minorities and immigrant populations: Methodological challenges and research strategies*. Amsterdam: Amsterdam University Press.
- Formea, C. M., Mohamed, A. A., Hassan, A., Osman, A., Weis, J. A., Sia, I. G., & Wieland, M. L. (2014). Lessons learned: Cultural and linguistic enhancement of surveys through community-based participatory research. *Progress in Community Health Partnerships: Research, Education, and Action*, 8(3), 331–336.
- Glaesmer, H., Brähler, E., & von Lersener, U. (2012). Kultursensible Diagnostik in Forschung und Praxis. Stand des Wissens und Entwicklungspotenziale. *Psychotherapeut*, 57, 22–28.
- Harkness, J. A., van de Vijver, F. J. R. & Mohler, P. Ph. (Eds.) (2003). *Cross-cultural survey methods*. Hoboken, NJ: Wiley.

- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. van de Vijver & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: John Wiley & Sons.
- Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L., Mohler, P. Ph. ... Smith, T. W. (Eds.) (2010). *Survey methods in multinational, multiregional, and multicultural contexts*. Hoboken, NJ: Wiley.
- Maehler, D., & Brinkmann, H. U. (Eds.) (2015). *Methoden der Migrationsforschung: ein interdisziplinärer Forschungsleitfaden*. Wiesbaden: Springer.
- Maehler, D. B., Shajek, A., & Brinkmann, H. U. (2018). *Handbuch Diagnostische Verfahren für Migrantinnen und Migranten*. Göttingen: Hogrefe.
- Martinez, M., Marin V., & Schoua-Glusberg, A. (2006). Translating from English to Spanish: The 2002 National Survey of Family Growth. *Hispanic Journal of Behavioral Sciences*, 28(4), 531–545.
- Mohler, P., Dorer, B., de Jong, J., & Hu, M. (2016). Translation: Overview. *guidelines for best practice in cross-cultural surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from <http://www.ccsr.isr.umich.edu/>
- Pan, Y., & Lubkemann, S. (2014). Standardization and meaning in the survey of linguistically diversified populations: Insights from the ethnographic observation of linguistic minorities in 2010 Census interviews. In Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M. & Bates, N. (Eds.), *Hard-to-survey populations* (pp. 575–589). Cambridge, UK: Cambridge University Press.
- Sauer, M. (2008). Methodik und Durchführung von standardisierten Befragungen türkeistämmiger Migranten in Deutschland durch die Stiftung Zentrum für Türkeistudien. In K.-S. Rehberg & Deutsche Gesellschaft für Soziologie (DGS) (Eds.), *Die Natur der Gesellschaft: Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006* (pp. 3901–3907). Frankfurt am Main: Campus Verl.
- Schenk, L., Ellert, U., & Neuhauser, H. (2007). Kinder und Jugendliche mit Migrationshintergrund in Deutschland. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 50, 590–599.
- Schoua-Glusberg, A. (2017, March). Issues in designing instruments and materials for immigrants and migrants in the US. Paper presented at the GESIS Symposium on “Surveying the migrant population: Consideration of linguistic and cultural aspects”, Mannheim, Germany.
- Survey Research Center (2010). *Guidelines for best practice in cross-cultural surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from <http://www.ccsr.isr.umich.edu/>
- Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M., & Bates, N. (Eds.) (2014). *Hard-to-survey populations*. Cambridge, UK: Cambridge University Press.
- Van de Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 17–45). New York, NY: Cambridge University Press.

The Development of a Research Programme to Translate and Test the Personal Well-being Questions in Sylheti and Urdu

Ruxandra Comanaru & Jo d'Ardenne

Abstract

A pilot research programme was undertaken in order to translate and cognitively test the personal well-being questions developed by the Office for National Statistics (ONS) in the UK into Urdu and Sylheti. These are both complex languages spoken by minority ethnic groups in the UK; monolingual speakers of these languages have been identified as being at risk of lower general health, and thus potentially of a lower well-being than the general UK population. The research programme involved two key stages in the production of these translations: translation workshops and cognitive testing of the translations. The translation workshops brought together experts in questionnaire development and personal well-being with native speakers of the languages from the community and bilingual interviewers in order to attempt to underpin the essence of the four personal well-being measures and arrive at the translation to be tested. The cognitive interviews that followed assessed these translations and the level of appropriateness of these measures with monolingual speakers in the UK. This article discusses the merits of each stage of this research programme in arriving at the best and most suitable translation for the personal well-being questions in Sylheti and Urdu.

1 Background

A pilot research programme was designed to translate and cognitively test the personal well-being questions developed by the Office for National Statistics (ONS) into Urdu and Sylheti. The main reason for the development of this programme for these particular languages was that monolingual speakers of these languages in the UK have been identified as being at risk of lower general health (Office for National Statistics, 2013a), and thus potentially of a lower well-being due to lower levels of integration into mainstream society and thus lower access to services. A team was brought together to develop and implement this research programme. This team included researchers with extensive expertise in the personal well-being indicators in the UK, questionnaire design experts, bilingual researchers, and bilingual interviewers. The project identified two pivotal stages: a translation workshop and a pretesting stage, using cognitive interviews with monolingual speakers of

Urdu and Sylheti, respectively. Furthermore, this collaborative iterative method of translating the well-being measures led to the identification of unforeseen challenges, but also allowed for their quick resolution.

2 Harmonisation of Questions in the UK

The UK has a number of government-led surveys on various topics, including the Census of Population; they are a rich source of social and economic information. However, these surveys were designed at different times, they were developed to serve different purposes, they were also commissioned by different departments, and thus they were developed mostly in isolation from each other. Hence, they often use different questions to investigate the same concepts, thus making the data obtained difficult to compare. The topics covered in these surveys, and therefore the official national statistics available, include income, expenditure, food, health, housing, transport, and many more.

A cross-governmental programme has been implemented to harmonise the measures used across the different government-led surveys. This process is known as *harmonisation* (Office for National Statistics, 2011). For several years now, ONS has led this programme of work, which has the aim of simplifying the use of survey measurements, such that users can draw clearer and more robust comparisons between data sources.

2.1 Well-being Questions

Well-being is a complex psychological concept that has been the focus of many research studies recently. In simple terms, well-being can be defined as “optimal psychological functioning and experience” (Ryan & Deci, 2001, p.142). Measurements of this concept can have major implications for policies at the local and national level; for example, more resources can be allocated locally and nationally for sectors of the population that are found to be experiencing a lesser well-being, and new policies can be implemented. Monitoring well-being over time can show which of these measures and policies have a positive effect. ONS has undertaken a programme of study and research to define the most important features that constitute well-being, and to develop means of testing and monitoring those over time at a national level. In 2010, ONS started a national programme to track the well-being of the UK population. Forty-one indicators have been identified which constitute the dimensions of an individual’s well-being. These include: relationships, health, finance, economy, education, and others. One of these indicators is personal (or subjective) well-being, underpinned by three constructs:

- evaluation (an overall assessment of satisfaction with life);
- eudaimonia (often referred to as the flourishing measure; the overall perception of how much purpose one’s life has, that is, how worthwhile the things are that one does in his/her life); and

- experience (a snapshot of positive and negative feelings, that is, happiness and anxiety at a particular point in time) (Dolan & Metcalfe, 2011).

The questions that represent these constructs (see further below) were selected with input from academics and various advisory groups, such as experts from the OECD, Eurostat, think tanks, and related social research experts. Today, ONS recommends the inclusion of these questions on all national surveys, so that sufficient data is obtained at regular time intervals to be able to track changes in the UK population. Currently, the four questions that underpin these concepts are included on more than 20 surveys in the UK. The questions are shown below:

The questions are usually introduced by a brief statement:

I would like to ask you four questions about your feelings on aspects of your life. There are no right or wrong answers. For each of these questions I'd like you to give an answer on a scale of nought to 10, where nought is 'not at all' and 10 is 'completely'.

Overall, how satisfied are you with your life nowadays?

Overall, to what extent do you feel the things you do in your life are worthwhile?

Overall, how happy did you feel yesterday?

Overall, how anxious did you feel yesterday?

National personal well-being measurements are representative of the perceived quality of life of the UK citizens, as well as variations in the measurements of well-being, related to changes in circumstances, policies, and wider events in society. The results from the data collected play a vital role in monitoring local and national well-being, informing the design of public policy, and appraising policy interventions. ONS monitors these measures with various subgroups of the population. Previous studies have successfully validated the personal well-being questions with different population subgroups, for example children and young people.

2.2 Linguistic Minorities in the UK

Other subgroups that the ONS and other governmental and non-governmental agencies are interested in monitoring closely with regards to their well-being are various minority groups in the UK. The 2011 Census identified that the greatest proportions of people currently residing in the UK and reporting that they do not speak English well or not at all are from Pakistan, Bangladesh, China, and Poland. These people form a vulnerable group, with less access to governmental and other resources and support due to their lack of proficiency in English. Thus, a measure of their well-being that would allow monitoring this indicator would be of high importance to governmental departments, local authority providers, and charities.

ONS has taken charge of organizing the translation of the personal well-being questions into these minority languages in the UK, so that they can be asked of monolingual non-English speaking people taking part in large national surveys. The translations will also

be useful for local authorities, community centres, support organisations, medical centres, and various other organisations. Due to low or non-existent levels of English, these primarily monolingual populations could be at a higher risk for low levels of integration and access to public services, which in turn could lead to higher distrust in public services or officials.

The present study aimed to find the appropriate translation of the personal well-being questions in Urdu (a Pakistani language) and Sylheti (a spoken language from Bangladesh), both languages spoken by significant minority ethnic groups in the UK. This initial research programme was designed to function as a pilot study for translating the personal well-being indicators. Future research programs will focus on the remaining languages (Chinese, Polish, and others).

There are about 400,000 Urdu speakers in the UK. Urdu is a form of Hindustani and has Persian and Arabic influences. It is the national language of Pakistan and is also an official language of India. About 19.05% of the British Pakistani population lives in London (224,000 according to the 2011 Census) and the British Pakistani make up 1.86% of the UK's population. Most people who speak Urdu in the UK come from the west Panjab and the Mirpur district of Azad Kashmir (Pakistan). Urdu is the fifth most spoken language in London and the fourth most common language in the UK. According to the 2011 Census, Urdu is the main language spoken by 0.5% (269,000) of the UK's population (Office for National Statistics, 2013b). About 25% of them declared that they could not speak English well or at all.

Similarly, there are about 400,000 people in the UK who speak Sylheti, but many refer to it as 'Bengali'. Sylheti is originally spoken in North India/Bangladesh and is derived from Sanskrit. Sylheti is sometimes considered to be a dialect of Bengali; however, its phonology and morphology differ from Bengali to the point that they are not mutually understandable. Sylheti has a traditional script, Siloti Nagri that was in use in Greater Sylhet (now part of Bangladesh and India) until the 1970s when socio-political pressures encouraged the disuse and destruction of the Siloti Nagri printing presses. Anecdotal evidence suggests that some speakers of Sylheti who are proficient in Bengali may use the Bengali script to write Sylheti; however, this is not standardised nor recognised as a formal Sylheti writing system. Sylheti is spoken by 95% of Bangladeshis living in the UK. According to the 2011 Census, over 450,000 UK residents said their ethnicity was Bangladeshi (about 53% of them were born in Bangladesh). The Census also found that Bengali (with Sylheti and Chatgaya)⁵ is the main language spoken by 0.4% (221,000) of the UK population and about 30% of them declared that they could not speak English well or at all. There are currently no estimates on the proportion of speakers of Sylheti as compared to Chatgaya; however, the majority of them are Sylheti speakers, thus this research programme focused on translating the questions into this language.

5 The 2011 Census asked people what their main language was. The response options presented on the form were tick boxes for English and "Other, write in (including British sign language)". The results were then aggregated and the top main languages identified were reported.

3 The Research Programme for Translating the Personal Well-being Measures into Urdu and Sylheti

In order to arrive at the best and most appropriate translations of the personal well-being measures in Urdu and Sylheti, a pilot research programme was devised and carried out. The programme involved two main stages: translation workshops and cognitive interviews to test the resulting translations. These steps have been found to be invaluable to the study for the following reasons: they revealed complexities relating to each language in the UK context, as well as cultural aspects of the Urdu and Sylheti-speaking communities.

3.1 Linguistic and Cultural Complexities

The two languages pose certain complexities which needed to be considered in the initial stages of the translation. Urdu, for instance, has various dialects which are more or less mutually understandable, while Sylheti is a spoken language with no official written form, script or dictionary.

Not all of the Urdu speakers in the UK are comfortable using formal Urdu, which might be associated with higher social status both in the UK culture and within the Urdu-speaking community. Instead, the main language of many Urdu speakers is a variation or a dialect of Urdu. Speakers using different dialects negotiate meaning in live interactions so they can communicate using the different forms and code-switching to English when necessary. In addition, some monolingual speakers might not be able to read or write Urdu. When considering translating the personal well-being measure in Urdu, we needed to consider the most appropriate terms for a population who is monolingual Urdu, who may or may not be comfortable speaking formal Urdu, and who may or may not be able to read and write Urdu.

Sylheti does not have a written form, and it is also the umbrella language for different dialects and Sylheti variations. Thus, the translations of the personal well-being measures needed to account for the fact that the language should be plain and simple in order to be understandable by all speakers. Also, a written version of the measures would be difficult to employ, given that it would have to either use transliteration (transcription of the Sylheti translation using the Latin script) or the Bengali script. Thus, it was agreed that the mode of question administration should be aural rather than written. Therefore, audio files were produced for the introduction and for each one of the well-being measures so that questions could be standardised despite the fact that they could not be written down.

Initial investigations into the language and culture of the two communities in the UK suggested that it would be appropriate to match the interviewers and the respondents by *gender*. Given the cultural context, the gender of the interaction dyad could influence respondents' willingness and comfort when taking part in a survey. Another important consideration was the participants' experience with surveys and research in general, i.e. answering questions on a numerical scale. For this reason, it was suggested that the translations should also have a level of informality to put these speakers at ease when answering a survey.

Thus, the translations for both languages needed to account for:

- the variations of the languages, i.e. dialects,
- potential lower levels of education,
- unfamiliarity with answering survey questions, for example using a scale (that is, participants might have never taken part in a survey and thus be unaccustomed with mapping their response on a numeric scale; they might also struggle to understand the purpose of the survey, and might feel that they are being tested),
- lack of a written form, in the case of Sylheti, and the formality of the written Urdu,
- gender.

For initial translations, the instructions given to the translations agencies were to make all attempts to keep the translations plain and simple, while keeping in mind the complexities described above.

3.2 Initial Translations

Figure 1 below provides a simplified representation of the research programme undertaken for these translations. Specificities relating to each language and cultural context led to slight variations in the outcomes or organisation of each of these stages for Urdu and Sylheti. These will be discussed in more depth below.



Figure 1 Research plan for the translation of the personal well-being measures

In preparation for the workshops, it was considered necessary to have two different initial translations to be used as starting points for the discussion. These translations were purposefully sourced from different organisations, such that the differences in the translations could constitute a good basis for discussion.

In the case of Urdu, we secured two versions of the translations for the personal well-being measures: One of them was the translation from the Civil Service People Survey, a survey that is already in the field and collects personal well-being data from Urdu native speakers, but which has never been cognitively tested in this language. Thus, it is not known whether the translations are culturally sensitive and understandable for the Urdu speakers in the UK. The second version of the translation was undertaken by an accredited translation agency which has specialised in translating questionnaires and other documents for social research. Both versions of the translations were in written form.

The Sylheti translations were undertaken in a different way: We have asked the translation agency for a written translation in Bengali (in Bengali script), as well as a female voice audio file and a male voice audio file in Sylheti. It was considered that given that Sylheti is a spoken language, the most appropriate form of administering the survey questions was using an audio file of the translation of the measures. This approach has been

previously used in the field when collecting field data from refugee participants for the German Socio-Economic Panel (SOEP) (Britzke & Schupp, 2017, Jacobsen, this volume). The resulting male and female audio translations presented some differences in pronunciation and phrases used for translating specific concepts. These differences were discussed with a researcher who was a native speaker and were deemed sufficient as a starting point for the discussions at the workshops.

3.3 Translation Workshops

In order to develop the most appropriate translations of the personal well-being measures, we adopted a technique which was closely related to the TRAPD model (Translate, Review, Adjudicate, Pretest, and Document) (Harkness, 2003; Harkness, Pennell, & Schoua Glusberg, 2004). The process is iterative in nature: Several initial translations are produced, which are then reviewed, and an adjudicator chooses the version that will then be tested. The team composed of translators, reviewers, and adjudicators must work closely together and document every stage of the process. The importance of the team translation stage of the process has been highlighted in many research projects undertaken recently with minority languages that investigated differences between the colloquial forms of the language and the formal variation. Cultural aspects on the language in context are also highly important and should be considered throughout the development process of the translated questions (Formea et al., 2014).

The translated questions are then to be pretested using “focus groups, cognitive interviews, split pretests with bilinguals and monolinguals, as well as respondent and staff debriefing” (Harkness, 2003, p. 41); and following this stage, the questions might undergo modifications again, before the translation is finalized. Using pretesting methods to ensure cultural and linguistic equivalence is increasingly becoming the norm in questionnaire translation. This is particularly challenging when the languages used belong to groups that are culturally very different from the main language and are not fully integrated to the mainstream society, such as Sylheti and Urdu in the UK. This poses two main challenges. First, establishing a rapport and good working relationship with members of the community who are native speakers of these languages; and second, being mindful of how cultural characteristics affect the way people respond to questions.

Thus, in order to assess the initial translations in both Urdu and Sylheti, we designed and carried out translation workshops, where the teams met to deliberate and debate the two translations for each language. The teams for each of the workshops were comprised of two bilingual survey interviewers for each language, bilingual NatCen researchers who acted as members of the community, but also as adjudicators, questionnaire development and cognitive testing experts from NatCen, and ONS personal well-being experts. The latter two did not have any knowledge of the languages, but were invited to the workshops to provide advice on the questionnaire design aspects, and insight into the personal well-being concepts and previous testing that had been carried out in the source language, i.e. English.

The workshops were found to be extremely useful from the point of view of the research team involved in the project and also the bilingual interviewers who were later trained in conducting the cognitive interviews for this project.

Positive feedback received from the team in the workshop related to the following four aspects. First, having two versions of the translated measures meant that the discussion could commence with agreeing which, if any, of the two versions was better and more appropriate for this context. In some cases, new words and phrases were suggested. These comments were noted and used in the development of the cognitive testing protocol. Second, having the ONS personal well-being experts present meant that questions related to the intended English meaning of the measures could be addressed on the spot. For example, in the case of the second well-being measure (Overall, to what extent do you feel the things you do in your life are worthwhile?) the various connotations of “worthwhile”, such as meaningful, with a purpose, etc. came up in both workshops. The well-being experts could steer the discussion and the translation in line with the English intended meaning. In addition, in the workshop three bilingual speakers were present, each with different life experiences and degrees of knowledge and involvement in the Urdu and Sylheti speaking. Although they all spoke their native language, their ties to the community and the variations of the language they spoke brought to light potential issues which could have arisen in the field. The members of the panel thus had the opportunity to discuss these differences in meaning and understanding, and arrive at a solution. Finally, the questionnaire design and cognitive testing experts could thus identify potential issues that might arise in cognitive testing and include these in the protocol for testing the new versions of the translations.

The conclusion from both workshops was that the initial translations were too formal to be used with the monolingual Urdu or Sylheti speaking communities in the UK. Certain words were removed or changed to address this issue. New translations were agreed for each one of the personal well-being questions, and these were circulated to all team members for review. Some other interesting findings included the fact that it was suggested that monolingual members in the community might not have ever taken part in a survey, and thus might struggle to answer a question on a scale. It was agreed that the cognitive interviewers would record this issue if it arose and attempt to guide participants to finding a suitable comparable way of answering the questions. Suggestions included using graphical representation scales, for examples using emojis or colour codes instead of the 0 to 10 scale and producing showcards. Based on these findings, we included a probe on the appropriateness of using showcards in the cognitive protocols and instructed the interviewers to explore whether there was a better way of representing the scale visually.

We also found that the showcard could use numerals in the Urdu and Bengali script, respectively, since they are different from the Latin numerals. However, it was also identified that since the potential participants who would take these surveys live in the UK, they may be able to understand and use Latin numerals. A probe relating to the use of different scripts for numerals was also added to the cognitive protocol.

Another important finding from the workshops was that the standard introduction to the four well-being measures would not be sufficient for respondents who have never participated in a survey before. Thus, it was suggested that before commencing the cogni-

tive interview, the interviewers would spend more time explaining the purpose and nature of large-scale surveys in an attempt to make the participants feel more at ease with the research process.

With regards to the four well-being questions, the team members discussed potential cultural differences that might occur in testing, such as the importance of religious beliefs and their potential impact on the responses that the participants would provide to these questions. The team members who were familiar with the cultural norms and religious beliefs of the communities suggested that it was possible for respondents to discuss their well-being in terms of the agency of a higher power, and outside of their personal control, which might have an impact on the way they responded to the personal well-being questions. It was agreed that this would be explored in the cognitive interviews if it came up.

The bilingual participants in the workshops agreed that the interviews need to be matched by gender in order to avoid making participants feel uncomfortable and allow them to speak freely. It was also mentioned that it might be considered inappropriate for different gender interviewer and participants to be alone in a room. For the cognitive interviews, we decided to pair the interviewers and interviewees by gender, and explore in probing whether this would be necessary in a real life situation.

Based on the thorough discussions which emerged during the workshops, new versions of the translations were agreed. These translations took into account cultural aspects that were relevant for these linguistic groups in the UK. The agreed translations were then formalised with the help of the bilingual researchers who performed the task of adjudicators. They also undertook the task of producing the final Urdu translation (and a transliteration of it to be used consistently in the cognitive interview in Urdu) and the audio recordings of the Sylheti translation (identical audio files with a male and a female voice were produced).

The research team produced cognitive protocols for both languages in English. Probes were included to explore various aspects of the translations that had been brought up in the workshops. The bilingual interviewers participated in an extensive cognitive interviewing training exercise. Also, a briefing session was organized with them to discuss in detail the purpose and scope of the cognitive testing, the scripted probes from the protocols, and the use of spontaneous probes when needed. The interviewers were also asked to interview each other in English, using the protocol provided, and they received extensive feedback about it. They were advised to discuss amongst themselves the protocols and how they would carry out the interviews in Urdu and Sylheti, respectively. The benefits of pre-testing methods in the production of the final survey questions were discussed during the training and the briefing sessions. They also received in-depth training and instructions on recruiting monolingual participants in the community and were advised to be in constant touch with the research team to provide feedback and updates on recruitment and fieldwork.

3.4 Cognitive Interviews in the Translation Context

An integral part of the translation process was the cognitive interviewing stage. As discussed, we identified a male and a female survey interviewer for each language, who participated in the workshops, and who were then trained and briefed to conduct cognitive interviews in their native languages. For each language one of the interviewers was born

in the UK and had English as a dominant language, but was fluent in spoken Urdu and Sylheti, respectively. Table 1 presents the demographic characteristics of the interviewers who worked on this project.

Table 1 Demographic information for the interviewers working on the project

Interviewer	Gender	Dominant language	Can write Urdu/Bengali?	Age	Country of birth
Urdu 1	Male	English	No	Under 35	UK
Urdu 2	Female	Urdu / English	Yes	Over 35	Pakistan
Sylheti 1	Male	Bengali/Sylheti	Yes	Over 35	Bangladesh
Sylheti 2	Female	English	No	Under 35	UK

Cognitive interviewing methods provide an insight into the mental processes participants use when answering survey questions, thus helping researchers to identify problems with question wording and design. These methods investigate four cognitive stages: how participants understand and interpret survey questions, how they recall information that applies to the question, the judgements they make as to what information to use when formulating their answers, and the response mapping process (Tourangeau, 1984). These processes were closely monitored with participants who were mostly monolingual speakers of Urdu and Sylheti.

In addition to conducting cognitive interviews in Urdu and Sylheti, ONS had previously conducted a suite of cognitive interviews on the personal well-being questions in English (Dolan, Layard, & Metcalfe, 2011). This meant that during the analysis phase of the cognitive interviews we could establish whether any issues arising were unique to the Urdu and Sylheti translations (and therefore indicative of a problem with the translations or cross-cultural equivalence) or whether similar issues were also documented for English speakers. Goerman and Caspar (2010) suggest that if the source and translated versions of the questionnaire cannot be developed at the same time (as proposed by Harkness, 2003), they should at least be tested in parallel with bilingual speakers, as well as monolingual speakers of the two languages. Due to time and budget constraints (addressed by Goerman & Caspar, 2010, as well), this was not possible to implement at this stage; however, the protocols for the cognitive interviews were informed by previous pretesting exercises conducted with English speakers, as well as by the issues that were discussed at the translation workshops.

The four bilingual survey interviewers, as listed in Table 1, were trained in cognitive testing methods. They were each advised to recruit five monolingual speakers of their own gender and language, and interview them cognitively using the protocol developed during the workshops. The interviewers were asked to use their community networks to find suitable participants, but were advised to avoid interviewing acquaintances, friends or family members due to the potential sensitive nature of the questions. A total of twenty cognitive interviews were conducted.

The cognitive protocols were produced only in English and discussed in depth at the briefing session with the interviewers. In a cognitive interview, the questions that need to be standardized are the survey questions which were provided to the interviewers either in audio format (for Sylheti) or in writing, including in transliteration (for Urdu). The cognitive probes were discussed with the research team, and the interviewers discussed and agreed amongst themselves the best wording to use to introduce the project and ask the cognitive probes during the interviews. The interviewers were given the opportunity to practice interviewing both in English and in their native language which each other and raise any questions they might have with the research team.

The protocols included a description of the purpose of the study and the potential usefulness of the personal well-being measures for the ethnic community (and the interviewers were advised to discuss this in as much depth as needed with their respondents), an example of the think aloud technique, the rights of the research participants to anonymity and confidentiality as well as the voluntary nature of the study. The introduction and the four personal well-being questions were then delivered in the standardised form produced after the translation workshop. In-depth probing of each of the questions followed. The interviewers were instructed to ask all probes in the protocol but also use spontaneous probes if necessary. The last part of the interview asked participants probes about the cultural appropriateness of the questions and whether they would be comfortable answering these questions if they were part of a survey. The gender aspect of the interviews was also included as a probe in the protocol – whether they would feel comfortable answering the questions if an opposite gender interviewer would ask them. The scripted probes explored: comprehension of key terms used in the questions, comprehension of items overall, and comprehension of the research process, i.e. answering on a scale, the relevance of conducting surveys, etc. Cultural sensitive probes were also included to explore whether interviews should be carried out by same gender interviewers, as well as the perceived appropriateness of the personal well-being questions.

Participants were interviewed in their own homes or in a community venue, whichever they preferred. The interviews were audio-recorded with the participants' consent. Respondents were eventually given a £15 voucher to thank them for their time and help.

Participants were recruited in different locations in England, that is, where the bilingual interviewers had contacts in the community that could help them identify monolingual speakers. The only recruitment quotas were around language: participants in the cognitive interviews had to be monolingual or predominantly monolingual speakers of Urdu and Sylheti, respectively. We also advised interviewers to attempt to recruit a varied group of people in terms of age, educational level, and length of residence in the UK; however, these were not set as quotas. Throughout the duration of the fieldwork, the research team and the bilingual interviewers were in constant contact to monitor recruitment as well as the actual fieldwork.

Once the majority of the cognitive interviews had been conducted, all panelists from the translation workshops were invited to a debriefing exercise. The anecdotal findings related to each of the four personal well-being measures were discussed with the panelists. The findings from debriefing meeting suggest that, overall, the translated measures were well understood by the participants; however, many of them struggled with the comprehension

of the research process – the purpose of these questions, the use of scales to provide an answer, the relationship between a quantified answer on a scale and their everyday life experiences, giving a response but qualifying it as god's will or plan for their lives, and so on. The bilingual interviewers were then asked to review the recording and submit detailed interviewer summaries in a template provided. The interview summaries were reviewed in depth by the researchers who then inputted them in an Excel pro-forma matrix. All interview summaries were transferred to the two matrices – one for the Urdu testing and one for Sylheti. Responses to each test question were recorded, along with observations made by interviewers, any think aloud data, findings from each of the scripted probes, and suggestions for the most appropriate translation for each of the personal well-being measures. Any uncertainties and queries were discussed in depth with each interviewer. Suggested final wordings of the questions were also recommended by the interviewers based on their respective interviews. Data could thus be read horizontally as a complete case record for an individual interview, or vertically by question, looking across all cases.

Once the matrices were completed, the data in the matrix were reviewed thematically in conjunction with reviewing the already documented issues that had previously been uncovered in testing the same questions with English speaking participants. We employed the use of the C-NEST⁶ strategy (Fitzgerald, Widdop, Gray, & Collins, 2011) to identify and correct issues in the translation of the well-being measures. Issues that occurred both in the translated questions and the source English question were noted but were not addressed in terms of recommending changes to the translation (as the source questions are extensively used in other surveys, and thus the aim was to retain conceptual equivalence as far as possible). Translation problems and issues related to the cultural portability of the four well-being measures were considered in depth at the analysis stage and guided their final translation and recommendations for use in the field.

The cognitive interviews suggested similar general findings for both languages: Participants suggested that the introduction to the question needed to be more detailed, to explain the purpose of conducting the study and asking these questions, to explain in more depth that there were no right or wrong answers, and that the responses provided by the participants would be kept confidential and would not be disclosed. The interviewees felt that this level of reassurance would prepare the respondents for the questions and help in building rapport. Furthermore, men suggested that they would not object to being interviewed by either a man or a woman, but women interviewees admitted to preferring to talk to a woman, as that would allow them to be more honest. In both languages, participants brought up religion and how people should not be ungrateful for what they have in life; however, a close examination of the responses to the questions and the probes showed that the answers were quite varied and in line with the life circumstances and experiences of the participants. Some participants had a difficult time choosing a response option on a scale and would have preferred to explain their answers in their own words. Probing revealed that this could have been a function of not having previously participated in any

6 C-Nest or Cross-national Error Source Typology has been developed to identify the sources of various errors that might come up when pretesting surveys in translation. These are: poor source question design, translations problems and cultural portability.

kinds of surveys. Interviewers had to probe explicitly and insist until the participants chose a particular response option on the scale, rather than giving a verbal response. Overall, all participants were able to read the Latin script numerals and declared that they could answer using a showcard which employed their use.

Additionally, the four personal well-being questions were explored in depth. In some instances participants made suggestions for alternative words or phrases and those were taken into account as well when agreeing the final translations. Participants were able to provide examples from their lives to complement the response options selected. These examples seemed to vary across genders. For example, male participants declared that for them being satisfied with life meant that they could provide financially for their families, while women talked about the success of their children and the harmony of the home life. Similar findings were identified in the pretesting of these questions with different sub-groups of English speakers – for example, children and young people gave examples of how being satisfied with life meant that they received good grades in school. This suggests that the questions worked well, leading the respondents to consider their own life circumstances and experiences when providing an answer.

The findings from the cognitive interviews were discussed again in a workshop with the same team members as the initial translation workshops. This debriefing session was important as it brought together the language experts and the researchers to discuss the cognitive interviews. Final translations of the personal well-being measures were agreed with the interviewers following the debriefing workshop and analysis of the cognitive interviews. The bilingual researchers who had previously assumed the role of adjudicators took up this role once more and created the final version of the translations. These were shared with the bilingual interviewers, who then confirmed that they were appropriate based on their interviewing experience. The final output for the pilot research programme included for both languages audio files with a male and a female voice as well as transliterations in Latin script, and furthermore a written translation in Urdu as well as Bengali and Siloti Nagri, respectively. Audio files were produced for both languages to allow for choice of the most appropriate form of presenting the questions. Our final recommendation was to complement the audio and the written translations and transliterations, depending on the needs of the interviewer and the interviewee.

4 Conclusions

The pilot research programme designed for the translation of the personal well-being questions in Urdu and Sylheti showed that the translation workshops and the cognitive testing stages were paramount to the success of the project. Sufficient time needed to be allocated to these stages in order to allow for extensive discussions around the meaning and purpose of the personal well-being measures, the cultural and religious characteristics of the language communities, and how these might affect the final outcome of the project, and the findings from pre-testing the translations with monolingual speakers of these languages.

Furthermore, in the present study, the bilingual interviewers were novice users of the cognitive interviewing methods. Although overall the outcome of the interviews was very positive, they did require extensive feedback and reassurance from the researchers. Also, given that they were survey interviewers, it was difficult for them to deliver the cognitive interviews which are a qualitative method. For example, the interviewers struggled with spontaneous probing, as well as with reporting the interview findings in a summary format. The researchers were reliant on them to report the interview findings objectively, which at times was difficult, because the interviewers had their own opinions about the best phrases to use in translation. We recommend that, whenever possible, the cognitive interviews should be conducted by bilingual researchers who are also experts in pretesting methods. Alternatively, sufficient time and resources need to be allocated to the training stage, potentially building up time for extensive practice exercises and feedback.

Some important and very relevant lessons learned during this process were that the composition of the team was highly important: The researchers who were experts in personal well-being and the questionnaire development and pretesting experts worked closely with the bilingual interviewers and the adjudicators (who were bilingual social researchers) for both workshops, that is, the translation workshop and the debriefing session. Furthermore, communication and rapport among the members of the team involved in the research programme was vital, because of the important role on the research puzzle that each member of the team had. All members of the team contributed invaluable information during the entire process, bringing in different but very valuable perspectives. The different professional and life experiences of the team members meant that many potential issues in the field could be foreseen before pretesting and steps could be taken to ensure appropriate responses. Good rapport among the team members meant that when unforeseen issues came up in the field, these could be addressed and solved immediately. The research programme needs to allow sufficient time for each stage to unfold naturally, without hastening.

Finally, it should be noted that the research programme to translate the personal well-being questions to Urdu and Sylheti was designed and carried out to reflect the linguistic characteristics of the context of these languages, i.e. the UK. The translation workshops and the cognitive testing helped to adapt the translations to the cultural and linguistic specificities of the environment, and thus yielded translations of these measures that were deemed to be locally applicable and relevant for these communities in the UK. The research programme undertaken also provided valuable information for the next stage, that is, the use of the translated personal well-being measures in the field.

References

- Britzke, J., & Schupp, J. (2017). *SOEP wave report 2016*. Berlin: DIW.
- Dolan, P., & Metcalfe, R. (2011). *Comparing measures of subjective well-being and views about the role they should play in policy*. London, UK: Office for National Statistics.
- Dolan, P., Layard, R., & Metcalfe, R. (2011). *Measuring subjective well-being for public policy*. Retrieved from <http://eprints.lse.ac.uk/35420/1/measuring-subjective-wellbeing-for-public-policy.pdf>

- Fitzgerald, R, Widdop, S., Gray, M., & Collins, D. (2011). Identifying sources of error in cross-national questionnaires: Application of an error source typology to cognitive interview data. *Journal of Official Statistics*, 27, 569–599.
- Formea, C. M., Mohamed, A. A., Hassan, A., Osman, A., Weis, J. A., Sia, I. G., & Wieland, M. L. (2014). Lessons learned: Cultural and linguistic enhancement of surveys through community-based participatory research. *Progress in Community Health Partnerships: Research, Education, and Action*, 8, 331–336.
- Goerman, P.L., & Caspar, R. A. (2010). A preferred approach for the cognitive testing of translated materials: Testing the source version as a basis for comparison. *International Journal of Social Research Methodology*, 13, 303–316.
- Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. J. R. van de Vijver, & P. Ph. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: Wiley.
- Harkness, J., Pennell, B.-E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 453–473). Hoboken, NJ: Wiley.
- Jacobsen, J. (2018). Language barriers during the fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany. In Behr, D. (Ed.), *Surveying the migrant population: Consideration of linguistic and cultural aspects*. Köln: GESIS.
- Office for National Statistics. (2011). *Harmonised concepts and questions for social data sources. Introduction to harmonised standards*. Version 4. Retrieved from <http://webarchive.nationalarchives.gov.uk/20160106185655/http://www.ons.gov.uk/ons/guide-method/harmonisation/about-harmonisation/guide-to-harmonisation/index.html>
- Office for National Statistics. (2013a). *2011 Census: Detailed analysis – English language proficiency in England and Wales, main language and general health characteristics*. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/cultural-identity/language/articles/detailedanalysisenglishlanguageproficiencyinenglandand-wales/2013-08-30>
- Office for National Statistics. (2013b). *2011 Census: Quick statistics for England and Wales, March 2011*. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011Censusquickstatisticsforenglandandwales/2013-01-30#tab=Main-language>
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual review of psychology*, 52, 141–166.
- Tourangeau, R. (1984). Cognitive science and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100), National Academy Press: Washington DC.
- Willis, G. B., Kudela, M. S., Levin, K., Norberg, A., Stark, D. S., Forsyth, B. H., ... Hartman, A. M. (2010). Evaluation of a multistep survey translation process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler ... T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp.141–156), John Wiley & Sons, Inc., Hoboken, NJ, USA.

The Place of Expert Review in Translation and Questionnaire Evaluation for Hard-to-Count Populations in National Surveys

Patricia Goerman, Mikelyn Meyers, & Yazmín García Trejo⁷

Abstract

Many researchers consider it a best practice to include respondent pretesting as part of the survey translation process (Survey Research Center, 2016). The U.S. Census Bureau has a Pretesting Standard that delineates requirements to ensure that any data collection instrument “works,” by verifying that it can be administered as intended by interviewers and understood and responded to appropriately by respondents (U.S. Census Bureau, 2015). Expert review is included in the standard but it is described as a sort of evaluation method of last resort due to its lack of inclusion of respondent pretesting in the process. In large survey organizations, there can be a variety of types of materials in need of translation ranging from actual survey questionnaires to other support materials. Due to resource limitations, expert review is a method that often comes into play in the absence of resources for respondent testing. Recent discussions have involved defining and modernizing the Census Bureau’s approach to translation methodological expert reviews and looking at how they can best fit into the overall translation process. This paper provides a review of the literature on expert review, a description of how this type of review is currently done at the U.S. Census Bureau, along with limitations, challenges, and a plan for future research in order to further develop the method.

1 Introduction

Methodological expert reviews are often used to revise survey materials before they are shown to respondents either for pretesting or when an instrument is fielded. A methodological expert review can have various goals: to incorporate a team-based review of an instrument that did not include review as part of the translation process, or to do a joint methodological review along with a review of the source text. A methodological expert review gives the opportunity to methodologists, translators, and other experts to join forces and provide feedback based on empirical evidence, literature, and experience

⁷ *Disclaimer:* This article was written to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

that will ultimately improve an instrument. Some best practices for incorporating review into the translation process itself have been documented (Survey Research Center, 2016).

Ideally, expert review of survey materials (whether source text or translated) is a preliminary step that precedes respondent testing, but when funding, resources or time do not allow for respondent testing, expert review may be the only “pretesting” method possible. As such, establishing sound best practices for implementing methodological expert reviews of translated survey materials is vital, and this paper seeks to fill a gap in the cross-cultural survey methodology literature in this regard.

2 Review of the Literature

2.1 Hard-to-Count Populations in National Censuses and the Need for Multilingual Questionnaire Development and Evaluation

The fundamental goal when developing surveys for use across linguistic and cultural groups is to ask the same questions across these diverse types of survey respondents. Survey designers strive for comparable understanding by respondents. For example, ideally survey methodologists aim to use questions and terms that an average respondent would be able to interpret and respond to, which will ultimately contribute to collection of more parallel data across diverse populations and will increase the likelihood that respondents understand questions and concepts as methodologists intended.

This topic is particularly relevant when it comes to including hard-to-count populations such as immigrants in national censuses. In a population census, the goal is to survey the entire population and it is therefore important to design survey instruments and supporting materials in as many relevant languages as possible. In order to include as many people as possible in the count, the U.S. Census Bureau, for example, has historically divided languages into “tiers” for which varying levels of support have been provided. In 2010, the highest level of support was provided for languages for which American Community Survey data showed that there were 100,000 or more occupied U.S. housing units with no persons aged 15 or older who spoke English “very well.” Spanish was the most commonly spoken language followed by Chinese, Vietnamese, Korean, and Russian.

In large-scale surveys, there can be a variety of types of materials in need of translation, including actual survey questionnaires, respondent letters and brochures, interviewer instructions, instructional videos, and training manuals for interviewers. In any survey life cycle, there will inevitably be constraints on resources. Materials may need to be prioritized in terms of data collection instruments versus supporting materials and supplementary materials that may not even be seen by respondents. Due to resource limitations, expert review is a method that often comes into play in the absence of ability to conduct respondent pretesting. Basically, if one cannot pretest a survey instrument with respondents through methods such as cognitive interviews or focus groups, researchers are left to rely on expertise and feedback from survey methodologists and language experts. This

paper discusses the ideal role and methodology for expert review in the context of questionnaire translation.

2.2 Review as a Part of the Survey Translation Process

Many researchers consider it a best practice to include review as part of the survey translation process itself (Survey Research Center, 2016). In fact, it is often recommended that translation be done via a team approach as opposed to being done by one translator or through methods such as back translation (Survey Research Center, 2016; Harkness, 2008a; Harkness, 2008b; Harkness, Pennell, & Schoua-Glusberg, 2004; Pan & de la Puente, 2005; Willis et al., 2010). The key players in the team translation method are translators and/or reviewers working together to come up with the final product. The process can involve multiple translators each translating parts of the instrument or each translating the whole instrument independently. The work then includes a review of the draft instrument amongst the group of translators. The team assigns an adjudicator who makes final decisions in the event of disagreement once the team compares their work.

Many researchers recommend that translation and review be embedded in a larger survey development process. Methods such as TRAPD (Translation, Review, Adjudication, Pretesting and Documentation) recommend respondent pretesting as the next step in the development of a translated survey instrument. The Cross-Cultural Survey Guidelines (CCSG) provide a TRAPD graphic (see Figure 1). Many survey researchers point to the critical importance of including respondents from various social locations or backgrounds through pretesting. After all, it can be difficult for questionnaire designers or translators to imagine how respondents with differing characteristics might interpret a survey question.

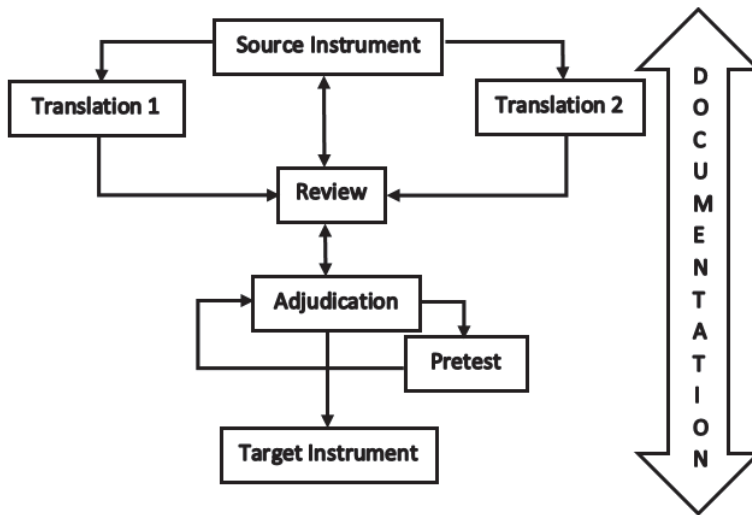


Figure 1 Illustration of the TRAPD model from the Cross-Cultural Survey Guidelines (Survey Research Center, 2016)

Despite team translation and pretesting being the industry standard, there are many agencies that do not have the resources to use the method for every type of survey material needed. In the event that an agency or organization has not been able to employ the team translation method followed by respondent pretesting, some agencies undertake an expert review of translations as a separate step involving multiple reviewers. In some cases, standalone reviews of translations are conducted in preparation for respondent pretesting rather than instead of respondent pretesting. There is a lack of consensus about whether and how to best conduct an expert review of a translation as a questionnaire evaluation method.

2.3 Methodological Expert Review of Translations as a Questionnaire Evaluation Method

Expert review can be defined as a method by which “questionnaire design experts appraise the questionnaire, applying generally accepted questionnaire design principles and knowledge based on their own pretesting experiences” (Willimack, Lyberg, Martin, Japac, & Whitridge, 2004). The goal is to predict interviewer and/or respondent difficulty with the questionnaire items and recommend ways to improve the instrument (U.S. Office of Management and Budget (OMB) Inventory, 2016b).

Many researchers discuss shortcomings of the expert review method (Presser & Blair, 1994; Tourangeau, 2004). In particular, Tourangeau points out that expert review tends to be carried out in a “nonrigorous, even subjective way” (p. 210). Many expert reviews involve assessment by experts but there is a lack of consistency in methods used across experts. Some researchers go so far as to say that “expert review, however systematic, does not provide transparent, empirical, or analyzable data, and cannot be considered a scientific method” (U.S. OMB Inventory, 2016b, p. 13). However, several cognitive appraisal coding assessments have been created which attempt to apply such consistency to the work of expert reviewers. Willis and Lessler’s (1999) Question Appraisal System and Lessler and Forsyth’s (1996) coding system are two such examples.

At the same time, expert review also has its advantages. Many researchers recommend that expert review be done early in the questionnaire design process since it can identify large issues prior to extensive development (Willis, 2005). In addition, it is a low-cost method compared to many other pre- and post-field testing methods (Tourangeau, 2004). Finally, expert review can be the only feasible pretesting for some languages and materials, and while team translation followed by respondent pretesting is the preferred approach, an expert review is the next best method available for ensuring high-quality translations, particularly when the expert review is conducted systematically by multiple reviewers with specialized expertise.

The U.S. Census Bureau has a Pretesting Standard that delineates requirements to ensure that any data collection instrument “works” by verifying that it can be administered as intended by interviewers and understood and responded to appropriately by respondents (U.S. Census Bureau, 2015). Most of the pretesting methods addressed in the Census Bureau standards involve respondent testing. In fact, the Census standards specify that respondent testing is both critical and required in order to identify and resolve any issues with content,

context effects, skip patterns, formatting, and navigation. Expert review is included in the standards but it is described as a method that does not generally satisfy the pretesting requirement. However, it is also described as meeting minimum testing requirements in the event that time and resources do not allow for respondent testing, which is often the case for translated survey instruments and supplementary materials.

3 Expert Review of Translations at the U.S. Census Bureau

The U.S. Census Bureau came out with a Translation Guideline in 2004 and a recommended translation review process in 2009 (Pan & de la Puente, 2005; Pan, 2009). The translation guideline was not a required procedure but rather was developed by Census Bureau staff who had previously collaborated in the development of the CCSG translation guidelines with the Comparative Survey Design and Implementation (CSDI) organization. The Census Bureau guideline was the work of linguists and survey methodologists and details “best practices” to be implemented by the Census Bureau when circumstances permit. Resource constraints have led to translations being conducted using methods besides the team approach, which is a departure from the “best practices” that would ideally be implemented. In response to these resource constraints, we have incorporated both team translation guidelines and survey methodology expertise into the expert review process.

The Census Bureau guidelines list types of people who should ideally be involved in the review process: subject-matter specialists, program managers, survey methodologists with knowledge of questionnaire design and pretesting, translators, and translation adjudicators.

4 Current Practices at the Census Bureau Center for Survey Measurement

The U.S. Census Bureau has a research center called the Center for Survey Measurement (CSM) that is made up of survey methodology experts who work on questionnaire design and pretesting. The Language and Cross Cultural Research (LCCR) group is a sub-team that works on questionnaire design and pretesting of translated instruments. The LCCR group has been receiving an increasing number of requests to do expert reviews of translated survey instruments in recent years.

Methodological expert reviews are somewhat different from review as part of the translation process itself in that they are sometimes done as a substitute pretesting method both for original English source wording as well as for translations. However, the line can get blurred when it comes to the methodological expert review of a translation. The current LCCR process involves having bilingual survey methodologists with language/cultural expertise and a background in survey methodology, social sciences, linguistics and/or translation review the translation. If not included on the official review team, subject-

matter experts and translators are often consulted with questions during the review process. Reviewers conduct an independent review of the materials in question and then meet for a consensus meeting, at which time an adjudicator makes any final decisions in the event that the team is unable to come to agreement. The types of participants involved in a particular project vary based on availability of staff, budget, timeline, and the types of materials being reviewed.

5 The LCCR Translation Methodological Expert Review Method

Step 1: Getting the Expert Team Together

The LCCR team is typically given a translation as a starting point. Some translations are done by an internal Census Bureau translation office and some are conducted by an outside agency. Typically, we do not have information about how the translation was done and whether any review has already taken place. In some cases, the translation may have been done by a single translator with no additional review. The LCCR team's expert reviewers begin by identifying relevant team members (typically two researchers) and a team lead who acts as both a reviewer and the adjudicator. The adjudicator is usually an experienced bilingual survey methodologist. While our teams often consist of a team lead and two supporting reviewers, the decision regarding how many people to staff on an expert review is often driven by staff availability and how many people are needed to fulfill all needed roles: survey methodologist, subject-matter expert, certified translator, program manager, and adjudicator. While it is ideal for bilingual program managers and independent translators, who did not complete the original translation, to participate as official members of the expert review team, in practice many surveys do not have a bilingual program manager, and it is not always possible for a translator who did not complete the original translation to participate. In those cases, the team lead works closely with program managers and translators as questions arise during the review process, but the program managers and translators do not act as official expert reviewers.

Step 2: Kick-off Meeting and Independent Reviews

The team lead creates an electronic repository for the documents under review and leads a kick-off meeting to provide the reviewers with background information about the content of the survey, the intended audience, any research conducted in the past and a thorough explanation about the process of the expert review along with deadlines and assignments. At such meetings, the team lead typically walks the reviewers through the source text and translations, the scope of the review, applicable context regarding survey operations, information about how the translation was produced, and any applicable reference materials (e.g., online medical dictionaries, existing translations of common batteries of survey questions, etc.).

Regardless of the format used to record reviewer comments, the review process should be completely independent, such that individual reviewers are blind to the comments of other reviewers until their review is complete. This ensures that reviewers, including the adjudicator, are not biased by referencing the comments of others.

Step 3: The Consensus Meeting

In the LCCR team, the team lead is typically responsible for combining reviewers' comments prior to consensus meetings into one document to facilitate group discussion. When time permits, consensus meetings are conducted with all reviewers, and comments are discussed one-by-one. Reviewers reach a consensus on comments during the meeting itself or decide on next steps, e.g., designating a particular reviewer to research a given term and report findings back to the group. The team lead rarely adjudicates disagreements between reviewers, as most disagreements can be resolved by coming to agreement through discussion or by further research in reference materials.

When the length and complexity of the materials as well as conflicting staffing demands do not permit all reviewer comments to be discussed individually during consensus meetings, the team lead prioritizes comments, looking for those that require group discussion: Where reviewers are in disagreement, or where a single reviewer had concerns about a particular item not shared by other reviewers, the team lead flags the item for group discussion. This process of prioritizing comments for discussion is necessarily subjective, and as such requires a team lead with familiarity with the target language, survey methodology, and subject-matter expertise.

Step 4: Tracking Decisions and Preparing the Recommendations

The team lead tracks decisions on reviewer comments during consensus meetings and distributes the document for team review either following each consensus meeting or after all consensus meetings are complete. Final documentation includes the source text, the original and proposed translations, back translations⁸ of new recommendations in the event that the client does not speak the target language, along with an explanation for proposed revisions. Typical explanations document the reasons that revisions are being recommended. For example, there may be 1) missing or additional concepts in the source text compared to the target language text, 2) mismatches between source text and target

8 The term “back translation” is best known as a description of a translation method where one translator translates source version A into target language B. A different translator then translates version B back into source language A so that the two source language versions can be compared by monolingual stakeholders to look for errors. This is not a highly recommended translation method (see Behr, 2017). We use the term “back translation” here to describe a sort of literal or word-for-word translation that can be used to explain to monolingual survey sponsors what newly recommended target language wording means. Our “back translations” of newly recommended wording are often provided along with an explanation of the meaning and context and reasons for recommending the change. This method enables a survey sponsor who does not speak the target language to participate in the decision-making process in terms of whether to accept recommendations.

language text register (i.e., formality or complexity of terms), 3) concerns that target language terminology may not be well understood by respondents of a variety of national origins when applicable (e.g., Spanish or Arabic), 4) grammatical errors, 5) readability problems, etc. In addition to including the recommendations from the expert review, final documentation also outlines the methodology used to conduct the expert review.

Challenges of conducting such expert reviews are manifold. Given the amount of information that must be conveyed about each individual recommendation (original source text, original target language text and back translation, recommended target language text and back translation, rationale for recommending a change), documentation is a major hurdle in any expert review. Establishing procedures for each reviewer to provide comments in a format that facilitates the combination of comments is vital, while careful note-taking during consensus meetings is also necessary to producing a high-quality final product. When possible, documentation from sponsors tracking which changes were implemented and why can be helpful for targeting future research on problematic items when schedules or resources permit respondent testing.

Step 5: Final Recommendations

Presenting final recommendations to clients is also challenging given that many stakeholders are not proficient in the target language and they may lack experience with survey translations. LCCR typically walks clients through our recommendations to ensure that all relevant information has been conveyed in a manner that is understandable and that allows stakeholders to make informed decisions about potential changes to their survey instruments. There are some types of findings, such as problems detected in both the original and source language versions of the questionnaire that can be out of scope for our translation reviews. While recommendations maybe valid, they often cannot be implemented due to constraints in the timing of the translation expert review during the survey life cycle, where the original English wording might already be final.

While expert review of a translation is a valuable step during the development of any translated survey instrument, it is particularly valuable when the translation was not conducted via the team approach. The expert review process allows an approximation of the team approach to be included in the review process itself. Additionally, expert review can improve translated survey instruments in advance of conducting respondent pretesting so that respondents do not waste time grappling with issues that could be addressed prior to respondent pretesting. When changes to translated survey instruments cannot feasibly be made prior to respondent pretesting (e.g., programming schedules do not allow updates to online instruments), expert review of translated survey instruments can inform the design of pretesting protocols to target suspected problems with the translation prior to conducting respondent pretesting. Finally, while testing translated survey instruments with respondents is the gold standard of respondent pretesting, when such testing is not possible due to timing or resource constraints, expert review is a vital step in the survey life cycle.

Selected Examples of Translation Expert Review Recommendations

Expert reviews can catch many important potential questionnaire issues, both in materials that are newly translated as well as in materials that were previously translated and have recently undergone additional pretesting. In a recent study, we conducted an expert review of translated Census Bureau Group Quarters materials in Spanish. Group quarters are places such as college dormitories, prisons, shelters or residential treatment centers. Reviewers noticed that the name of the form itself may present some issues. The form in English was called “Individual Census Report.” Spanish-speaking reviewers were concerned that the term “report” in both English and Spanish might sound threatening to respondents and recommended instead “Individual Census Questionnaire.” In addition, the term “facility” was translated as “facilidad” throughout the questionnaire materials. Reviewers pointed out the fact that “facilidad” in Spanish is a noun that means “ease” and “ability” and it does not mean facility and/or group quarters as in English. They felt that the use of “facilidad” in the instrument in Spanish in this context was an informal Anglicism, which is inappropriate in an official government survey. The review team recommended changing these terms and then conducting respondent testing as a next step.

In another recent study, LCCR conducted an expert review of a Spanish translation prior to cognitively testing the translation with Spanish-speakers. The English and translated wording were as follows (Table 1):

Table 1 English – Spanish translation of “break the law”

Source Text	Original Translation	Back Translation
Is the census used to help the police and FBI keep track of people who break the law, or is it not used for this?	¿El censo se usa para ayudar a la policía y al FBI a mantener un registro de las personas que infringen la ley o no se usa para este fin?	[Is the census used to help the police and the FBI maintain a registry of people who infringe on the law, or is it not used for that purpose?]

Expert reviewers commented that the term “infringir” [infringe] in the translation was comparatively a higher register term than “break” in the English source text, and recommended using a lower register term like “romper” [break] for consistency with the English. This change was not implemented prior to cognitive testing. Findings from cognitive testing indicated that three of ten Spanish-speakers expressed concerns about the term “infringir” and felt that it would be difficult for many Spanish-speakers to understand. Findings from this cognitive testing project validated the recommendations made by expert reviewers.

6 Limitations to the Current Process

There are several limitations to our current process. For example, we do not always have the same number and types of staff members available to perform an expert review. Some surveys are about complex, specialized topics such as health conditions or social programs. It can be difficult to find bilingual survey methodologists, linguists or translators with particular knowledge of the subject-matter.

Some instruments under review are “legacy” instruments involving time series that are not easily changed. Similarly, in many cases, expert reviews of a translation occur separately from a review of the English source language version and at a time when only minimal changes can be made to the English version. This limits the scope of recommendations that can be implemented.

7 Current Issues/Challenges

While this paper outlines the procedures that LCCR uses to carry out methodological and translation expert reviews, more work remains to be done comparing practices within and across agencies and organizations, and more importantly evaluating those models to develop more formal best practices.

Separately from the issue of which procedures result in the most successful expert reviews is the question of why expert reviews are often the only form of pretesting for many translated materials. While this decision is often driven by resource and time constraints, new project schedules and budgets should ideally address the needs of translated materials at the outset rather than fitting the design, review, and testing of these materials into schedules and budgets after the design, review, and testing of English source materials have already been scoped out. Ideally, source and target language materials should be developed and tested in parallel in order to allow for a higher quality final product and to ensure that testing of target language materials is not reduced due to cost or schedule over-runs.

8 A Program of Research for the Future

LCCR has recently begun experimenting with our expert review procedures in an effort to formalize best practices. In one recent project, we experimented with the inclusion of bilingual field interviewers as expert reviewers. In this evaluation, we coded the interviewers’ comments to determine at what point additional reviewers provided diminishing returns as well as the extent to which interviewers who have contact with the general public but lack survey methodology, program, and subject-matter expertise are able to contribute actionable comments to the review process (Goerman, Meyers, & García Trejo, 2018).

As it becomes more common and necessary for surveys to provide translations of their materials in order to cover more of the population, LCCR is being asked to do more expert reviews of translated survey materials. Staffing constraints sometimes require that reviews be completed without at least three independent reviewers and a series of consensus meetings. LCCR is experimenting with models of expert review that require less staff participation while meeting a minimum quality standard for those projects with smaller budgets or tighter schedules. CSM continues to advocate for respondent testing as the gold standard of evaluating translated survey materials, while acknowledging that real-world constraints make establishing best practices for expert reviews a much-needed initiative in cross-cultural survey research.

References

- Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, 20, 573–584.
- Survey Research Center (2016): Guidelines for best practice in cross-cultural surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Retrieved from http://ccsg.isr.umich.edu/images/PDFs/CCSG_Full_Guidelines_2016_Version.pdf
- DeMaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 89–108). New Jersey: Wiley.
- Forsyth, B. H., & Lessler, J. T. (1991). Cognitive laboratory methods: A taxonomy. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 393–418.) New York: John Wiley & Sons.
- Goerman, P., Meyers, M., & Garcia Trejo, Y. (2018). Survey translation and the place of expert review in the question evaluation tool kit. Presentation given at the Federal Committee on Statistical Methodology (FCSM) Research and Policy Conference. Washington, D.C. March 7 – 9, 2018
- Groves, R. M., Fowler, Floyd J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley and Sons.
- Hansen, S. E., & Couper, M. P. (2004). Usability testing to evaluate computer-assisted instruments. In S. Presser, J. M. Rothgeb, M. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 337–360). Hoboken, NJ: John Wiley & Sons.
- Harkness, J. A. (2008a). Round 4 ESS translation strategies and procedures. Retrieved from http://www.europeansocialsurvey.org/docs/round4/methods/ESS4_translation_guidelines.pdf

- Harkness, J. A. (2008b). Comparative survey research: Goals and challenges. In E. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 56–77). New York, NY: Psychology Press Taylor & Francis Group.
- Harkness, J. A., Pennell, B.-E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. Rothgeb, M. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 453–473). Hoboken, NJ: John Wiley & Sons.
- Lessler, J., & Forsyth, B. (1996). A coding system for appraising questionnaires. In N. Schwartz, & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.
- Pan, Y., & de la Puente, M. (2005). *Census Bureau guidelines for the translation of data collection instruments and supporting materials: Documentation on how the guideline was developed*. Research Report Series #2005-06. Statistical Research Division, Washington, DC: U.S. Bureau of the Census.
- Pan, Y. (2009). *A process for reviewing translations of data collection instruments and related materials*. U.S. Census Bureau Survey Methodology Study Series #2009-09. Retrieved from <https://www.census.gov/srd/papers/pdf/ssm2009-09.pdf>
- Pan, Y., Sha, M. M., Park, H., & Schoua-Glusberg, A. (2010). 2010 Census language program: Pretesting of Census 2010 questionnaires in five languages. Research Report Series (Survey Methodology #2009-01) Retrieved from <http://www.census.gov/srd/papers/pdf/rsm2009-01.pdf>
- Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? In P. Marsden (Ed), *Sociological methodology* (pp. 73–104). San Francisco: Jossey-Bass.
- Ramirez, C. (2002). Strategies for subject matter expert review in questionnaire design. International Conference on Questionnaire Development, Evaluation and Testing Methods (QDET), Charleston, SC.
- Tourangeau, R. (2004). Experimental design considerations for testing and evaluating questionnaires. In S. Presser, J. Rothgeb, M. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 209–224). Hoboken, NJ: John Wiley & Sons.
- U.S. Census Bureau Statistical Quality Standard A2: Developing data collection instruments and supporting materials (2015). Retrieved from <https://www.census.gov/about/policies/quality/standards/standarda2.html>
- U.S. Office of Management and Budget (2016a). *OMB Statistical Policy Directive No. 2 Addendum: Standards and Guidelines for Cognitive Interviews*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/directive2/final_addendum_to_stat_policy_dir_2.pdf
- U.S. Office of Management and Budget (2016b). *OMB Statistical Policy Working Paper 47: Evaluating Survey Questions: An Inventory of Methods*. Statistical and Science Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Retrieved from <https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/04/spwp47.pdf>

- Willimack, D. K., Lyberg, L., Martin, J., Japac, L., & Whitridge, P. (2004). Evolution and adaptation of questionnaire development evaluation, and testing methods for establishment surveys. In S. Presser, J. Rothgeb, M. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 385–407). Hoboken, NJ: John Wiley & Sons.
- Willis, G., & Lessler, J. T. (1999). *The BRFSS-QAS: A guide for systematically evaluating survey question wording*. Rockville, MD: Research Triangle Institute.
- Willis, G. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications, Inc.
- Willis, G. B., Kudela, M. S., Levin, K. L., Norberg, A., Stark, D. S., Forsyth, B. H., ... Hartman, A. M. (2010). Evaluation of a multistep survey translation process. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 137–152). Hoboken, NJ: John Wiley & Sons.

Measuring Migrants' Educational Attainment: The CAMCES Tool in the IAB-SOEP Migration Samples

Silke L. Schneider, Roberto Briceno-Rosas, Verena Ortmanns, & Jessica M. E. Herzing

Abstract

Education is one of the most frequently used variables in social science research. However, it is challenging to measure educational attainment with a high degree of validity and comparability in migrant surveys. In migrant surveys, respondents were educated in various different educational systems. Rather than providing specific response options for the qualifications available in every country of origin, migrant surveys often use generic response options (such as “secondary education”) that supposedly work equally well for respondents educated in all kinds of educational systems. Given the lack of universal understanding of such generic categories, we have doubts whether this approach leads to reliable, valid, and comparable data. To improve the measurement of educational qualifications obtained abroad, GESIS has developed a new tool in the project “Computer-assisted measurement and coding of education in surveys” (CAMCES). In this paper, we present how migrants’ education is usually measured in the German IAB-SOEP Migration Samples and the alternative measurement using the CAMCES tool implemented in the IAB-SOEP Migration Samples 2015 and 2016. We analyze the coverage of educational systems in the CAMCES tool, compare the level of item nonresponse and non-coded responses affecting the standard and the CAMCES instruments, and examine the consistency of the resulting education variables. The paper concludes by discussing benefits and limitations of either measurement approach, and by giving an outlook of possible applications of the CAMCES tool.

1 Introduction

Educational attainment is a central socio-economic variable and captured in virtually all surveys of individuals (Smith, 1995). When surveys ask for the highest educational qualification obtained, they usually use a closed-ended question format that offers respondents a list of country-specific educational categories for self-classification. The advantage of this response design is that it is not very costly for researchers in terms of time and labor, and not very burdensome for respondents. However, this approach also brings along some challenges, especially in cross-national surveys (Braun & Mohler, 2003; Braun & Müller, 1997; Schneider, 2009). Each educational system has its unique idiosyncratic institutions,

and their certificates often have proper names which rule out translation (Schneider, Joye, & Wolf, 2016). The state-of-the-art for cross-national surveys is thus to use country-specific questionnaire items for education.

Depending on their age at migration, migrants have often obtained their educational qualifications in their country of origin. One measurement approach, especially in general population samples, is to ask respondents with foreign educational qualifications to indicate the “equivalent” qualification in the educational system in which the survey takes place. This is likely burdensome for migrant respondents, especially for those who did not (yet) have much contact with the educational system or labor market in their host society. Another approach is to offer a list of categories that are generic descriptions of educational levels that are assumed to be understood universally by respondents from various backgrounds. This is, however, a strong assumption: Language ability and cultural differences in the understanding of constructs underlying a survey question may introduce measurement error (Kleiner, Lipps, & Ferrez, 2015). For example, generic terms such as “primary” or “mandatory education” correspond to schooling of different durations in different countries, and “secondary” education includes vocational training in some countries but not in others. Neither measurement approach – using survey country education categories or using generic categories – takes into account institutional differences between educational systems across the world, and large measurement errors are likely.

In order to achieve acceptable measurement quality and thereby comparability across respondent groups from different origins, we argue that it is desirable to measure educational attainment using country-specific response categories in migrant surveys. The project “Computer-assisted measurement and coding of educational qualifications in surveys” (CAMCES, funded by the Leibniz Association and implemented at GESIS – Leibniz Institute for the Social Sciences), tackled the issue of measurement and coding of educational attainment in cross-cultural (including migrant) surveys. The goal of the project was the development of a tool for computer-assisted surveys that facilitates the measurement and coding of educational qualifications across countries. To achieve this goal, we designed a short questionnaire module asking respondents to, firstly, indicate where they were educated and then to choose their highest qualification from a list of educational qualifications specific to the respective educational system. The country-specific response options are generated via innovative interfaces to an underlying database of educational qualifications, which was also developed in the project. The database (as of October 2017) covers nearly all European educational systems, some neighboring countries, and the countries of origin of the largest migrant and current refugee groups in Germany.

This paper presents the implementation of the CAMCES tool in the German IAB-SOEP Migration Samples and the results from our evaluation of the resulting data. In the next section, we present and discuss the standard measurement instrument employed for migrants’ education in the IAB-SOEP Migration Samples and the alternative measurement procedure proposed with the CAMCES tool. In section 3, we describe the implementation of the tool in the IAB-SOEP Migration Samples and our evaluation strategy. Results of this evaluation are presented in section 4. We conclude with a discussion about the benefits and limitations of the proposed tool in comparison with the standard measure and an outlook on future developments.

2 Measurement of Foreign Educational Qualifications

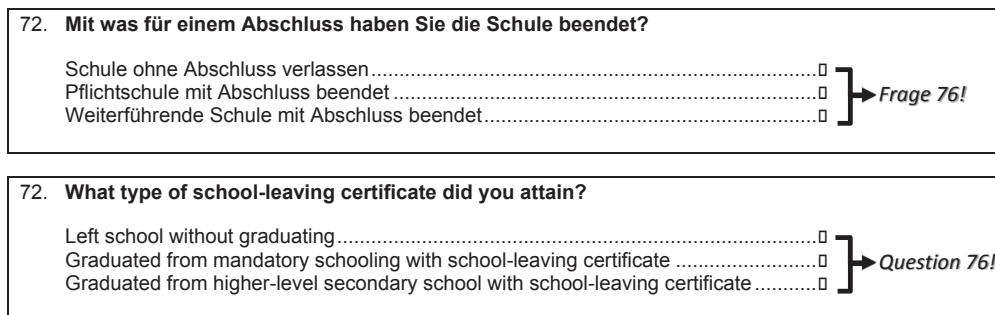
2.1 The IAB-SOEP Standard Instrument for Measuring Migrants' Educational Attainment

The IAB-SOEP Migration Samples are special samples of the German Socio-Economic Panel (SOEP) initiated by the German Institute for Economic Research (DIW) and the Institute for Employment Research (IAB) (Brücker et al., 2014). These special samples, which have been in existence since 2013, focus on migrants who have immigrated to Germany since 1995 and their children. They have a panel design and respondents are interviewed annually. All face-to-face interviews are conducted in German by trained interviewers using laptops with the survey software. The whole questionnaire has been translated into English, Polish, Turkish, Romanian, and Russian, and the translated versions are used to support respondents with weak knowledge of the German language (TNS Infratest Sozialforschung, 2014).

The IAB-SOEP Migration Samples measure educational attainment of migrants using generic response options that are not adapted to the origin of the respondents. This generic instrument is strongly inspired by the German educational system. Firstly, in terms of format, following common practice in German surveys, it measures educational attainment using two questions: The first question asks about the highest school-leaving certificate and the second one asks about the highest post-school qualification obtained by the respondent, i.e. from vocational training or higher education. This division of the education question is not common in other countries, where general schooling and post-school education are often covered in just one item.

Secondly, in terms of content, the IAB-SOEP Migration Samples use response categories reminding us of the German educational system. However, rather than explicitly mentioning the names of German qualifications, they describe typical German education categories in vague terms, in an attempt to be more universally understood and translatable (see Figure 1 and Figure 2).

This approach to measuring education may be problematic because the meaning of education-related terms is highly contextual. For example, the term *Pflichtschule* (“mandatory schooling”) may not be understood in the same way by all respondents because the length of mandatory schooling depends on the country. For instance, mandatory schooling has a duration of 5 years in Bangladesh, 6 years in Iraq, 8 years in Croatia, and 9 years in Greece (The World Bank, 2017). Also, some countries do not have mandatory schooling, or if it exists but is not enforced, respondents may not know what it actually is in their country of origin. What is probably meant with the term mandatory schooling is what corresponds to the completion of mandatory school in Germany, i.e. the completion of general lower secondary education after 9 or 10 years of schooling (equivalent to the German school-leaving certificate of *Hauptschule*). Similar problems occur with the term *Weiterführende Schule* (“higher-level secondary school” – note that the English translation is clearer than the German original in this case), as there are several certificates for the completion of secondary education in Germany, only some of which grant access to tertiary education.



Source: TNS Infratest Sozialforschung, 2014a and 2014b

Figure 1 IAB-SOEP question on respondents' foreign school-leaving qualification (German original and translation into English)

There are similar issues with the categories in the second item, which measures vocational and tertiary education, i.e. post-school education (see question 83 in Figure 2). These categories mirror central elements of the German vocational and higher education systems, which may not be relevant and rather confusing for migrant respondents. For example, different types of vocational trainings are presented to respondents: *im Betrieb angelernt* (“in-house training at a company”), *eine längere Ausbildung im Betrieb gemacht* (“completed an extended apprenticeship at a company”), and *Besuch einer berufsbildenden Schule* (“attended a vocational school”). It might be difficult for respondents to understand the meaning of these categories without knowing the German system of vocational training. Respondents might have problems in distinguishing “in-house training” and “extended apprenticeship at a company”. The latter category also lacks specification regarding the duration implied by the term “extended”. While the question designers might have had the duration of German apprenticeships in mind, ranging from two to four years, respondents may regard vocational training of six months or a year already as “extended”.

With respect to the higher education categories, the categories remind us of the German *Fachhochschule* and the classical university. Such a subdivision does not exist in many countries and the differentiation may thus be understood in unintended ways.

Finally, the question misses a response category for advanced vocational training, which exists in Germany (e.g. *Meister* or *Techniker*) and many other countries. If respondents with such qualifications choose vocational training, their educational attainment will be underestimated, and if they choose one of the higher education categories (likely “university / college with a more practical orientation”), it will be overestimated.

83. Was für eine Ausbildung war das? In welchem Jahr haben Sie diese Ausbildung beendet?

☛ *Mehrfachnennungen möglich!*
 ☛ *Bitte machen Sie auch eine Jahresangabe, wenn Sie die Ausbildung nicht abgeschlossen haben.*

	Ja	Jahr
Ich wurde in einem Betrieb angelehrt.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Ich habe in einem Betrieb eine längere Ausbildung gemacht.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Ich habe eine berufsbildende Schule besucht.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Ich habe eine Hochschule / Universität mit eher praktischer Ausrichtung besucht.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Ich habe eine Hochschule / Universität mit eher theoretischer Ausrichtung besucht.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Ich habe ein Promotionsstudium absolviert.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Sonstige Ausbildung.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

83. What kind of education or training was it? When did you complete this education or training?

☛ *Multiple answers possible!*
 ☛ *Please state the year even if you did not complete the training.*

	Yes	Year
I received in-house training at a company.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
I completed an extended apprenticeship at a company.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
I attended a vocational school.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
I attended a university / college with a more practical orientation.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
I attended a university / college with a more theoretical orientation.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
I completed doctoral studies.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Other.....	<input type="checkbox"/> →	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Source: TNS Infratest Sozialforschung, 2014a and 2014b

Figure 2 IAB-SOEP question on respondents' foreign post-school education (German original and translation into English)

As this discussion shows, designing universally understandable response options for the education question in cross-cultural surveys is a minefield.⁹ In general, dealing with these ambiguous terms in these two questions, regardless of language, can be difficult for migrant respondents. Even if well translated (which we did not check beyond the comparison of the German and English versions), these items are unlikely to generate comparable responses to the education questions, and thus answers cannot be unambiguously related to the German educational system or ISCED (UNESCO-UIS, 2012).

2.2 The CAMCES Tool as an Alternative Measurement Instrument

The CAMCES tool aims to alleviate some of the challenges associated with the measurement of educational attainment in surveys of migrants and other cross-cultural surveys, by considering different educational systems and referring to specific educational qualifications. The tool combines (1) a short questionnaire module, (2) an international database of educational qualifications and their classification, and (3) two survey interfaces that dynamically show relevant response options from the database in the questionnaire (combination box and search tree). The tool thereby offers response categories relevant for the specific educational systems in which respondents completed their education.

Two versions of the CAMCES interfaces were developed to support different survey modes: a CAPI version for computer-assisted personal interviews and a CAWI version for web surveys. With regard to the CAPI version, the interviewer needs to hand over the device to respondents educated abroad who are then asked to report their educational qualification by themselves. This is necessary, firstly, because the interviewer may not speak the respective language and, secondly, because the interfaces are, due to their dynamic nature, more suitable for self-completion. This paper focuses on the CAPI version because the IAB-SOEP Migration Samples are conducted as CAPI surveys.

2.2.1 The Questionnaire Module

The CAMCES questionnaire module has, as its core, the question asking for respondents' educational attainment but it also includes questions to identify the relevant educational system and some optional questions. The whole questionnaire module was developed in German (for Germany) and has been translated into English (for the UK), French (for France), Spanish (separate versions for Spain and Venezuela), and Dutch (for the Netherlands) using the team approach (Harkness, 2003; Harkness, van de Vijver, & Johnson, 2003; Harkness, Villar, & Edwards, 2010).¹⁰ The questionnaire module consists of the following three main elements:

9 This is basically why a standard terminology, the International Standard Classification of Education (ISCED), was developed for international statistical comparisons. It is not directly useable in questionnaires, though, because it is not commonly understood in the intended way by respondents.

10 The final CAMCES questionnaire module can be downloaded in these languages from <https://www.surveycodings.org/education/question-module-measuring-educational-attainment>. The downloadable versions differ somewhat from the questionnaire implemented in the IAB-SOEP

- a) A short introductory text that defines the concept of formal education. It aims at generating a consistent interpretation of the subsequent questions by respondents. (This intro can be skipped if only the search tree interface (see section 2.2.3) is implemented.)
- b) A question that aims to identify the educational system in which the respondents obtained their highest educational qualification. The educational systems are defined in the database (see section 2.2.2).
- c) The actual question on the highest educational qualification obtained, including a number of instructions. The response options are, depending on the educational system mentioned previously, dynamically fed in from the database (see section 2.2.2) via the interfaces (see section 2.2.3).

Additionally, the module offers some optional questions asking for:

- a) The year in which the highest educational qualification was obtained.
- b) The number of years the respondent has spent in formal education in total to obtain a direct measure of the actual years of education.
- c) Multiple qualifications: The questionnaire module accounts for the possibility that surveys may want to ask respondents for more than their highest educational qualification. Respondents with multiple qualifications may not know which one is the highest, or what respondents assess as their highest qualification may not concur with the assessment of the researcher or with the criteria of an international education classification. Furthermore, respondents can have two qualifications classified in the same ISCED level, and ISCED does not specify any hierarchy, for instance between vocational and general qualifications within levels. Therefore, the question module contains optional questions and routing instructions for repeating (or “looping”) questions b) and c) for respondents to report multiple qualifications. The wording of these questions is then adjusted by not referring to the “highest” qualification any more from the second loop onwards.

For CAPI surveys, the CAMCES questionnaire module includes specific instructions for interviewers: They have to turn around the computer to allow respondents to report their educational qualification themselves by using the CAMCES interfaces (meaning a switch to CASI mode).

2.2.2 The Database

The second element of the CAMCES tool is the underlying database. The database consists of three major parts:

- a) It lists the educational systems, usually corresponding to countries. In multi-lingual countries like Belgium or Switzerland, the list further differentiates between the languages in which the educational qualification could have been obtained. In countries where educational systems differ markedly by region, such as Scotland compared to the

Migration Samples 2015 and 2016 since, firstly, they are not adapted for surveys of migrants, and secondly, we used the results reported in this paper for improving the questionnaires (especially instructions to respondents).

rest of the United Kingdom, the list is further broken down by region. This ensures that only relevant response options are shown to respondents. Currently, the CAMCES educational database (as of June 2018) contains 71 educational systems including country/language combinations (corresponding to 57 countries).

- b) The central part of the database is a table listing educational qualifications in the relevant languages, with detailed identification codes. If information is available on outdated qualifications, they are covered in the database, too. In addition, alternative expressions are included when several terms – e.g. official and colloquial – are used for the same qualification. The structure of the search tree (see section 2.2.3) is also defined in this part of the database. As of June 2018, the CAMCES database contains nearly 2500 unique educational qualifications and more than 2500 alternative expressions.
- c) Each educational qualification is linked with detailed codes for the International Standard Classification of Education (ISCED). The following versions of ISCED are covered: ISCED 1997 (OECD, 1999; UNESCO-UIS, 1997 [2006]), ISCED 2011 (UNESCO-UIS, 2012; UNESCO-UIS, European Union, & OECD, 2015), and an alternative version of ISCED 2011, developed by the CAMCES team.¹¹

For countries that are not (yet) covered in the database and where low numbers of immigrant groups are expected, we suggest using an open-ended question for respondents to name their qualification and using the optional question on years of education as a second indicator of educational attainment.

2.2.3 The Interfaces

The final “ingredient” of the CAMCES tool are the survey interfaces that make the database accessible to respondents while completing the questionnaire.¹² One can search the database either by typing (text string matching) or by using a structured list. The respective user interfaces for these tasks are the “combination box” and the “search tree”. By using these interfaces, more response categories can be offered than with a simple showcard.

The combination box allows respondents to dynamically search their educational qualification in the database by typing and then selecting the best-matching result (see Figure 4). If they do not find an adequate match, they can submit the text they have typed. It is thus a combination of a simple text box and a drop-down box as tested by Couper and Zhang (2016), which is why we call it a combination box. The entry part looks like the empty text field of a typical open-ended question. However, in order to signal to respondents that the text field actually includes a search box, it is preceded by a magnifier glass

11 The differences compared to the official ISCED codes are that, first, we developed valid codes for educational qualifications not documented in the official ISCED mappings (e.g. for outdated qualifications). Second, when we had reasons to doubt the official ISCED code, for instance after consultations with country experts (see also chapters in Schneider, 2008), we determined an unofficial code. The alternative ISCED codes are intended to improve cross-national comparability compared with official ISCED codes. Most of the codes are, however, identical to the official ISCED codes.

12 For more research on such interfaces, see Couper and Zhang (2016) who focus on lists of prescription drugs and Tijdens (2014, 2015) who focuses on lists of occupations.

and contains a watermark reading “search” in the language of the educational system in which the respondent reported to have been educated. In the instructions, respondents are asked to enter the response into the text box. The instructions also dynamically show the specific language in which the qualification should be entered (“Please enter the qualification in Croatian”, see Figure 3). With each letter that the respondent types, the response is matched with educational qualifications and alternative expressions in the database (text string matching). The results retrieved from the database are presented to the respondent below the text field. The number of results narrows further down as the respondent continues typing. Respondents can then select the best match from the resulting list of educational qualifications. If no matching educational qualification is found or if the respondent does not make a selection among the list of results, the typed text is saved as a text string, as it would be for a typical open-ended question.

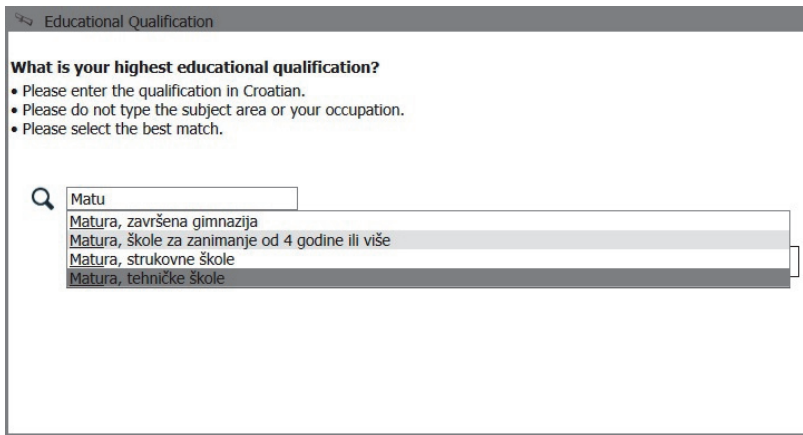


Figure 3 Interface “combination box” of the CAMCES tool version 1.0 (example shows qualifications for Croatia in an English-language survey environment)

The search tree (see Figure 4) is similar to showcards or response lists in standard educational attainment questions. However, a search tree can accommodate a larger number of response options because of its nested structure. For measuring occupations, Tijdens (2014, 2015) uses a three-level search tree. In order to be as manageable for respondents as possible, and because the number of educational qualifications per education system is not as high as the number of occupations in modern societies, we opted for a two-level search tree. The first level offers country-specific summary terms such as “secondary education” or “university degree”. When respondents click on a first-level entry, the search tree expands and reveals the different qualifications within this group of qualifications (for example Bachelor’s, Master’s and PhD degrees within the university degrees section).

The search tree is used as a fallback interface for respondents who do not make a valid selection using the combination box. This can happen, firstly, if respondents leave the combination box empty (item nonresponse); secondly, if the text entered does not generate any matches with the database, or thirdly, if respondents enter text that generates matches, but do not select any of the matches offered. If the script of the language of the

selected educational system does not correspond to the script of the keyboard used during the survey, for instance, when interviewing Russian or Greek migrants in a CAPI survey in Germany, only the search tree interface is used. For this purpose, the database includes, in the table identifying educational systems, also the scripts used in the respective countries and languages. This ensures that the search tree interface is shown rather than the combination box if the script used in the educational system deviates from the one used in the questionnaire.

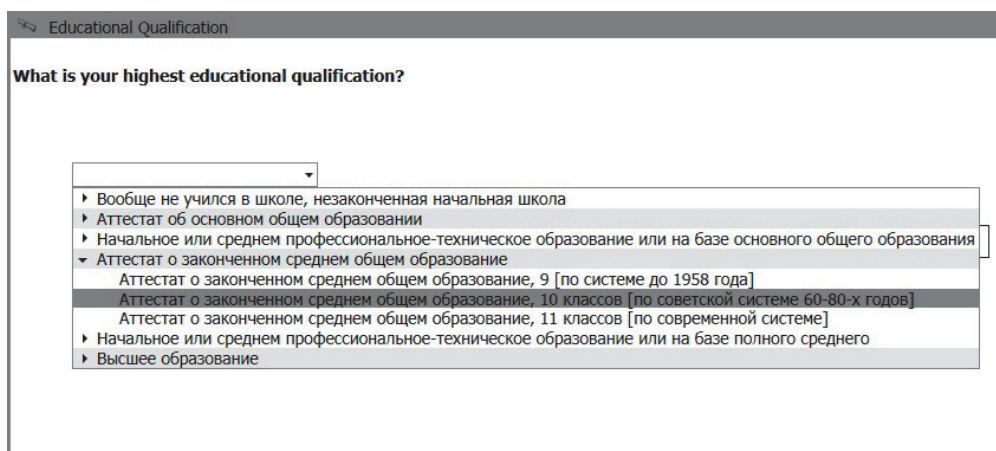


Figure 4 Interface “search tree” of the CAMCES tool version 1.0 (example shows qualifications for the Russian Federation in an English-language survey environment)

When respondents select a qualification from the combination box or from the search tree, the respective detailed identification code for the qualification is saved in the survey data set. After data collection, this detailed code can be recoded to education classifications such as ISCED (see section 2.2.2) via syntax files provided with the CAMCES tool.¹³

For the CAPI version, the education question itself (element c) in section 2.2.1) including the response interfaces were programmed as a separate piece of software. This piece of software can be loaded by the CAPI software and the resulting code can be fed back to the CAPI software. That is, the identification code of the relevant educational system resulting from the respective questionnaire item (see element b) in section 2.2.1), is transmitted from the CAPI software to the CAMCES software. The detailed qualification code as well as information about the type of interface that has generated the code is then transmitted back to the CAPI software. The CAMCES software also allows survey agencies or researchers to specify the question text, instruction text, colors, etc., and it generates log files with paradata such as time stamps for each respondent.¹⁴

13 See <https://www.surveycodings.org/education/data-processing>.

14 The full documentation is available at <https://www.surveycodings.org/education>.

3 Implementation of the CAMCES Tool in the IAB-SOEP Migration Samples and its Evaluation

3.1 Implementation in the IAB-SOEP Migration Samples in 2015 and 2016

The CAMCES tool was implemented in the initial IAB-SOEP Migration Sample in 2015 and in the refresher sample in 2016 on a trial basis. The initial sample consists of people who migrated to Germany between 1995 and 2011 and who were first interviewed in 2013 (Kroh, Kühne, Goebel, & Preu, 2015 - also referred to as M1). The refresher sample consists of people who migrated to Germany between 2009 and 2013 and who were first interviewed in 2015 (Kühne & Kroh, 2017 - also referred to as M2). In the first interviews of the respective samples M1 and M2 in 2013 and 2015, the standard education measure (see section 2.1) was implemented. In the second interviews in 2015 and 2016, the CAMCES questionnaire module was placed at the very end of the questionnaire, and all respondents who reported to have obtained a foreign vocational or higher education qualification were routed to it. In 2015, 945 out of the responding 989 migrants (96%) and in 2016, 637 out of 659 (97%) reported to have completed a vocational or higher educational qualification abroad. The same version of the CAMCES CAPI software and database was implemented in the IAB-SOEP Migration Samples in 2015 and 2016.¹⁵

The CAMCES questionnaire module used in the IAB-SOEP Migration Samples somewhat differs from the version described in section 2.2.1, which is not specifically designed for migrant surveys (e.g. the IAB-SOEP version of the CAMCES questionnaire module only asks about *foreign* qualifications, whereas the more general CAMCES version asks about all educational qualifications). We used the CAMCES questionnaire module with looping (see point f) in section 2.2.1). The most important educational systems to be covered by the CAMCES database were agreed with the SOEP team in advance (mostly countries of origin of the largest migrant groups in Germany); they were tailored to the initial sample (M1) with no addition of countries for the refresher sample (M2).

There are a number of differences worth noting between version 1.0 of the combination box interface shown in Figure 4, and the version implemented in the IAB-SOEP Migration Samples, resulting from optimizations in response to the results reported later in this paper: While version 1.0 dynamically adapts the instruction to report the educational qualification in the language of the respective educational system (“in Croatian” in Figure 4), the version implemented in the Migration Samples in 2015 and 2016 generically asked to “enter the qualification in the original language”. The other instructions also changed. Furthermore, the version fielded in the Migration Samples did not show the magnifier glass and did not contain the watermark reading “search” in the language in which the response needed to be entered. The text string matching algorithm has also improved since 2016. In order to minimize missing data, the search tree interface did not allow continuing the

15 The current version 1.0 of the tool was implemented in the BAMF-IAB-SOEP Refugee Samples M3 and M4, first interviewed in 2016, and in 2017. At the time of writing, this data could not yet be analyzed.

survey without making a valid selection, which respondents could only avoid by clicking a button “other qualification” in the interface.

3.2 Evaluation of the CAMCES Tool in the IAB-SOEP Migration Samples

In our analyses, we aimed to evaluate how the CAMCES tool generally worked, how the resulting data compared to the standard IAB-SOEP Migration Samples education measures, and which interface of the tool (combination box or search tree) was the most adequate for surveys with a sample of migrants. The analysis steps in detail were:

- a) Firstly, we determined to what extent the different educational backgrounds of the migrants in the sample were covered by the CAMCES database (country coverage).
- b) Secondly, we compared the level of item nonresponse (missing data) between the standard measure (see section 2.1) and the measures generated by the different CAMCES interfaces.
- c) Thirdly, we analyzed in more detail responses generated by the CAMCES tool not producing a valid qualification code and, thus, potentially requiring manual post-coding.
- d) Fourthly, we analyzed whether the educational qualification reported to be the highest qualification by respondents (i.e. the first loop, where the questionnaire explicitly asked about the highest qualification) actually was the highest qualification of all qualifications reported by the respondent in any loop.
- e) Finally, we evaluated the content of the data generated by the tool by comparing it to the IAB-SOEP Migration Samples standard measure (see section 2.1). We examined to what extent the CAMCES tool produced data that was consistent with previous measurements of educational attainment of the same respondents (this is possible due to the panel design).

4 Results

4.1 Coverage of the Relevant Educational Systems in the CAMCES Database

We started by analyzing the extent to which the tool was able to cover the diversity of the educational backgrounds of migrants in the survey. Table 1 shows how many respondents were covered by the CAMCES tool in each subsample, and which interface they were routed to, based on the script relevant for the educational system they indicated to have been educated in.¹⁶ The results clearly indicate differences in the composition of the samples. Substantially more respondents' educational systems were covered by the CAMCES

16 If a respondent was educated in an educational system not yet covered by the database, information was gained in simple text form in response to an open-ended question so that post-coding would be required. This data was not analyzed for the purposes of this paper.

database in 2015 (M1) than in 2016 (M2). The refresher sample M2, focusing specifically on recent immigrants, thus required an extension of the database, which has been completed in the meantime. A higher proportion of respondents was routed to the combination box in 2016 than in 2015.

Table 1 Coverage of educational systems in CAMCES database and interface routing

	2015 (M1)	2016 (M2)	Total
Total respondents responding in respective panel wave	989 100%	659 100%	1648 100%
of which: Respondents with foreign vocational qualification	945 96%	637 97%	1582 96%
of which: Country covered by CAMCES database	818 87%	448 70%	1266 80%
of which: Respondents routed to combination box	410 50%	340 76%	750 59%
Respondents routed to search tree	408 50%	108 24%	516 41%

Figure 5 indicates in which countries how many respondents have received their highest educational qualifications in the samples in 2015 and 2016. It shows many educational systems in which only a few respondents were educated, which is proof of the diversity of educational backgrounds of migrants in Germany. The change in interface usage reported above is caused by the relatively lower number of migrants from the former Soviet Union (to whom the search tree was offered with Cyrillic script) and the relatively higher number of migrants from other European countries using mostly Latin script in the refresher sample (M2, 2016) compared to the initial sample (M1, 2015).

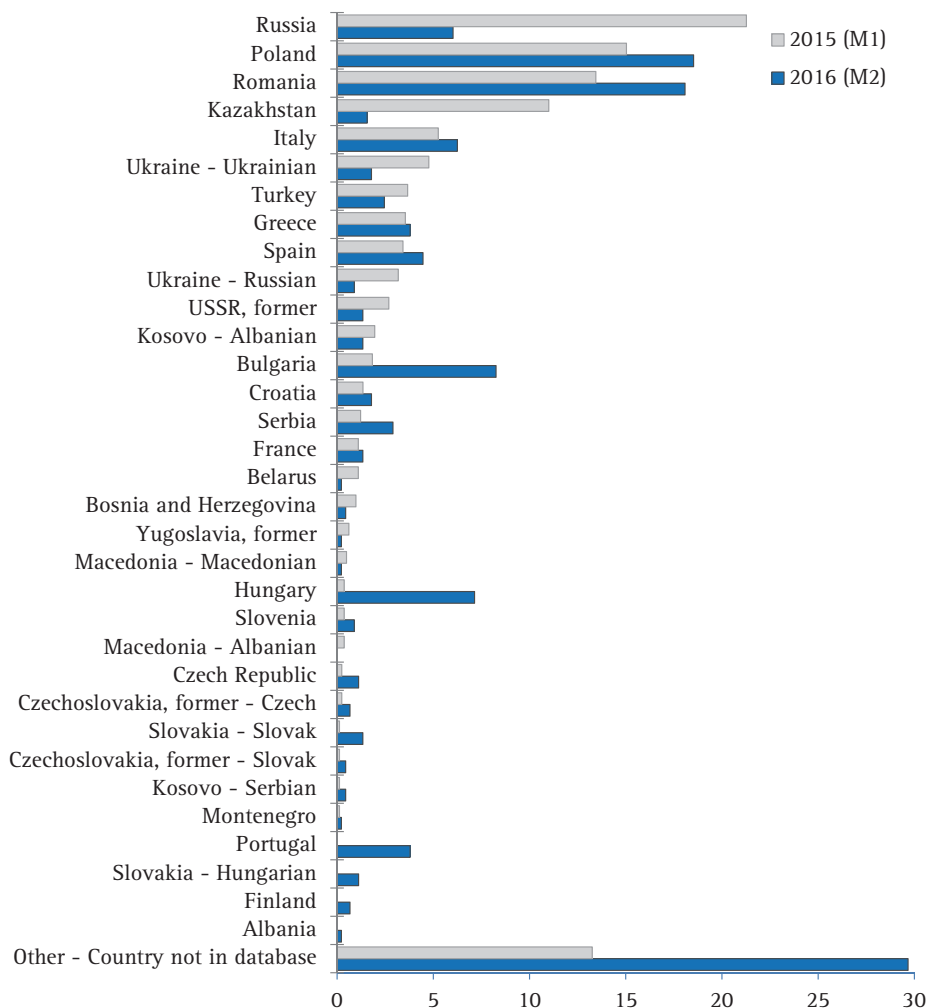


Figure 5 Countries in which respondents obtained their highest educational qualification in the 2015 and 2016 Migration Samples (%)

4.2 Analysis of Item Nonresponse and Non-Coded Responses

In this section, we compare item nonresponse and non-coded responses produced with the CAMCES tool with the corresponding outcomes in the standard IAB-SOEP Migration Samples instrument. First, we cover cases with no response to the education question using the combination box (item nonresponse). Second, we look into cases with some response but which could not be coded automatically (non-coded responses).

4.2.1 Item Nonresponse

In Figure 6, we see that both standard education items in the IAB-SOEP Migration Samples had a very low proportion of missing data due to item skipping (less than 1%). We compared these results with the CAMCES measurements using the combination box in the first loop.¹⁷ Item nonresponse resulting from the combination box was very high, with 20% (83 cases) in 2015 and 37% (125 cases) in 2016. The search tree did not allow continuing the survey without valid selection, unless respondents indicated to have “another” qualification, which is analyzed in the following section.

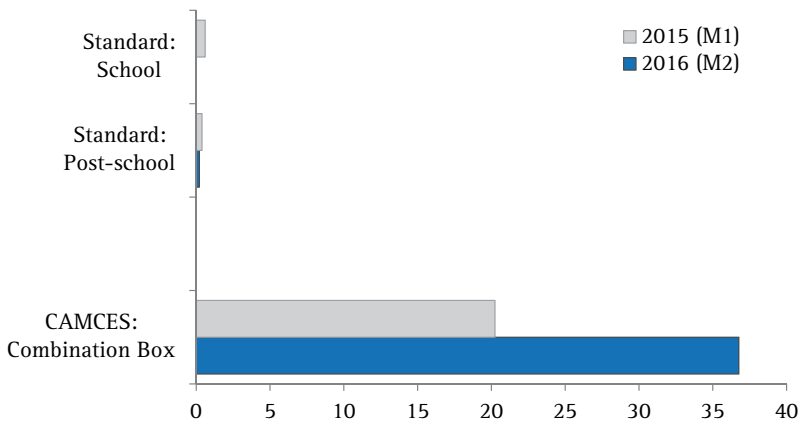


Figure 6 Item nonresponse for the standard measures and the CAMCES combination box in %

In order to understand the differences in item nonresponse between the standard items and the combination box, it is important to consider the characteristic of each response design. In the standard items, response categories were read out by the interviewer, while in the combination box respondents needed to type a letter or word themselves in order to receive response suggestions. Because of the higher response burden, we can expect satisficing and thus higher nonresponse for the combination box. It is also known that open-ended questions – and the combination box is similar to an open-ended question from the respondents’ point of view – produce more nonresponse than closed-ended questions (Reja, Lozar Manfreda, Hlebec, & Vehovar, 2003). Since the search tree was offered as a fallback to all those not generating a valid code when using the combination box, no missing data was eventually generated with the CAMCES instruments. However, this result puts the usefulness of the combination box, at least in interviewer-administered surveys, into question (see section 5 for a discussion of this point).

¹⁷ Respondents who reported having their highest educational qualification from a country that was not covered by the CAMCES database (see section 4.1), or had missing data on that question, were excluded from this and all further analyses (127 cases in 2015, and 189 cases in 2016). Further, the analyses do not include respondents who did not get the interfaces due to a technical error in the implementation of the tool (3 cases in 2015 and 4 in 2016).

A possible explanation for the increase of 17 percentage points in item nonresponse from 2015 to 2016 in case of the combination box might be the interviewers (for an overview on interviewer effects see Blom & West, 2017), which overlap to some degree across survey waves. The interviewers may have learned in the 2015 survey that the combination box is followed by the search tree if no valid response is given. In order to speed up the interview (remember that the CAMCES tool was implemented at the very end of the questionnaire when both interviewer and respondent may have been tired), the interviewer might have asked the respondent to skip the combination box, or might have delayed the switch to the CASI mode accordingly.

4.2.2 Responses not Automatically Coded

We now turn to responses that could not be coded automatically by the tool, because no valid entry was selected from the database. For the combination box, this meant that the respondent wrote text into the search field without subsequently selecting any of the database entries offered as a response (or there were no matching database entries). For the search tree, this meant that respondents clicked the button “other”. Non-coded responses in CAMCES are comparable with the response category “other” in the IAB-SOEP standard instrument as the respondents provided an answer but this answer required post-coding.

In the IAB-SOEP Migration Sample 2015, the CAMCES tool allowed to automatically code 91% of the educational qualifications reported by respondents as their highest qualification (741 cases). In 2016, the results were, with 93%, very similar. As in the case of item nonresponse, the search tree as a fallback for those who did not make a valid selection in the combination box, reduced the final amount of not automatically coded cases substantially. Figure 7 shows the proportion of not automatically coded cases broken down by interface and compared to the standard item.¹⁸ All instruments yielded lower non-coded responses in 2016 than in 2015. Across response formats, we observe that the pattern is similar to the pattern regarding item nonresponse presented above. The IAB-SOEP standard item on post-school education produced the lowest proportion of “other” responses (between 2 and 3%). The search tree produced up to 6 to 10% of non-coded responses. The combination box had, by far, the largest proportion of not automatically coded responses (52% in 2015 and 35% in 2016).

18 The standard question on the school-leaving certificate (the first of the two education questions) is not included in the comparison because it does not have an answer option “other certificate”.

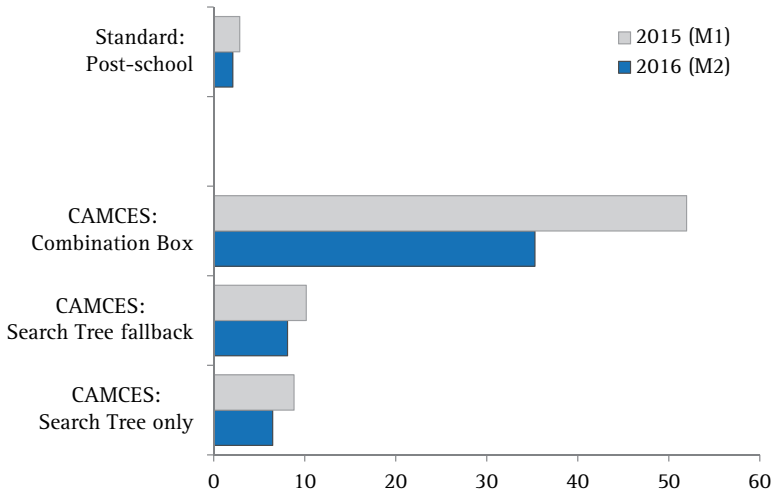


Figure 7 Responses not automatically coded or “other” answers per interface in comparison to the standard item in %

4.2.3 Reasons for Non-coding using the Combination Box

The large proportion of non-coded answers for the combination box requires a closer look. An advantage of the combination box is that it records respondents’ verbatim answers. In this section, we analyze their text entries in a qualitative way to understand why there was no automatic coding in so many cases. We developed a coding scheme identifying the six most plausible reasons for the failure of automatic coding. In Figure 8, we present the results. The responses could be coded into multiple categories because there could be several plausible reasons per response.¹⁹

19 For this analysis, we looked at data from all loops rather than the first loop only, so that cases refer to responses rather than respondents. Then, 35% (286 cases) of the combination box entries in 2015 and 33% (149 cases) in 2016 contained information not automatically coded.

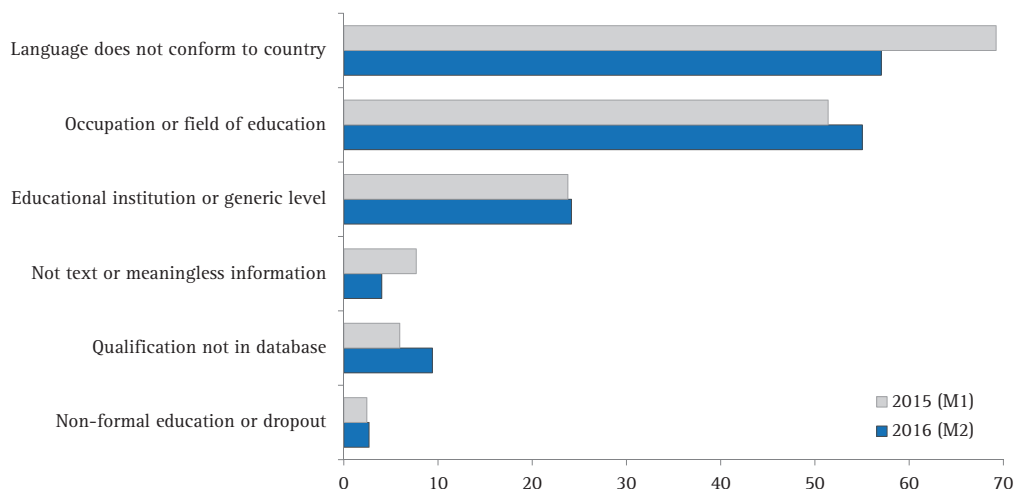


Figure 8 Reasons for non-coded responses of the combination box in %

In 2015, around 70% (198 cases) of the non-coded entries contained text in German rather than the language implied by the educational system the respondent reported to have been educated in. This was by far the most common reason why the automatic coding using the combination box failed. Responding in German (e.g. typing “Abitur” in the combination box) did not lead to automatically coded responses because the search algorithm limited the search in the database to the country and language in which respondents had obtained their qualification. Therefore, German qualifications did not appear on the list of proposed qualifications if the respondent previously had indicated that she or he had been educated, for instance, in Italy. This is intended behavior of the CAMCES tool because we do not want respondents to report the German qualification they think is equivalent to their foreign qualification but the foreign qualification itself. Apparently, this intention was not made clear enough to respondents (and/or interviewers): In the version of the CAMCES tool implemented in 2015, only a general hint in the instruction indicated to “enter the qualification in the original language”. In 2016, we improved the interviewer instructions to emphasize that respondents should enter the educational qualification themselves (i.e., switch to CASI mode) and that respondents should use the language of the country in which they had obtained their education. This resulted in a reduction of entries in German to 85 out of 149 (57% of non-coded responses), but it may also explain some of the increase in item nonresponse between 2015 and 2016 because respondents may, rather than overlooking the instruction, not have been able to remember and report their education in the original language (see section 4.2.1). As 57% is still high, we subsequently improved the interface of the combination box by implementing a watermark with the term “search” and improving the instructions (see section 2.2.3).

Further reasons for not automatically coded entries were that about half of the respondents in 2015 (148 cases) and 2016 (82 cases) reported their field of education and training, or their occupation, rather than their highest educational qualification. This indicates that

respondents have difficulties differentiating between the concepts of educational attainment, field of education, and occupation. Therefore, for the 2017 data collection, we added a new instruction in the question: “Please do not type the subject area or your occupation.”

Further issues were that the responses were too vague to generate a match in the database, and, in rather few cases (22 in 2015 and 6 in 2016) contained text without meaning. In 17 cases in 2015 and 14 cases in 2016, respondents typed an alternative expression for a qualification covered by the database, or a qualification that had not been listed in the database yet. Those entries were subsequently added to the database. Very few entries (7 cases in 2015 and 4 cases in 2016) referred to non-formal education or an education program the respondent dropped out of, which were intentionally not covered by the CAMCES database.

4.3 Usage of Loops

In addition to their highest foreign educational qualification, respondents were asked to report any further foreign educational qualifications. This feature was implemented to test whether respondents’ assessment of their highest educational qualification would actually match the highest educational qualification as defined in the ISCED classification. This allowed us to evaluate whether substantial improvements could be made by asking for further educational qualifications using loops, or if it is sufficient to just ask about the highest qualification.

In Figure 9, we present the distribution of the ‘actual’ highest educational qualification (rather than the one reported to be the highest qualification by respondents) across the four loops and across interfaces. In 2015, 23% of the respondents (220 cases) reported two or more qualifications. Within this group, we observed that only 65% (124 cases) reported their highest educational qualification in the first loop, in which we explicitly asked for the highest educational qualification. The proportion decreased with each loop: 32% of the respondents (60 cases) reported their highest qualification in the second loop, and 3% (6 cases) reported their highest qualification in the third loop. We see a similar picture in the data from 2016. In this wave, 30% (193) of the respondents reported two or more qualifications.

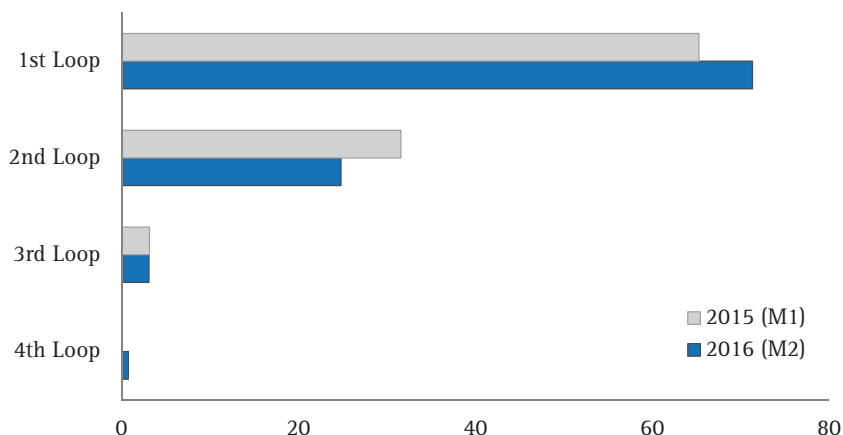


Figure 9 Loop in which the actually highest educational qualification was reported, in %

These results indicate that respondents’ assessment of what is their highest educational qualification does, in a substantial number of cases, not match what would be regarded as their highest qualification when coded in ISCED.²⁰ This means that asking only for the highest qualification is probably insufficient for reliably capturing the highest educational qualification of respondents based on the ISCED classification when using the CAMCES tool. Therefore, we recommend that surveys using the CAMCES tool provide at least two loops when measuring the educational attainment of respondents to reduce this source of error.

4.4 Comparison of the CAMCES and Standard IAB-SOEP Measures

A very interesting question is to what degree the two different measurement instruments lead to consistent data. We thus next compare the data resulting from the standard IAB-SOEP instrument (see section 2.1) and the CAMCES tool (see section 2.2) for all cases with valid data using both measures (722 cases in 2015 and 414 cases in 2016).²¹

20 Nonetheless, it is important to highlight that in both samples, the proportion of respondents reporting a single qualification was much higher than the proportion of respondents reporting multiple qualifications. This means that the relevance of the loops for measuring the highest educational qualification is lower when considering all respondents. We can only speculate as to whether those reporting only one qualification actually only have one, or whether they omitted reporting their further qualifications (or were motivated by interviewers to do so), for instance, in order to shorten the interview. This is quite likely, as recent research shows (Eckman & Kreuter, 2018).

21 The way how education was coded for the generic IAB-SOEP measure is shown in Table 4 in the appendix. In the case of the CAMCES measures, we used the highest educational qualification measured irrespective of the loop. Table 5 in the appendix shows how the detailed education codes resulting from the CAMCES tool were recoded to a scheme close to, but a bit more informative than, the one used for the IAB-SOEP measure. We based our codes on ISCED 2011 (UNESCO-UIS, 2012).

On the one hand, one could expect consistency to be an indicator of the validity of the new CAMCES-based measurement. On the other hand, we do not expect that the standard and the CAMCES measures agree completely. After all, doubts about the validity of the standard measure (see section 2.1) motivated the development of the CAMCES tool. Therefore, we will discuss the areas in which we find plausible reasons for under- or over-reporting of educational attainment using either measure. However, we should keep in mind that the reliability of education measures, as for any survey measure, is not perfect (Porst & Zeifang, 1987), and even applying the IAB-SOEP or CAMCES measures twice would lead to inconsistencies.

This section will, firstly, look at the distribution of each variable and then at the joint distribution. As Table 2 shows, the distribution of education differs substantially across measures for some categories. The CAMCES measure finds substantially more respondents with vocational upper secondary (ISCED 3) education, and substantially fewer with post-secondary non-tertiary (ISCED 4) education. ISCED level 4 is in most countries very small (e.g. OECD, 2017, Indicator A1, showing an OECD average of 5%), which adds credibility to the CAMCES measure. The IAB-SOEP measure cannot identify short-cycle tertiary education (ISCED level 5) because no generic response option corresponding to this level was provided (see section 2.1).

Table 2 Distributions of education derived from standard IAB-SOEP and CAMCES measures

	IAB-SOEP measure		CAMCES measure		
	Frequency	%	Frequency	%	% coded like IAB-SOEP measure
ISCED 0	20	1.8	5	0.4	1.3
ISCED 1			10	0.9	
ISCED 2	137	12.1	122	10.7	10.7
ISCED 3 vocational	176	15.5	237	20.9	20.9
ISCED 3 general	179	15.8	148	13.0	13.0
ISCED 4	191	16.8	94	8.3	8.3
ISCED 5	0	0	108	9.5	9.5
ISCED 6			87	7.7	
ISCED 7	409	36.0	306	26.9	34.6
ISCED 8	24	2.1	19	1.7	1.7
Total	1136	100.0	1136	100.0	100.0

Table 3 reports the joint distribution of both measures in a cross tabulation and allows a more thorough analysis of their correspondence. Shaded cells show which codes conceptually correspond across the two measures, i.e. where the most overlap of responses should be found.

Table 3 Cross-tabulation of ISCED derived from CAMCES and IAB-SOEP measures (row percentages)

	ISCED derived from IAB-SOEP standard measure								Total
	0-1	2	3 vocational	3 general	4	5	6-7	8	
0	0.0	20.0	20.0	40.0	0.0	0	20.0	0	100
1	10.0	30.0	20.0	30.0	10.0	0	0	0	100
2	5.7	35.2	27.0	12.3	17.2	0	2.5	0	100
3 vocational	2.5	14.3	24.9	24.5	28.7	0	5.1	0	100
3 general	2.7	16.9	14.9	33.1	17.6	0	14.2	0.7	100
4	2.1	7.4	22.3	12.8	39.4	0	14.9	1.1	100
5	0	12.0	28.7	10.2	26.9	0	22.2	0	100
6	0	4.6	2.3	9.2	3.4	0	79.3	1.1	100
7	0	2.3	1.3	6.9	2.0	0	84.6	2.9	100
8	0	0	5.3	0	0	0	31.6	63.2	100
Total	1.8	12.1	15.5	15.8	16.8	0	36.0	2.1	100

All in all, only 49% of cases are classified consistently across measures. Interestingly, there is no upward or downward bias in the distribution of education resulting from the CAMCES measure, leading to a higher ISCED code than the IAB-SOEP measure in 25% of the cases, and leading to a lower code in 26% of the cases.

We find the highest mismatch between measures for the lowest education category: None of the cases coded as ISCED 0 (less than primary education) and 10% coded as ISCED 1 (primary education) using the CAMCES measure are coded in ISCED 0-1 in the IAB-SOEP measure. The case numbers are very low though and measurement error probably rather high here – these respondents are not actually expected to get into the CAMCES module because the questionnaire routing only routed respondents reporting foreign vocational or higher education qualifications into the module, which are as a rule higher than ISCED level 1. There must be measurement error either in the routing question or in either or both of the education measures.

Regarding ISCED level 2 (lower secondary education), a large proportion (27%) of respondents coded here using the CAMCES tool are coded as vocational ISCED 3 (upper secondary) in the generic IAB-SOEP measure. On the one hand, it is possible that what these respondents report as “extended apprenticeship at a company” or “vocational school” in the IAB-SOEP measure on vocational training does not constitute formal vocational education and training at the upper secondary level in ISCED, for example, because the program is either not classified as formal education, or not long enough (note that the IAB-SOEP instrument does not indicate what “extended” is supposed to mean). Some countries also offer vocational education and training at the lower secondary level, or vocational upper secondary programs that are too short for the resulting qualification to count as a

completion of upper secondary education. The CAMCES tool can differentiate the resulting qualifications from vocational education and training at the upper secondary level, while the IAB-SOEP instrument cannot. On the other hand, there may also be underreporting in the CAMCES measure, especially in the combination box, when respondents only think about their schooling rather than all levels and types of formal education.

Furthermore, 12.3% of respondents are coded as having completed lower secondary (ISCED 2) education in CAMCES and general upper secondary (ISCED 3) education in the standard IAB-SOEP measure. This may result from the fact that the response option “graduated from higher-level secondary school” (which was the vaguer “weiterführende Schule mit Abschluss beendet” in the most-often used German version of the questionnaire) in the IAB-SOEP measure may have been interpreted as having completed education classified as *lower* secondary education (ISCED 2) in the country of origin of the respondent. This might happen especially when the response option “graduated from mandatory schooling with certificate” in that country refers to less than lower secondary education. The cases classified at ISCED 4 (post-secondary non-tertiary education) in the IAB-SOEP measure but ISCED 2 in CAMCES will result from a combination of the issues described in this and the previous paragraph.

Respondents reporting a vocational upper secondary qualification (ISCED 3) using the CAMCES tool distribute almost evenly across three IAB-SOEP categories: 1) vocational (24.9%) and 2) general (24.5%) upper secondary (ISCED 3) and 3) post-secondary non-tertiary education (28.7%, ISCED 4). The high number of respondents coded as ISCED 4 may again result from the ambiguity of the response option “graduated from higher-level secondary school”/“weiterführende Schule mit Abschluss beendet” in combination with completed vocational education and training (see above). In Germany, this category includes respondents who have a certificate of upper secondary general education (“*Abitur*” or “*Fachhochschulreife*”) and completed vocational training (apprenticeship or vocational school). The respective combination of generic foreign qualifications measured with the IAB-SOEP instrument was thus also coded as ISCED 4 (see Table 4, appendix). However, this combination of educational qualifications is not common in other countries. In many, especially Eastern European countries, there are generally and vocationally oriented secondary schools, which is reflected in the CAMCES database but not in the IAB-SOEP measure. Using the IAB-SOEP measure, respondents having completed vocationally oriented secondary schooling in such countries may report completed secondary schooling on the first education item, then coded as ISCED 3 general, or completed secondary schooling on the first item and vocational schooling on the second item, then coded as ISCED 4 (no response option for vocationally oriented secondary schooling is available in the standard IAB-SOEP measure). Both approaches lead to a different coding than reporting the respective qualification using the CAMCES tool, where the measure results in ISCED 3 vocational.

Around a third of cases (33.1%) classified as ISCED 3 general education using the CAMCES tool are coded in the same way using the IAB-SOEP standard measure. The other cases are coded as ISCED 2, ISCED 3 vocational or ISCED 4, and even ISCED 6-7 in the IAB-SOEP measure. Respondents from countries where ISCED 3 is part of mandatory education, which we have assumed to refer to ISCED 2 (see Table 4, appendix), will be miscoded as ISCED 2 in the IAB-SOEP measure. ISCED 3 vocational could be the result of qualifications

reported as “extended apprenticeship at a company” or “vocational school” in the IAB-SOEP measure not constituting vocational education and training at the upper secondary level in ISCED and hence, in the CAMCES database. Respondents mentioning completed higher education (ISCED 6-7) in the IAB-SOEP measure and who reported a general upper secondary (ISCED 3) qualification in the CAMCES tool likely focused their attention on schooling only when using the CAMCES instrument, leading to underreporting. Social desirability bias in the IAB-SOEP measurement is an alternative explanation, for example when a respondent who has dropped out of university before obtaining the degree reported the degree anyway. A final explanation could be the broader meaning of “college” in the English translation compared to “Hochschule/Universität” in the German version, since “colleges” in many English-speaking countries also include institutions offering upper secondary (ISCED 3) and post-secondary non-tertiary (ISCED 4) education, while “Hochschule/Universität” clearly refers to higher education (ISCED 6-8) only. Respondents using the English version of the questionnaire might thus overreport their education using the IAB-SOEP standard measure compared to the CAMCES measure and the German version of the IAB-SOEP measure. This problem may also have occurred in other translated versions of the IAB-SOEP standard education measure.

How can the CAMCES measure arrive at ISCED 4, while the IAB-SOEP measure detects ISCED 3 general (12.8%) or vocational (22.3%) only, or ISCED level 6/7 (tertiary education, Bachelor and Master level, 14.9%)? ISCED 4 programs are often of a marginal character and very heterogeneous across countries, such as short programs after completion of upper secondary education, or bridging programs to enter higher education that are equivalent to general ISCED 3 programs. If the short programs are not regarded as one of the vocational options offered in the IAB-SOEP measure by respondents, respondents will only report their next lower qualification, resulting in underreporting. In the case of bridging programs, these will often be reported as completed secondary schooling, and in this case, it is rather their classification in ISCED that is doubtful than the IAB-SOEP measure. Some respondents, however, may also perceive those post-secondary qualifications as higher education already, especially if they were offered at an institution of higher education, and report them accordingly, leading to overreporting.

Moving on to tertiary education, while the CAMCES tool allows measuring and coding the rather diverse set of qualifications classified as short-cycle tertiary education (ISCED 5), the IAB-SOEP measure does not offer any response category relating to this level (see section 2.1). Respondents thus likely use various response options that they regard as closest or report their next lower qualification only. In most cases, this leads to underreporting.

With regards to Bachelor’s and Master’s level education, this is where both measures agree the most. As a concept, higher education is thus more universally understood than different levels of schooling or vocational training. In some cases, however, using the IAB-SOEP measure, respondents report completed upper secondary education only (9.2% of those reporting Bachelor level and 6.9% of those reporting Master level qualifications in CAMCES). Again, this may result from social desirability bias, this time in the CAMCES instrument, especially amongst higher education drop-outs. There may also be tertiary qualifications that are not regarded as higher education (as implied by the terms “Hochschule/Universität” in the German questionnaire) by respondents, thus leading to under-

reporting using the IAB-SOEP measure. For doctoral level education, the match is also rather high, but more than a third of respondents classified as ISCED 8 using the CAMCES measure (5 out of 13 cases) did not mention completed doctoral studies in the IAB-SOEP measure, which is a bit curious indeed: We would have expected the PhD to be the most universally understood term included in the IAB-SOEP instrument.

5 Summary and Discussion

The complexity of educational backgrounds of respondents in surveys increases as survey samples become more ethnically diverse, mobility between countries increases, and educational systems change. The measurement of migrants' level of education is hampered by the multitude of origins amongst migrants and strong differences between educational systems across countries as well as the lack of a universally understood education-related terminology.

In this paper, we introduced the CAMCES tool as a new measurement instrument for educational attainment in cross-cultural (including migrant) surveys and compared it with the standard measurement instrument as used in the IAB-SOEP Migration Samples. The CAMCES tool allows respondents to report their foreign educational qualification in terms of the educational system and language they received their education in rather than asking them to classify themselves in abstract education categories that can be interpreted in various ways. The tool consists of a short questionnaire module, with the education question being connected to an underlying database of educational qualifications via two alternative response interfaces. One interface, the combination box, allows a text search in the database (response options are only provided when respondents start typing), and the other interface, the search tree, works like a nested showcard with an unfolding second level. The results from the implementation, on a trial basis, of the CAMCES tool in the German IAB-SOEP Migration Samples M1 and M2 show that the CAMCES database covered most of the countries in which the sampled migrants had been educated, even though the samples possessed a highly diverse educational background. However, the 2016 sample (M2) included respondents educated in quite a number of countries for which the database had initially not been prepared. The foreign qualifications of these respondents will be reassessed in the 2018 panel wave for the M1 and M2 samples employing a more complete CAMCES database.

The results also show that looping the questionnaire module on educational qualifications at least once is important because – in the absence of a showcard with hierarchically ordered response options – respondents do not always report their *highest* qualification when asked to do so. The main issue that we have identified in the implementation of the CAMCES tool in the IAB-SOEP Migration Samples is the higher item nonresponse and the higher proportion of non-coded entries, both especially affecting the combination box, when compared with the standard measure. We presented further analyses of respondents' entries in the combination box and concluded that the language mismatch (respondents entering the foreign qualification in German rather than in the language of the country in

which they were educated) was a major obstacle to responses being automatically coded. Respondents might have overlooked the instruction to do so, or they might have been unable to report their highest foreign educational qualification in the original language. It may thus be advisable to only use the search tree interface in surveys of migrants. We also suspect that interviewers, trying to speed up the interview, affected the implementation of the combination box, especially with regard to the high item nonresponse. Interviewers may have ignored the instruction to hand the device over to the respondent, and when they themselves could not enter the response in the requested language they may have used German instead or may have skipped the item. Therefore, further analysis of this effect is needed to differentiate error introduced by interviewers from error generated by respondents. The combination box worked better in the web survey version of CAMCES which was tested in the Dutch LISS panel in early 2016 (Schneider, Briceno-Rosas, Herzing, & Ortmanns, 2016), which, however, represents a general population rather than a migrant sample, and is self- rather than interviewer-administered.

The analysis of the consistency between the IAB-SOEP standard measure and the CAMCES measure shows a substantial overlap, but also allowed to identify cases of under- and over-estimation of educational attainment in either measure. The substantial degree of ambiguity in generic response options such as “graduated from mandatory schooling” or “extended apprenticeship at a company” and their translations into various languages may lead to substantial inconsistencies between both measures. The combination box of the CAMCES tool in turn may lead to underreporting if respondents have only schooling but not other types and levels of education in mind. Of course, if both instruments arrived at the same result, there would be no reason to invest into a complex tool like CAMCES, and we think that the results reported here give more credibility to the CAMCES approach than the generic approach, especially as regards the search tree. We expect the CAMCES measure to be more precise, and classification of specific foreign qualifications in ISCED 2011 to lead to more valid ISCED codes than “guessing” at the ISCED code that would best represent what respondents have in mind when reacting to the IAB-SOEP response options (in various languages). Generally, the CAMCES measure appears to produce more cross-nationally comparable data.

We argue that the CAMCES tool, especially the search tree interface, represents a promising addition or even an alternative for the measurement of educational attainment of migrants as it improves the precision, validity, and comparability of the measurement. While a generic measurement instrument is easier to implement in a survey of migrants than an instrument providing context-sensitive education response options relative to the country in which the respondent was educated, supposedly universal but abstract response options are open to interpretation (see section 2.1). The CAMCES tool, in contrast, measures foreign qualifications directly. By making country-specific response options available for the education question in cross-cultural surveys, the CAMCES tool is far more detailed and specific than the standard instrument. By means of this more detailed measurement, it is possible to conduct a more precise mapping to cross-nationally comparative education classifications. Coding detailed education data to such classifications using a freely available database allows the comparison of qualifications across educational systems without the efforts and costs of ex-ante harmonization or further post-coding. It also avoids post-

coding errors when processing the data, and reduces the efforts related to documentation of post-coding. It can be used by many surveys, which then do not need to develop their own showcards any more for the education question, which is especially difficult when surveying migrants educated in various educational systems. However, this increased precision comes at the price of more complex preparation of the CAPI system, taking more interview time, and a somewhat higher response burden due to the language switch, which is probably only justifiable for studies of migrants rather than general population samples.

The possibilities offered by the CAMCES tool are promising for the SOEP Migration Samples and other cross-cultural surveys. An updated version²² of the tool has been implemented in the IAB-SOEP Migration Samples M1 and M2 in 2017 and in the BAMF-IAB-SOEP Survey of Refugees (Brücker, Rother, & Schupp, 2017)²³ in Germany. The 2017 data will show whether the improvements of the combination box interface reduced the number of responses given in German to a substantially lower level – if not, the combination box likely has to be regarded as unsuitable for migrant surveys in which respondents are not interviewed in their language of origin. Moreover, the CAMCES database is used in a survey of the project “ReGES – Refugees in the German Educational System” conducted at the Leibniz Institute for Educational Trajectories (LifBi).²⁴

More countries will be added to the CAMCES database within the context of the SERISS project, which runs until mid-2019 (see www.seriss.eu), so that we hope that the missing data issue for countries not yet covered in the database in the IAB-SOEP Migration Samples 2015 and 2016 will get solved in the near future. This development will focus on the search tree, which will be kept as simple as possible. With further development and wider implementation, the CAMCES tool will keep improving the measurement of educational attainment in cross-cultural surveys. The CAMCES tool is openly and freely available at www.surveycodings.org/education.

Funding

This work was supported by the Leibniz Association through the Leibniz Competition (formerly SAW Procedure), grant number SAW-2013-GESIS-5, as well as the European Union’s Horizon 2020 research and innovation program under grant agreement No 654221, the IAB/SOEP Migration Sample at the German Institute for Economic Research (DIW), Berlin, and the study “Refugees in the German Educational System” (ReGES) conducted at the Leibniz Institute for Educational Trajectories (LifBi) in Bamberg, Germany.

22 Based on the knowledge gained so far from the IAB-SOEP Migrant Samples and other studies such as cognitive interviews (Briceno-Rosas, del Busto, & Taoli, 2016; Lenzner et al., 2015), we improved respondent instructions and interface designs in the latest version (used in the 2017 studies).

23 BAMF is the Federal Office of Migration and Refugees, whose Research Center on Migration, Integration, and Asylum has entered into a cooperation with IAB and SOEP to establish a new refugee sample.

24 For further details, see <https://www.lifbi.de/en-us/furtherstudies/reges.aspx>

References

- Blom, A. G., & West, B. T. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*, 175–211. doi:10.1093/jssam/smw024
- Braun, M., & Mohler, P. P. (2003). Background variables. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 101–115). Hoboken: John Wiley & Sons.
- Braun, M., & Müller, W. (1997). Measurement of education in comparative research. *Comparative Social Research*, *16*, 163–201.
- Briceno-Rosas, R., del Busto, K., & Taoli, A. (2016). *Report on cognitive interview study in Venezuela for the project “Computer-Assisted Measurement and Coding of Educational Qualifications in Surveys (CAMCES)” at GESIS – Leibniz Institute for the Social Sciences*: Report by LASCO- Laboratoria de Ciencias Sociales. Retrieved from https://www.surveycodings.org/sites/default/files/files/Report%20CAMCES%20Cognitive%20Interviews%20Venezuela_final.pdf
- Brücker, H., Kroh, M., Bartsch, S., Goebel, J., Kühne, S., Liebau, E., . . . Schupp, J. (2014). *The new IAB-SOEP Migration Sample: An introduction into the methodology and the contents*. SOEP Survey Papers, Series C - Data Documentations. Berlin: DIW/SOEP. Retrieved from http://doku.iab.de/fdz/iab_soep/surveypaper_en.pdf
- Brücker, H., Rother, N., & Schupp, J. (2017). *IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen*. Berlin: DIW/SOEP. Retrieved from https://www.diw.de/documents/publikationen/73/diw_01.c.563710.de/diwkompakt_2017-123.pdf
- Couper, M. P., & Zhang, C. (2016). Helping respondents provide good answers in web surveys. *Survey Research Methods*, *10*, 49–64. doi:10.18148/srm/2016.v10i1.6273
- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken: John Wiley & Sons.
- Harkness, J. A., van de Vijver, F. J. R., & Johnson, T. P. (2003). Questionnaire design in comparative research. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 19–34). Hoboken: John Wiley & Sons.
- Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 117–140). Hoboken: John Wiley & Sons.
- Eckman S., & Kreuter F. (2018). Misreporting to looping questions in surveys: recall, motivation and burden. *Survey Research Methods*, *12(1)*, 59–74.
- Kleiner, B., Lipps, O., & Ferrez, E. (2015). Language ability and motivation among foreigners in survey responding. *Journal of Survey Statistics and Methodology*, *3*, 339–360. doi:10.1093/jssam/smv015
- Kroh, M., Kühne, S., Goebel, J., & Preu, F. (2015). *The 2013 IAB-SOEP Migration Sample (M1): Sampling design and weighting adjustment*. SOEP Survey Papers Series C - Data Documentation. Berlin.

- Kühne, S., & Kroh, M. (2017). *The 2015 IAB-SOEP Migration Study M2: Sampling design, nonresponse, and weighting adjustment*. SOEP Survey Papers Series C – Data Documentation. Berlin.
- Lenzner, T., Neuert, C., Otto, W., Landrock, U., Adams, F., Disch, K., . . . Menold, N. (2015). *Computer-Assisted Measurement and Coding of Educational Qualifications in Surveys (CAMCES)*. Kognitiver Pretest. GESIS Projektbericht. Version: 1.0. GESIS - Pretestlabor. Retrieved from <http://doi.org/10.17173/pretest25>
- OECD. (1999). *Classifying education programmes. Manual for ISCED-97 implementation in OECD countries*. Paris: OECD Publishing.
- OECD. (2017). *Education at a glance 2017 - OECD indicators*. Paris: OECD Publishing.
- Porst, R., & Zeifang, K. (1987). A description of the German General Social Survey test-retest study and a report on the stabilities of the sociodemographic variables. *Sociological Methods & Research*, 15, 177–218. doi:10.1177/0049124187015003002
- Reja, U., Lozar Manfreda, K., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. In A. Ferligoj & A. Mrvar (Eds.), *Developments in applied statistics (Metodološki zvezki, ISSN 1318-1726, 19)*. Ljubljana: University of Ljubljana, Faculty of Social Sciences. Retrieved from <http://mrvar.fdv.uni-lj.si/pub/mz/mz19/reja.pdf>
- Schneider, S. L. (2008). *The International Standard Classification of Education (ISCED-97). An evaluation of content and criterion validity for 15 European countries*. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES).
- Schneider, S. L. (2009). *Confusing credentials: The cross-nationally comparable measurement of educational attainment*. Doctoral Thesis. Nuffield College, University of Oxford. Retrieved from <http://ora.ox.ac.uk/objects/uuid:15c39d54-f896-425b-aaa8-93ba5bf03529>
- Schneider, S. L., Briceno-Rosas, R., Herzing, J. M. E., & Ortmanns, V. (2016). *Overcoming the shortcomings of long list showcards: Measuring education with an adaptive database lookup*. Paper presented at the International Conference on Social Science Methodology (RC33 Conference), Leicester, United Kingdom, September.
- Schneider, S. L., Joye, D., & Wolf, C. (2016). When Translation is not enough: Background variables in comparative surveys. In C. Wolf, D. Joye, T. W. Smith, & Y.-c. Fu (Eds.), *The SAGE handbook of survey methodology* (pp. 288–307). London: SAGE Publications Ltd.
- Smith, T. W. (1995). Some aspects of measuring education. *Social Science Research*, 24, 215–242. doi:10.1006/ssre.1995.1018
- Socio-Economic Panel (SOEP). (2016). *Data for years 1984-2015*. Version 32, SOEP.
- Socio-Economic Panel (SOEP). (2017). *Data for years 1984-2016*. Pre-published Data, Version 33, SOEP.
- The World Bank. (2017). Duration of compulsory education (years). Retrieved from <https://data.worldbank.org/indicator/SE.COM.DURS>
- Tijdens, K. (2014). Dropout rates and response times of an occupation search tree in a web survey. *Journal of Official Statistics*, 30, 23–43. doi:10.2478/jos-2014-0002

- Tijdens, K. (2015). Self-identification of occupation in web surveys: Requirements of search trees and look-up tables. *Survey Insights: Methods from the Field*, Retrieved from <http://surveyinsights.org/?p=6967>.
- TNS Infratest Sozialforschung. (2014). *Erhebungsinstrumente des IAB-SOEP-Migrations-samples 2013: Übersetzungshilfen (englisch, polnisch, türkisch, rumänisch, russisch)*. Berlin: DIW/SOEP. Retrieved from http://panel.gsoep.de/soep-docs/surveypapers/diw_esp0219.pdf
- UNESCO-UIS. (1997 [2006]). *ISCED 1997 International Standard Classification of Education*. Montreal: UNESCO Institute for Statistics.
- UNESCO-UIS. (2012). *International Standard Classification of Education ISCED 2011*. Montreal: UNESCO Institute for Statistics.
- UNESCO-UIS, European Union, & OECD. (2015). *ISCED 2011 operational manual - guidelines for classifying national educational programmes and related qualifications*.

Appendix

Table 4 ISCED codes for the IAB-SOEP education measure

IAB-SOEP instrument 2013 / 2014		Vocational training and higher education (question 83 and 84)
ISCED categories	School-leaving certificate (question 72)	
ISCED 0-1 Less than lower secondary education	left school without graduating	no answer other in-house training at a company
ISCED 2 Lower secondary education	graduated from mandatory schooling with certificate	no answer other in-house training at a company
ISCED 3 General upper secondary education	graduated from higher-level secondary school	no answer other in-house training at a company
ISCED 3 Vocational upper secondary vocational education	left school without graduating	completed an extended apprenticeship at a company attended a vocational school
	graduated from mandatory schooling with certificate	completed an extended apprenticeship at a company attended a vocational school
ISCED 4 Post-secondary non-tertiary education	graduated from higher-level secondary school	completed an extended apprenticeship at a company attended a vocational school
ISCED 5 Short-cycle tertiary education	-	-
ISCED 6-7 Tertiary education Bachelor and Master levels	any	university / college with a more practical orientation university / college with a more theoretical orientation
ISCED 8 Doctoral level	any	doctoral studies

Note: Only if respondents indicated in item 84 that they received a certificate, i.e. successfully completed their vocational training or higher education, they were coded in the respective education category of item 83.

Table 5 Categories and recodes of the CAMCES education variable

ISCED categories		Detailed alternative ISCED 2011 generated by the CAMCES tool	
		Code	Label
ISCED 0-1	Less than lower secondary education	0	Less than primary education
		100	Primary education
ISCED 2	Lower secondary education	243	General lower secondary completed, without direct access to upper secondary education
		244	General lower secondary completed, with direct access to upper secondary education
		253	Vocational lower secondary completed, without direct access to upper secondary education
		254	Vocational lower secondary completed, with direct access to upper secondary education
ISCED 3 general	General upper secondary education	343	General upper secondary completed, without direct access to tertiary education
		344	General upper secondary completed, with direct access to tertiary education
ISCED 3 vocational	Vocational upper secondary education	353	Vocational upper secondary completed, without direct access to tertiary education
		354	Vocational upper secondary completed, with direct access to tertiary education
ISCED 4	Post-secondary non-tertiary education	443	General post-secondary non-tertiary education completed, without direct access to tertiary education
		444	General post-secondary non-tertiary education completed, with direct access to tertiary education
		453	Vocational post-secondary non-tertiary education completed, without direct access to tertiary education
		454	Vocational post-secondary non-tertiary education completed, with direct access to tertiary education
ISCED 5	Short-cycle tertiary education	560	Short-cycle tertiary education
ISCED 6	Bachelor's level	660	Bachelor's or equivalent level
ISCED 7	Master's level	760	Master's or equivalent level
ISCED 8	Doctoral level	860	Doctoral or equivalent level

Language Barriers during the Fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany

Jannes Jacobsen

Abstract

The IAB-BAMF-SOEP Survey of Refugees is one of the first large-scale quantitative surveys in Germany focusing on refugees exclusively. It is able to provide valuable insights on the recent cohort of refugees who arrived in Germany as of the year 2013. However, due to the fact that most respondents of the target population are not proficient in German, the research partners who conducted the survey faced several obstacles. One crucial aspect in this regard was the written and audio translation of field instruments. Therefore, this paper gives some insights into the translation and fielding procedures and presents numbers on the use of written and audio translations in the IAB-BAMF-SOEP Survey of Refugees. We found that especially a written translation was very helpful, but also the audio files proved to facilitate participation in the survey. However, due to the fact that those tools are new and innovative, further research on their effects on survey quality needs to be carried out.

1 Introduction²⁵

In 2015, around 890,000 refugees came to Germany (Bundesministerium des Inneren [BMI], 2016). This inflow was unique in the younger German history. To respond to these events in Germany, the German Socio-Economic Panel (SOEP) (Wagner, Frick, & Schupp, 2007), the Institute for Employment Research (IAB), and the Migration, Integration, and Asylum Research Center at the Federal Office for Migration and Refugees (BAMF-FZ) implemented a panel study²⁶, the so-called IAB-BAMF-SOEP Survey of Refugees in Germany²⁷, with humanitarian migrants that arrived in Germany between January 2013 and January 2016 as its target group (for first analyses, see Brücker, Rother, & Schupp, 2016; Brücker et al., 2016; Brücker, Rother, & Schupp, 2017).

The following description of the refugee sample relies on the methodological documentation of the IAB-BAMF-SOEP Survey of Refugees (Kroh et al., 2017a, 2017b). The sample

25 *Acknowledgements:* Thanks go to our partners at the Institute for Employment Research (IAB), the Migration, Integration, and Asylum Research Center at the Federal Office for Migration and Refugees (BAMF-FZ), and the fieldwork company Kantar Public with whom this study was developed.

26 This study contains around 4,500 respondents in around 3,300 households.

27 doi:10.5684/soep.v33.1

was drawn from the central register of foreigners (AZR) at the BAMF (see Babka von Gos-tomski & Pupeter, 2008). This register is updated regularly. However, due to the inflow of refugees at the German borders, especially in 2015 and 2016, there was a delay between the time of arrival of refugees and their proper registration. To tackle this delay, the sample for the refugee survey was drawn in tranches. The first tranche was drawn in January, the second in April, and the third and fourth in June 2016. In addition, the sample was split into two subsamples, namely M3 and M4. The former had a focus on adult migrants while the latter focused on refugee families. In general, both subsamples had a higher sampling probability for females, persons over 30, and refugees with granted asylum or refugee status. By applying weights that account for such a disproportional design, the samples remain representative in respect to the target group and can be used jointly.

Being a household panel, the sampling design followed the concept of so-called “anchor persons”. This means that even though we sampled on an individual level the whole household of the sampled (anchor) person was included in the survey (Kroh et al., 2017b). Each household of an anchor person and subsequently each member in this household were included in the sample. A letter was sent to each household explaining the project and emphasizing that this project was not related to any asylum procedures that a household might currently be involved in.

2 Translation and Fieldwork

Besides the particularities related to the sampling procedure and design, the research partners further needed to consider the fact that the respondents likely did not speak German well enough in order to take part in the survey in the German language. Therefore, all the material (advance letters, flyers, and questionnaires) was provided in seven different languages, including German (see Table 1). The translation process and the provision of the field material was the responsibility of Kantar Public.²⁸ For the translation of the material, professional translators for each language were used. The process for the questionnaire was as follows: To begin with, a German version of the questionnaire was developed by the research partners. It was then translated into English by two translators working separately. Either the English or the German version was then the starting point for all other translations²⁹: Two translators each produced a translation. Then one translator created a reconciled version and this was handed to the other translator who could correct remain-

28 <https://www.tns-infratest.com/sofo/kantar-public.asp>

29 This mixture was due to the fact that the corresponding language was not German for all the participating translators.

ing mistakes.³⁰ In the end, all the material was provided to the respondents both in German and in one of the languages listed in Table 1.³¹

During the interview, both languages (German and the respondent’s language) were displayed on the screen. Since the mode of the interview was computer-assisted personal interviewing (CAPI), both interviewer and respondent were able to look at the screen jointly to overcome language barriers. For cases where this was not sufficient, Kantar Public also provided audio files for each language. These audio files had been produced by the same translators who had already developed the written translation. On top, if anything in the interviewing process was problematic, Kantar Public also provided a hotline where professional interpreters were available to help during the process of fieldwork.

Table 1 lists all the translated versions of the questionnaire that were eventually provided by Kantar Public. Furthermore, it provides information on the use of these translated questionnaires during fieldwork.

Table 1 Distribution of languages used during fieldwork, in percent

Language	Percent
German/English	16.1
German/Arabic	65.2
German/Farsi	12.6
German/Pashto	1.0
German/Urdu	1.7
German/Kurmanji	3.4
N	4,527

Source: IAB-BAMF-SOEP Survey of Refugees, own calculations

Leading by far, the German/Arabic version was used most frequently with 65%. This reflects the regions of origin of the respondents quite well: The Middle East, and especially Syria, was the biggest source of migrants in this sample. Nevertheless, it is quite striking that the second most used language version was German/English with around 16%. If we cross-tabulate citizenship and language version used (see Table 2), we see that many respondents who used the German/English version came from countries such as Russia, Ukraine, Georgia, Armenia, or the Balkan states (41.1%) – countries from which respondents are not very likely to be granted asylum. These countries account for around 7% of the sample. Therefore, a bias due to not providing a questionnaire in languages of these countries should be limited for future waves of the panel. The second highest share of

30 There are some studies that deal with the quality of questionnaire translation and bias in terms of a transfer of meaning. With the mentioned procedure of questionnaire translation, a possible bias should be minimized. However, for a comprehensive overview regarding occurring obstacles, see Harkness et al. (2010).

31 The translated instruments can be provided by the SOEP team on request.

respondents using the German/English version were African nationals such as Eritreans and Somalians (28.4%). The remaining users of the English version distribute in small numbers over the rest of the populations.

Table 2 Language version and country of origin of respondents who used a given language the most, in percent, absolute numbers in parenthesis

Language versions	Countries of origin	Percent (absolute numbers)	N
German/English	Russia, Ukraine, Georgia, Armenia & Balkans	41.1 (299)	728
German/Arabic	Syria	73.5 (2,171)	2,952
German/Farsi	Afghanistan	88.4 (505)	571
German/Pashto	Afghanistan	95.7 (45)	47
German/Urdu	Iran/Pakistan	85.7 (66)	77
German/Kurmanji	Iraq	65.8 (100)	152
Total			4,527

Source: IAB-BAMF-SOEP Survey of Refugees, own calculations

3 Controlling the Use of Translations

It is obvious that, given the translation process and the different modes of translation (written and audio), we require an assessment of these aspects. Furthermore, these aspects will also need to be taken into account in substantive analyses. To tackle the translation and fieldwork particularities, the research partners provided an interviewer questionnaire in which they asked the interviewers to rate the respondent’s German language proficiency, provide information on the extent to which written translations and/or audio files were used, and to rate the helpfulness of these “tools”. In the following section, I will present these assessment instruments and present a first descriptive analysis.³²

- a) How well could you conduct the interview in German?³³
 Wie gut konnten Sie das Interview in deutscher Sprache durchführen?
 Scale: 1 (very well) to 5 (very bad)
- b) During this interview, how often were the translated texts used?
 Wie häufig wurden in diesem Interview die übersetzten Texte verwendet?
 Scale: 1 (with every question) to 5 (not at all)
- c) In this interview, how helpful were the translated texts?
 Wie hilfreich waren in diesem Interview die übersetzten Texte?
 Scale: 1 (very helpful) to 4 (not at all helpful)

32 The following questions were provided to the interviewers in German only (the English translations provided here merely serve documentary purposes).

33 We used this question also as an assessment of the German proficiency of the respondent. As shown further down, this assumption is legitimate.

- d) During this interview, how often were the audio files used?
Wie häufig wurden in diesem Interview die Audiodateien verwendet?
Scale: 1 (with every question to 5 (not at all)
- e) In this interview, how helpful were the audio files?
Wie hilfreich waren in diesem Interview die Audiodateien
Scale: 1 (very helpful) to 4 (not at all helpful)

Table 3 Distribution of respondent’s German language proficiency, as rated by the interviewer, in percent

German Language proficiency	Percent
very well	11.2
quite well	15.0
okay	22.1
quite bad	22.9
very bad	28.8
N	4,527

Source: IAB-BAMF-SOEP Survey of Refugees, own calculations

Table 3 shows that there is a clear trend towards assessing the German language skills of the respondents as (quite) bad (without the middle category). Nevertheless, it is remarkable that there seems to be a group of refugees in the sample (around 11%) that already masters the German language. Even though there is thus a share of respondents that seems to speak German very well, these numbers nevertheless indicate that translations are necessary. If we correlate these numbers with the four biggest refugee nationalities in our sample – respondents from Syria, Iraq, Afghanistan, and Eritrea –, we can conclude that especially refugees from Syria speak German quite well or better (with over 28% of the Syrians in the sample, the interview could be conducted in German quite well or better – as reported by the interviewers). Previous publications indicate that especially Syrians have high educational degrees. We therefore assume that this is the reason for their high German proficiency (see Schupp et al., 2017).

As mentioned before, for those who did not speak German well enough for survey participation in the German language, the research partners provided a written translation. Table 4 shows that this mode was used in particular by respondents with (quite) bad German-language skills. We see a significant relationship between the language assessment by the interviewer and the use of the written translation. In other words, the less a respondent spoke German, the more likely the written translation was used. Nevertheless, a small share with strong German language skills also used the written translation. We assume that this is due to the fact that a written translation was provided in every case per default. Therefore, also those with high proficiency might have used it to balance out remaining difficulties. Furthermore, there was a share of respondents with low German proficiency who did not use the written translation at all. Presumably those are respondents who were

not able to rely on their mother tongue, who did not speak English or who were illiterate and therefore not able to read.

In around 58% of the interviews, the written translation was used in the entire questionnaire. Connected with an average of 1.4 on the scale of helpfulness of the written translation (see above variable number c), we conclude that this written mode was necessary and extremely useful.

Table 4 Respondent's German language proficiency, as rated by interviewer, over use of written translation, in percent

German Language proficiency	Use of written translation					Total	N
	with every question	with two thirds	with half of the questions	with less than half of the questions	not at all		
very well	4.4	1.5	1.1	1.8	2.5	11.2	509
quite well	6.0	2.8	1.8	1.9	2.5	15	678
okay	11.9	3.9	2.5	1.3	2.5	22.1	1,000
quite bad	15.9	2.1	1.4	1.2	2.4	22.9	1,038
very bad	19.9	1.6	1	1.4	4.9	28.8	1,302
Total	58.1	11.8	7.8	7.6	14.7	100	
N	2,630	535	352	344	666		4,527

Source: IAB-BAMF-SOEP Survey of Refugees, own calculations

A similar pattern is detected when analyzing the audio files of the translations, in relation to the respondent's German language proficiency (see Table 5). The less a respondent spoke German, the more likely it was that the audio files were used. The mean assessment shows a score of 1.6 (for the scale, see above variable number e), given by those who gave another answer than "not at all". We therefore consider the audio files to be helpful as well. Nevertheless, we detect some contradictory patterns. On the one hand, as mentioned before, respondents with low German proficiency tended to use the audio "tool". On the other hand, there was also a high share of respondents with low proficiency who did not use the audio "tool" at all. This might be due to the fact that the research partners were not able to provide all the languages that were needed and therefore the audio files, in these cases, were redundant.

In sum, the two translation modes of the questionnaire (written and audio) were assessed as very helpful by the interviewers, even though they were used to quite different degrees. Especially the written translation stands out with having been used in almost every interview; it was also assessed by the interviewers as a bit more helpful than the provided audio files. The vast difference in usage is clearly due to the fact that the written translation was presented as a default while the audio files had to be activated by the interviewer or the respondent.

But in both cases we also detect contradictory patterns such as respondents with high German proficiency choosing these translation “tools” and respondents with low proficiency not choosing them. Besides the aforementioned reasons, I would like to repeat that the variables rely on information of the interviewers and that the questions were designed to ask for a raw estimation. Some ambiguities might also be due to this particular design.

Table 5 Respondent’s German language proficiency, as rated by interviewer, over use of audio files, in percent

German Language proficiency	Use of audio translation					Total	N
	with every question	with two thirds	with half of the questions	with less than half of the questions	not at all		
very well	1.1	0.3	0.2	0.6	9.0	11.2	509
quite well	0.2	0.7	0.4	1.2	12.4	15	678
okay	0.8	1.4	1.1	2.2	16.6	22.1	1,000
quite bad	1.2	1.7	1.2	2.3	16.6	22.9	1,038
very bad	3.9	2.3	1.1	2.7	18.8	28.8	1,302
Total	7.2	6.3	4	9.1	73.4	100	
N	326	286	181	411	3,323		4,527

Source: IAB-BAMF-SOEP Survey of Refugees, own calculations

As mentioned before, Kantar Public also provided professional interpreters to help during the fieldwork. Nevertheless, we observed that also many informal interpreters helped during the interviewing process. The term “informal” refers to the fact that the research partners did not provide these interpreters. We need to be aware of the fact that the use of interpreters might have created some bias. To control for such a distortion, questions on the presence of an informal interpreter were included in the interviewer questionnaire.

a) Due to language barriers, did other people help with translating the questions?

Haben wegen Sprachschwierigkeiten andere Personen bei der Übersetzung der Fragen geholfen?

1 (yes, a professional interpreter); 2 (yes someone different); 3 (no, nobody).³⁴

b) [asked when informal interpreters were used] In which language did the other person translate?

In welche Sprache hat die andere Person übersetzt?

c) [asked when informal interpreters were used] How old is the other person, approximately?

Wie alt ist die andere Person etwa?

³⁴ German Translation: 1 (ja, ein professioneller Dolmetscher); 2 (ja, eine andere Person); 3 (Nein, niemand)

- d) [asked when informal interpreters were used] Is the other person male or female?
Ist die andere Person männlich oder weiblich?
- e) Into which language did the interpreter translate?
In welche Sprache hat der Dolmetscher übersetzt?
- f) How old was the interpreter, approximately?
Wie alt war der Dolmetscher in etwa?
- g) Was the interpreter male or female?
Ist der Dolmetscher männlich oder weiblich?

Table 6 Use of interpreters, in percent

Type of interpreter	Percent
Professional interpreter	1.7
Informal interpreter	34.1
No interpreter	64.3
N	4,527

Source: IAB-BAMF-SOEP Survey of Refugees, own calculations

As Table 6 indicates, we see that the professional interpreters barely played a role during fieldwork in contrast to informal interpreters. In about one third of the interviews, help from an informal interpreter was sought. This figure is a reason to look deeper into the question about who sought such help in order to give some recommendations on how to deal with these cases.

By taking a look at which nationals used informal interpreters the most, we can report that respondents from Syria, Afghanistan, and Iraq were the most frequent users, which is most likely due to the many dialects that are spoken in the Arabic world and the high proportion of their group in the refugee population. This argument is supported by the fact that most of the informal interpreters interpreted into some form of Arabic. This information was provided by the interviewer (see question e in the aforementioned list). For the analysis of the survey data, we should consider taking these numbers into account and to control for an interpreter bias in the (multivariate) analysis. A problem with using interpreters is the fact that they might have a strong effect on the respondent’s behavior. Therefore, it is important to consider the questions regarding the interpreter’s characteristics especially when it comes to sensitive questions on for example religion or gender. It might be intimidating to answer questions about relationships, role concepts, and values when not only the interviewer but also a third party is present who might be part of the family or the circle of friends. Additionally, we are not able to control whether the given translation actually matches the semantic meaning of a given item.

Furthermore, on first sight it seems legitimate to conclude from this relatively high usage of informal interpreters that it is worthwhile to include more, especially Arabic, dialects in future surveys. However, due to obvious budget constraints research projects are not able to provide all the necessary languages. Therefore, it is more promising and more

inclusive to provide a wide range of languages instead of focusing on only one (and its dialects) so that we do not, unintendedly, narrow down the target group.

So far, in our assessment of the German language proficiency of the respondents we have relied on the information provided by the interviewer. However, the problem with the interviewer assessing the language skills is that it is unclear in reference to whom or what the assessment has been carried out. To tackle this problem, I compared the external evaluation with an assessment as undertaken by the respondents themselves. In the questionnaire, the research partners provided a language module testing German language skills in reading, writing, and speaking.³⁵ The three variables on the language competencies were strongly correlated, which is indicated by a Cronbach's Alpha of 0.93; therefore, I was able to produce an additive index. If we correlate this index of the assessment with the interviewer evaluation of the respondent's German language proficiency, a correlation of .58 is measured. Therefore, we have an indication that our measurement of language abilities, by the interviewers and by the respondents themselves, is appropriate.

4 Conclusion

In sum, we can conclude that different modes of translation (written and audio) are highly valuable when conducting interviews with refugees. Using these different modes, the research partners were able to reduce a response bias in respect to language barriers. Even though we are not able to assess the translation itself, by applying an interviewer questionnaire we can control and test for emerging biases during the interview. Especially with panel data this is quite important in order to improve further waves. Even though we are in need to revise our instruments constantly, for future waves it is likely that the respondents improve their German-language skills and therefore the need for bias control will decline.

References

- Babka von Gostomski, C., & Pupeter, M. (2008). Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters. *Methoden – Daten – Analysen*, 2, 149–177.
- Bundesministerium des Inneren [BMI]. (2016, May 15). 890.000 Asylsuchende im Jahr 2015. Press release from 30.9.2016. Retrieved from <http://www.bmi.bund.de/Shared-Docs/Pressemitteilungen/DE/2016/09/asylsuchende-2015.html>
- Brücker, H., Rother, N., & Schupp, J. (2017). IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnisse sowie kognitiven Potenzialen. *DIW Politikberatung Kompakt*, 123.

35 Scale: 1 “very well” to 5 “not at all”

- Brücker, H., Rother, N., & Schupp, J. (2016). *IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse*. DIW Politikberatung kompakt, 116.
- Brücker, H., Rother, N., Schupp, J., Babka von Gostomski, C., Böhm, A., Fendel, T.,... Valizadeh, E. (2016). Forced migration, arrival in Germany, and first steps toward integration. *DIW Economic Bulletin*, 46, 541–556.
- Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P.,... Smith, T., (Ed.) (2010). *Survey methods in multinational, multiregional, and multicultural contexts*. New Jersey: John Wiley & Sons.
- Kroh, M., Böhm, A., Brücker, H., Jacobsen, J., Kühne, S., Liebau, E.,... Trübswetter, P., (2017a). Die IAB-BAMF-SOEP-Befragung von Geflüchteten: Studiendesign und Feldergebnisse der Welle 1 (2016). In J. Schupp, H. Brücker & N. Rother, *IAB-BAMF-SOEP-Befragung von Geflüchteten* (pp. 4–19). DIW Politikberatung Kompakt, 123.
- Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017b). Sampling, nonresponse, and integrated weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4). *SOEP Survey Papers*, 477.
- Schupp, J., Brücker, H., Brenzel, H., Jacobsen, J., Jaworski, J., Kosyakova, Y.,... Siegert, M. (2017). Bildung, Sprache und kognitive Potenziale. In J. Schupp, H. Brücker, & N. Rother, *IAB-BAMF-SOEP-Befragung von Geflüchteten* (pp. 18–85). DIW Politikberatung Kompakt, 123.
- Wagner, G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): Scope, evolution and enhancements. *Schmollers Jahrbuch*, 127, 1, 139–170.

Surveying Migrants in the Context of the Low-income Panel PASS

Birgit Jesske

Abstract

The panel study “Labor Market and Social Security” (PASS), established by the Institute for Employment Research, is a central data source for research on unemployment, poverty, and the welfare state in Germany. Since 2007, approximately 12,000 persons in more than 8,000 households have been interviewed annually. The study’s design particularly allows for evaluating the situation of recipients of basic income support, called unemployment benefits II (UB II) in Germany. With the influx of refugees to Germany, the structure of UB II recipients changed and as a consequence Arabic-speaking persons needed to be integrated into PASS. In 2016, that is, in PASS wave 10, we were faced with several challenges to meet the needs of this new population. This paper focuses on fieldwork issues: How did we access the special population of Arabic-speaking respondents and which contact strategies were necessary to reach refugees? And how did we manage to conduct the interviews given the respondents’ different cultural backgrounds and their special living conditions as refugees in Germany? The sampling structure of PASS allows for comparing the refugee sample and the general UB II recipients. A great willingness to participate and high motivation was apparent among the refugees. It was possible to obtain access to the refugee households both by telephone and face-to-face. Whereas non-response occurred to a larger extent in the general refresher sample due to refusals, in the refugee sample it was caused by lower reachability.

1 PASS as a Household Survey in the Low Income Sector

The Labor Market and Social Security Panel (PASS) is a central data set for labor market and poverty research in Germany. It is also used for the evaluation of social policy measures adopted in the framework of volume II of the Social Security Code (SGB II).³⁶ The panel was established in 2006 at the Institute for Employment Research (IAB) and the annual survey waves that have been taking place since 2007 make a detailed description and analysis of the living conditions of unemployment benefit (UB II) recipients and low-income households possible. Given the fact that basic income benefits are tied to persons

36 Social Security Code (SGB), volume II, Basic Benefits for Job-Seekers.

living together and building a needs-based community, the conceptual design of PASS is a household survey (Trappmann, Beste, Bethmann, & Müller, 2013).

In light of the research interests mentioned above, two subsamples are formed in PASS so that the UB II recipients can be compared to a control group consisting of a sample of the general population. The benefit recipients (needs-based communities) sample is drawn from the process data of the Federal Employment Agency (BA). This subsample is refreshed annually based on new additions to the benefit recipients (general refresher sample). The general population sample is drawn from the population registry of selected municipalities. Up to now, the latter subsample was only refreshed in wave 5. The PASS survey started in 2007 with 12,794 households surveyed and, including the addition of yearly refresher samples of the UB II population (approximately 1,000 households each), has now a size of, on average, 10,000 households per survey wave or per year. The persons interviewed in PASS are all persons living in a household and being of at least 15 years of age. The Institute for Applied Social Sciences (infas) has been conducting the surveys since 2009 (wave 4) on behalf of the IAB and has also been responsible for the data processing.

1.1 Migrants in PASS

The population investigated in PASS includes a high proportion of migrants. Here, the disproportionately more frequent precarious living conditions to be found among the migrant group play a role, as do lacking linguistic abilities or vocational qualifications, which make it more difficult for migrants to integrate into the German labor market. To prevent the risk of systematic unit non-response particularly in this target group, the PASS survey has taken on a multilingual format since its beginning. However, given budget constraints, the PASS questionnaires were only translated into Turkish and Russian until wave 9. Initially, these two groups represented the numerically most important migrant groups among benefit recipients in Germany. The influx of refugees that began in 2013 has changed the composition of UB II recipients. Since then, Syrians and Iraqis have become the largest group among the migrants receiving UB II. In light of this fact, Arabic was chosen to be an additional survey language in survey wave 10 (2016)³⁷, and Syrians and Iraqis had to be included as a refugee sample in PASS. The inclusion of this new migrant group with its cultural particularities did not remain without consequences for the entire survey process and the field strategies in PASS. First experiences with interviewing refugees (Brücker, Rother, & Schupp, 2016; Johannson, Schiefer, & Andres, 2016; Gonser, Jesske, & Pfeiffer, 2017) have shown that for this target group it is not sufficient to simply provide a questionnaire that has been translated into their native language. The cultural particularities and the particular life situation as refugees in Germany require additional measures to be taken to facilitate contacting and interviewing refugees. How did we deal with these challenges in PASS and what experiences did we make during the first contact with this target group in survey wave 10? Answers to these questions will be provided in the following. By comparing this special sample with the general refresher sample, we can identify whether

37 Since survey wave 10, Turkish is no longer offered as a survey language. The proportion of Turkish-speaking persons in PASS has undergone a sharp decline in recent years.

special contacting strategies are necessary in dealing with this new migrant group and, if so, which.

1.2 Refugee Sample with Syrians and Iraqis in Wave 10

Syrian and Iraqi households were oversampled in the refresher sample within the UB II population, as we needed to reach a sufficient number of cases to allow for independent substantive analyses for Syrians and Iraqis. The oversampling was also advantageous for a detailed analysis of the contacting strategies in PASS survey wave 10. With the goal of reaching 500 completed interviews in Syrian/Iraqi households, a gross sample of 1,564 households was drawn. As already mentioned above, the general refresher sample, consisting of a population of UB II households of predominantly German nationality, serves as a basis for comparison with the refugee sample. The general refresher sample in PASS survey wave 10 was based on a gross sample of 2,870 households, and we aimed to eventually complete approximately 800 household interviews. The survey for wave 10 took place between the end of February 2016 and the beginning of September 2016. Households for the refugee sample were contacted from May to December 2016. The later start of fieldwork for this target group was due to the time which had to be spent on translating and preparing the Arabic version of the questionnaire. By the end of fieldwork, 521 Syrian/Iraqi households had been interviewed in PASS and 664 households in the general refresher sample. The distributions of the final disposition codes show a high degree of motivation and a great willingness to participate among the Syrian/Iraqi households in the refugee sample. This group achieved a response rate of 33.3% as compared to 23.1% in the case of the households in the general refresher sample (RR1, AAPOR response rate: The American Association for Public Opinion Research [AAPOR], 2016). Whereas in the general refresher sample the majority of non-responses were due to refusals, the non-responses in the case of Syrian/Iraqi households were the result of an inability to reach the households in question (non-contacts). The proportion of non-responses due to refusal amounted to 9.8% in the refugee sample as compared to 37.7% in the general refresher sample. Households that could not be contacted, either because they were not reached despite several contact attempts or because their address information was incorrect, comprised 43.6% in the refugee sample and 11.2% in the general refresher sample (Jesske, Knerr, & Kraft, 2017).

2 Study Design and Response-Enhancing Measures in PASS and the Particularities Involved in Interviewing Migrants

The sources and causes for (measurement) errors and distortions have been thoroughly investigated in survey research (Biemer, 2010; Groves & Couper, 1998; Dutwin et al., 2014). The result has been a large spectrum of measures and strategies that can be used to prevent selective non-responses, in particular with respect to special target groups within the total population. Measures and strategies are a component of study design, when, for instance, the survey methods to be used are selected. Furthermore, measures and strategies play a

role while working on a sample during fieldwork, for instance, when contacting target groups that are difficult to reach or when making a standardized interview possible with a person that has difficulties to follow a questionnaire.

2.1 Mixed-mode Study Design in PASS

The PASS study design involves a mix of methods in which both telephone interviews (CATI) and face-to-face interviews (CAPI) are possible. PASS interviewees can freely choose the interview method. Even different methods within a single household are possible. All the households in the refresher sample always start in CAPI. Households that cannot be reached despite several attempts to make contact in person or whose address is incorrect can be switched to CATI in the course of the survey. The flexibility of switching between survey methods ensures that efforts to make contact are maximally exploited in order to reach as many households in the sample as possible. The PASS survey period of more than six months (February to September) provides the opportunity to repeat and alternate different efforts at making contact. Moreover, tracking measures are employed in PASS, amongst others, to locate households that have moved. In addition, address searches and updates using the available registries (population registry bureaus, BA process data, Addressfactory (company within the Deutsche Post AG)) take place. Following Couper and Ofstedal (2009), tracking is understood as a comprehensive package of measures including the use of advance letters (incentives form part of this package), inquiries regarding the willingness to participate, and comprehensive contact information for the purpose of making contact in the current wave as well as in subsequent survey waves. In PASS, all households in the refresher sample receive advance information in written form on the purpose of the study, along with a detailed data protection statement. In addition, the letter announces an incentive of EUR 10, which the persons surveyed receive, along with a thank-you note, following their participation. In PASS, all letters and other written documents are also available in each of the survey languages.³⁸

2.2 Tailored Contact Strategies for the Refugee Sample

The usual PASS strategies were partially modified in wave 10 for the purpose of contacting the Syrian/Iraqi households. Both CATI and CAPI surveys were allowed as initial methods for Syrian or Iraqi households so that the Syrian/Iraqi households could be addressed in their mother tongue in Arabic (Standard Arabic): In CATI, contact was facilitated by bilingual interviewers. At its Bonn location, infas has a multilingual interviewing team, specifically trained and experienced for interviewing non-German respondents. Depend-

38 The study design and strategies employed in PASS are described in detail in the technical report published by the Research Data Center (FDZ) of the IAB following each survey wave (http://fdz.iab.de/de/FDZ_Individual_Data/PASS.aspx).

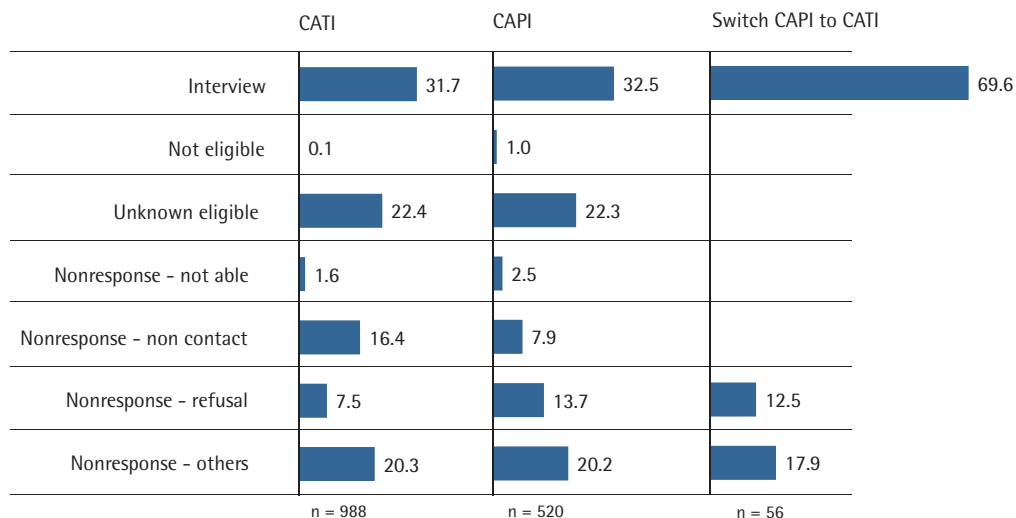
ing on the study design, this team can also be deployed as CAPI interviewers.³⁹ However, a specifically trained Arabic-speaking interviewing staff for CAPI surveys is not available in Germany at present. All households with a telephone number available started in CATI. For all other households, the CAPI interviewers attempted, even without having knowledge of Arabic, to establish contact with Syrian/Iraqi households in person. When doing so, they were allowed to draw on the support of other German-speaking persons living in the household or in the neighborhood. Furthermore, they had greater flexibility in conducting the interviews. The latter issue will be dealt with in greater detail in the next section. The assumption was that interviewers in CAPI can establish contact even if they do not have any knowledge of Arabic. Specific measures and strategies were adopted in order to assist the interviewers in their work and in dealing with the particular target group of refugees. Thus, in addition to the advance letter and the data protection statement in Arabic, a contact flyer was prepared for the Syrian/Iraqi households explaining the purpose of the survey and the function of the interview. The contact flyer was available to CAPI interviewers in both English and Arabic and it could be presented upon entering the household. Also CATI interviewers used sections from the flyer in order to better steer the conversation during the interview. If, despite the strategies mentioned above, the CAPI interviewers were unable to communicate with the Syrian/Iraqi households in person, they were asked to try to obtain a telephone number. This number could then be provided to the CATI interviewers for the purpose of making further contact.

Of the 1,564 addresses in the refugee sample of Syrian/Iraqi households, 576 households started in CAPI and 988 in CATI. 56 household were eventually transferred from CAPI to CATI for further treatment. Among the households that had initially been contacted with CATI, it was possible to conduct telephone interviews in two-thirds of the cases. For all other households, in both CAPI and CATI, it was possible to successfully interview nearly one-third of the households. Figure 1 shows the distributions of the final disposition code⁴⁰ for the respective survey methods. Whereas the proportion of refusals was almost twice as high in CAPI than in CATI (13.7% vs. 7.5%), the non-responses in CATI tended to be caused by a lower ability to reach the households. 16.4% of the households in CATI could not be reached until the end of fieldwork. In CAPI, this was the case for only 7.9% of the households. Syrian/Iraqi households showed a great willingness to participate, in particular when one compared their success rate to the one applying to the UB II households in the general refresher sample (as described above). The higher level of motivation of the refugee

39 infas has employed interviewers from the CATI field in collective housing, including in the case of refugee surveys: *Reallabor-Befragungen von Geflüchteten in der Rhein-Neckar-Region* on behalf of ZEW and Heidelberg University (2016), *Religion Monitor* on behalf of the Bertelsmann Foundation (2016).

40 Classification of groups according to AAPOR: Not eligible (NE) to participate in PASS are, in principle, deceased persons, persons who have moved to another country, and young persons under 15 years of age, as well as those who have already been interviewed. Unknown eligible (UE) are households that have moved, incorrect addresses, and wrong telephone numbers. The non-response (NR) group comprises all non-responses. These non-responses are differentiated into non-contact (NC), not able (NA), refusal (R), and others (O). The “others” group includes all households with whom the interview, for various reasons, could not be conducted in the course of the fieldwork period, but who were, in principle, willing to participate.

households became apparent in a great willingness to participate again in the next wave (97.7% panel consent). Willingness to participate again is the main challenge in longitudinal studies. The high panel consent of Syrian/Iraqi households gives hope that refugees can be integrated in the PASS panel study.



AAPOR final disposition code, RR1, household level; figures in percent

Source: Panel Study Labour Market and Social Security (PASS), 2016 Wave 10, contact data (infas)

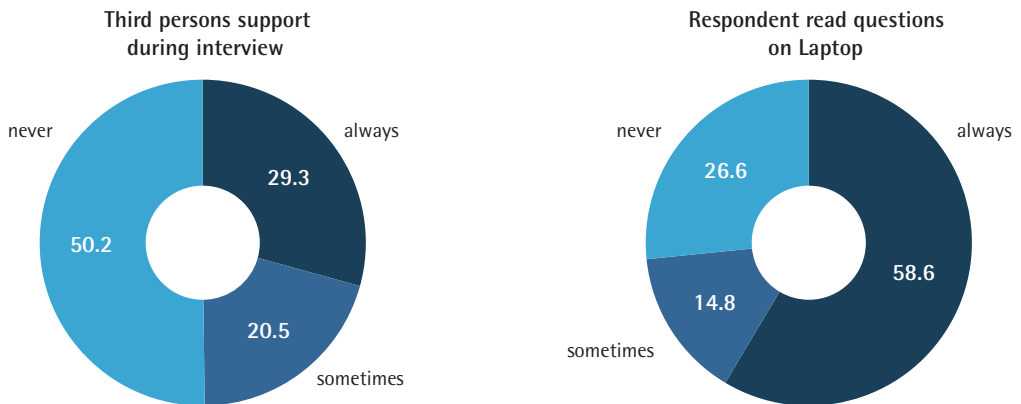
Figure 1 Final disposition code for sample Syrians/Iraqis via mode

2.3 Tailored Interviewing Strategies for the Refugee Sample

In general, there is a standardized questionnaire available in PASS. The survey instruments are identical for both telephone and face-to-face interviews. When conducting the interviews, interviewers follow the rules of standardized interviewing (Fowler & Mangione, 1990). 1:1 translations of the PASS questions were undertaken to produce the foreign-language versions of the questionnaire, which means that the whole German questionnaire was translated without any adaptation to Arabic-specific contents or specific conditions for respondents from Syria/Iraq. Interviews in the foreign languages should, in principle, follow the questionnaire and be conducted in a standardized fashion. Assistance and translations by third parties have not been allowed in the foreign-language interviews in Turkish and Russian during the last waves in PASS. However, in order to facilitate interviewing the Arabic-speaking households, deviations from the standard rules for conducting interviews were introduced in CAPI in wave 10. The main reason for this was to enable interviewers with no knowledge of Arabic to conduct interviews with persons who were not able to follow a long interview in German. The interviews in PASS have an average length of 45 minutes with the head of household and 30 minutes with each additional person of at least

15 years of age. A key deviation from the standard interview was a situation in which the interviewer and the respondent were not, as usual, sitting opposite but next to each other, thereby making it possible to jointly follow the questionnaire on the screen. The technical possibility of changing the language displayed on the screen was advantageous for this procedure. Thus, for each question, the interviewer and the respondent could decide whether the German or the Arabic version should be displayed. For Arabic, the change in the direction of the writing was, of course, heeded. Therefore, the interviewers could guide respondents through the questionnaire, even if they did not understand the texts displayed in Arabic. They just started with the German question on the screen and then switched to the Arabic version so that the respondents could read the question themselves. A further deviation from the standard interview involved support from third parties. In the case of Arabic-language interviews, assistance from third parties was allowed. Besides helping interviewers to steer the communication and explain tasks to the respondents, third parties also could help reading out questions and sometimes explaining their meaning. This was necessary due to the fact that not all respondents from the refugee sample were capable of reading and sometimes they needed help to understand questions with higher complexity.

The option to read together on the screen was chosen in almost 60% of the interviews conducted, whereas steady support from third parties occurred in almost one third of the interviews (see Figure 2). Support included help in reading out and explanations as described above. The consequences of this procedure for the quality of the data cannot be shown yet, since they are the object of further analyses. The feedback of the interviewers on the individual interviews provided some first indications of possible problems and further needs to revise the PASS questionnaire. Selected questions from the interviewer feedback will be described below.



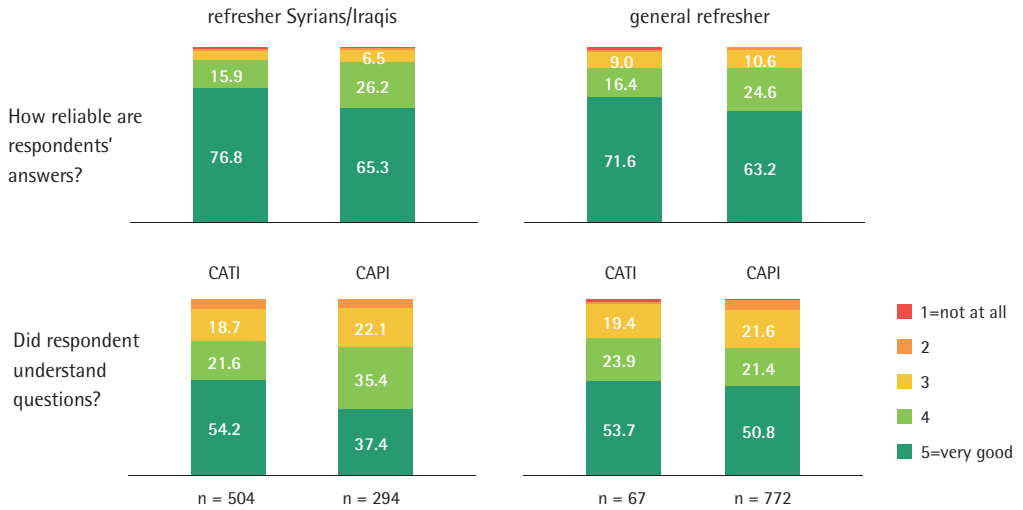
Refresher Syrians/Iraqis with CAPI mode, interviewers' rating on individuals level n=263; figures in percent

Source: Panel Study Labour Market and Social Security (PASS), 2016 Wave 10, paradata (infas)

Figure 2 Interview setting during CAPI

2.4 Interviewers' Assessment of Interview Situation with Refugees

At the end of each interview, the interviewers were asked to provide an assessment of the interview situation and of the respondents' behavior. These assessments took place at the end of both the CATI and the CAPI interview. Comparing the interviewer assessments from the two survey modes allowed us to identify possible effects of the deviations from standardized interviewing. Thus, for instance, the interviewers were supposed to estimate how often respondents had asked for explanations during the interview. In the refugee sample, respondents requested explanations more rarely in the CATI interview than in the CAPI interview (Resp never/almost never ask for explanation: CATI 57.9%, CAPI 42.5%). But this difference between the survey modes could also be observed for the general refresher sample (Resp never/almost never ask for explanation: CATI 73.1%, CAPI 58.9%). The greater need for explanation may depend less on the de-standardization in the CAPI interview though. It seems that explanations in the CATI interviews in general were less needed than in the CAPI interviews. However, the extent of the need for explanation was greater in the refugee sample than in the general refresher sample. This finding confirms reports from interviewers saying that questions in PASS were difficult for the migrants to understand. Further assessments of the interviewers – for instance about the reliability of the respondents' answers – also revealed a difference between the survey modes, but not between the samples, as is shown in Figure 3. It is only in the case of the assessment of the questions' understandability that an effect emerged among the Syrian/Iraqi households that were interviewed in CAPI. While a very good understanding of the questions was observed for half of the respondents in the general refresher sample and half of the refugees surveyed in CATI, it seems to have been considerably worse for the refugees surveyed in CAPI. It is possible that the respondents' reading the questions on the screen was not, after all, sufficient enough to enable them to answer the questions. Moreover, in light of this finding, one can assume that the CATI interviewers may have provided additional assistance to the Syrian/Iraqi households during the interviews. The PASS questionnaire includes some questions with particularities on the topic of unemployment benefits, which are difficult to understand for refugees unfamiliar with the German welfare system. Furthermore, the translation into Standard Arabic requires a higher level of education since Standard Arabic is the written form of Arabic and not used in everyday language.



Refresher sample, individual level, mode of interview; figures in percent

Source: Panel Study Labour Market and Social Security (PASS), 2016 Wave 10, paradata (infas)

Figure 3 Interviewers' rating on respondents' comprehension

3 Discussion

A great willingness to participate in PASS and high motivation were apparent among refugees from Syria and Iraq. It was possible to achieve a considerably higher response rate in the refugee sample than in the general refresher sample. This experience gives reason to be optimistic, also with regard to the next survey wave and the involvement of more migrants in PASS. It was possible to gain access to the refugee households both by telephone and face-to-face. It was also helpful to adapt contacting and survey strategies for this target group. In CAPI, deviations from the standardized interview were allowed to enable CAPI interviewers without foreign-language abilities to interview persons in the refugee households. The interviewer assessments, which were provided at the end of each interview with refugee households in both modes, indicate difficulties in understanding as well as further need for clarification but do not necessarily refer to differences in the collected data. The consequences on the data collected still have to be analyzed more closely. However, feedback from the interviewers on difficulties dealing with particularities of the German welfare system or other complex questions suggest that optimizing the translation of individual questions surely makes sense.

Non-responses occurred to a larger extent in the refresher sample due to refusals, whereas they were caused in the refugee sample by lower reachability. The difficulty of reaching respondents in CATI was particularly disadvantageous due to the fact that refugees very frequently provide cell phone numbers. Cell phone numbers bear the risk of

changing more frequently and being invalid. In order to remain in contact with refugees, it will be necessary to develop additional tracking strategies in PASS. The reachability of respondents in the follow-up surveys will be of major significance in order to avoid panel attrition. The integration of refugees in the panel can only succeed if longitudinal data about them can be collected over the course of several survey waves.

References

- Biemer, P. (2010). Overview of design issues: Total survey error. In P. Marsden & J. Wright, (Eds.), *Handbook of survey research* (2nd, pp. 27–58). Bingley: Emerald.
- Brücker, H., Rother, N. & Schupp, J. (Ed.) (2016). *IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse*. IAB-Forschungsbericht 14/2016, Nürnberg.
- Couper, M. P., & Ofstedal, M. B. (2009). Keeping in contact with mobile sample members. In P. Lynn (Ed.), *Methodology of longitudinal surveys*. Chichester: Wiley.
- Dutwin, D., Loft, J. D., Darling, J., Holbrook, A., Johnson, T., Langley, R. E., ... & Zukerberg, A. (2014). AAPOR Report: Current knowledge and considerations regarding survey refusals. Released September 8, 2014. Retrieved from: <https://www.aapor.org/Education-Resources/Reports/Survey-Refusals.aspx>
- Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing*. Newbury Park, CA: SAGE.
- Gonser, M., Jesske, B., & Pfeiffer, F. (2017). *Werkstattbericht Reallabor-Befragungen von Geflüchteten in der Rhein-Neckar-Region*. infas Bonn, Mannheim.
- Groves, R., & Couper, M. (1998). *Nonresponse in household interview surveys*. New York (N.Y.): Wiley.
- Johannson, S., Schiefer, D., & Andres, N. (ed.) (2016). *Forschungsbereich beim Sachverständigenrat deutscher Stiftung für Integration und Migration (SVR)(2016): Was wir über Flüchtlinge (nicht) wissen*. Berlin.
- Jesske, B., Knerr, P., & Kraft, L. (2017). *Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS. 10. Erhebungswelle – 2016 (Haupterhebung)*. FDZ Methodenreport 09/2017. Nürnberg.
- The American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. (9th ed.). Lanexa: Kansas.
- Trappmann, M., Beste, J., Bethmann, A., & Müller, G. (2013). The PASS panel survey after six waves. *Journal for Labour Market Research*, 46(4), 275–281.

Surveying Immigrants

The Role of Language Attrition and Language Change in the Application of Questionnaires

Patrick Brzoska

Abstract

Quantitative questionnaires are increasingly applied across different languages and cultures. They play an important role in research on immigrant communities as well as in surveys of the general population of which immigrants are a significant part. Because immigrants have often limited proficiency of the language of the host country, usually questionnaires in their first language (“mother tongue”) are needed. For this purpose, it is common to use questionnaires developed or adapted for the population of the respective countries which immigrants originate from. The applicability of these questionnaires, however, may be limited because of differences between both populations in terms of language usage, amongst others resulting from language attrition. This, for example, can be illustrated by means of the Revised Illness Perception Questionnaire (IPQ-R) applied to Turks in Turkey and Turkish immigrants residing in Germany. Questionnaires, therefore, must be thoroughly tested and often re-adapted to the language style and level of language proficiency of immigrants to avoid different forms of bias. An analytical framework of questionnaire equivalence can guide the process of testing and re-adaptation.

1 Introduction

Quantitative questionnaires are often applied across different countries and languages. Because of growing diversity and international migration, practitioners and researchers in many fields are also faced with the need to collect self-reported information from immigrant communities within a single country. Consequently, the need for multi-cultural and multi-language adaptations of research instruments is increasing, further stimulated by growing interest in cross-cultural research in general (Porter & Gamoran, 2002; Harkness, 1998; Hanna, Hunt & Bhopal, 2008).

Survey research among immigrants is associated with different challenges encompassing the entire research process. Many of these challenges go beyond the problems encountered in cross-national and cross-cultural survey research in general. Given their special characteristics, immigrants are often difficult to define and hard to reach. This makes the application of sophisticated sampling techniques and field work necessary in order to define adequate sampling frames and to collect high-quality data (see also Jacobsen

or Jesske in the present volume). One particular challenge concerns the application of quantitative questionnaires. As many immigrants have limited proficiency of the host country's main language(s), it is usually necessary to also provide questionnaires in their first language, that is, their 'mother tongue'. In the social, behavioral, and health sciences, frequently used questionnaires are available in different language versions. They have often been validated in the countries where immigrants originate from. In survey research on immigrants, it is common to make use of these questionnaires—often without further re-adaptation. For example, questionnaires developed for the population of Turkey are usually applied for the study of Turkish immigrants residing in Germany or the Netherlands. However, the first language spoken by immigrants may differ from the language spoken in their countries of origin in terms of their lexical, syntactical, and semantic characteristics as a result of a differential language development over time, amongst others resulting from language attrition.

The present paper illustrates how these differences may affect the functioning of questionnaires on several levels and how they may be responsible for a low performance of questionnaires frequently reported in migration research. For this purpose, the paper uses the study of illness perceptions in Turks residing in Turkey and Turkish migrants residing in Germany as an example. In the following, first a brief introduction into the language usage among immigrants and the role of language attrition is given, thereby also illustrating why users of Turkish as a first language living outside Turkey may be particularly prone to this phenomenon. Second, the construct of illness perceptions is described, which in this paper serves as an exemplary outcome to examine the transferability of psychometric questionnaires between immigrants and their population of origin. Third, results from studies on the functional equivalence of the Revised Illness Perception Questionnaire (IPQ-R) applied to Turks in Turkey and Turkish immigrants in Germany are presented, showing that the functioning of the questionnaire among Turkish immigrants may be limited despite a good performance among Turks in Turkey. Finally, recommendations are given on how differences in language usage between immigrants and their population of origin can be considered by means of thorough testing and re-adaptation when research instruments are to be used across both populations.

2 Language Attrition and Differential Language Development

In survey research, it is well acknowledged that the psychometric properties of questionnaires may differ when they are applied across users of the same language residing in different countries. Commonly used survey instruments such as the Short Form Health Survey 36 (SF-36) assessing health-related quality of life are therefore available in different varieties of the same language spoken in different regions (e.g., Spanish for Spain, Argentina, and the US or English for the US, Taiwan, and Australia). The differences between versions are reflected, for instance, in a different vocabulary or in different idiomatic language use. These differences are necessary as the meaning of words or the familiarity with certain expressions may differ between geographical regions. In addition, some words and expres-

sions possibly do not exist in some countries, regions or cultures (Spielberg, Moscoso, & Brunner, 2005). This is of particular relevance in cross-cultural research on health, since expressions describing health and illness are known to vary across (sub)cultures (Anastasi, 1972).

These differences in language usage between speakers of the same language are also relevant for immigrant research. With increasing time that immigrants spend in the host country, their usage of the first language may become different from that of the source population with regard to grammar, vocabulary, and semantics. While some differences only comprise lexical extension (e.g., to adapt to the new circumstances of living), others may reflect a deterioration of language skills resulting from language attrition (Pavlenko, 2004; Backus, 2006; Seliger & Vago, 1991).

Language attrition refers to the “loss of any language or any portion of a language by an individual or a speech community [...] [including] the declining use of mother tongue skills by those in bilingual situations or among ethnic minorities in (some) language contact situations where one language, for political or social reasons comes to replace another” (Lambert & Freed, 1982, p. 1). It particularly affects complex syntactical and lexical language features that are acquired late in the language development process, and that are complex and less frequently used. Language attrition may especially occur in contact situations involving two or more language families (*ibid.*).

Turkish is an Altaic language and is thus fundamentally different from the languages of the main host countries that Turkish immigrants reside in, such as German, Dutch, French, and English. Being languages of the Indo-European language family, their syntactical and lexical characteristics differ substantially from Turkish. This could increase the susceptibility of Turkish speakers to language attrition. As an agglutinative language, Turkish joins morphemes to form words/expressions and to fulfill grammatical and syntactical functions – a linguistic process referred to as ‘agglutination’. An example for this feature is an expression such as *arabanızdaydım* (‘I was in your car’), combining a noun (*araba*) with morphemes indicating possession (*-nız-*), location (*-da[y]-*), time (*-d-*), and personal pronoun (*-im-*). In contrast to Turkish, many Indo-European languages such as English and German are fusional. In fusional (or inflectional) languages, “the separation between morphemes is not readily apparent. Characteristically, in such languages inflectional morphemes each express two or more categories (for example, number/case in the noun, or tense/person/number in the verb)” (Harris & Zu, 2006, p. 513).

Due to geographical distance and isolation from the development of the Turkish language in Turkey, Turkish immigrants are considered by some authors to lose the “feeling for their language” (Özata, 1993, p. 102), which comprises a loss of their first language’s lexical, semantic, and syntactical properties. As a result, they have developed a new variation of the Turkish language, which is referred to as “contact Turkish”, a “mixed-language” or “Immigrant Turkish” (Özata, 1993; Schroeder, 2003; Backus, 2006, 1992; Tekinay, 1983b, 1983a, 1987). Some even consider it to be a new dialect of Standard Turkish primarily based on oral communication (Rehbein, 2001).

These assumptions are also supported by empirical studies which provide evidence that the Turkish language used by first-generation Turkish immigrants is influenced by the lexical, syntactical, and grammatical structure of the respective host country’s language in

various aspects (Boeschoten & Verhoeven, 1985; Tekinay, 1983b, 1987). This may become evident in different forms of code-switching. They may occur within syntactical entities where expressions of the contact language replace equivalents of the Turkish language such as in the examples *şıpasiren yapmak* (from German 'spazieren gehen', i.e. 'go for a walk'; instead of Turkish *gezmeye gitmek*), *überşutun yapmak* (from German 'Überstunden machen', i.e. 'to work overtime'; instead of Turkish *fazla mesai yapmak*) and *kaufu gitmek* (from German 'einkaufen gehen', i.e., 'go shopping'; instead of Turkish *alışverişe çıkmak*) (Tekinay, 1983b, 1987). Similar observations were made by Boeschoten and Verhoeven (1987). They identified patterns of code-switching in terms of insertion among first-generation adult Turkish immigrants in the Netherlands who tended to incorporate Dutch words into their Turkish language communication. Code-switching patterns can also include mixing on the syntactical level of the language comprising declension suffixes such as in the example *farşuleye* (from German 'in die Fahrschule', i.e., '[to go] to driving school'; instead of Turkish *sürücü kursuna*), prepositions such as in *ubanla* (from German 'mit der U-Bahn', i.e., 'by subway'; instead of Turkish *metroyla*), and suffixes such as in *kinderler* (from German 'Kinder', i.e., 'children'; instead of Turkish *çocuklar*) (Boeschoten & Verhoeven, 1985; Tekinay, 1983b, 1987).

In the first place, code-switching concerns the active use of language in oral communication and in most instances is less relevant for the understanding of psychometric questionnaires. Studies from the Netherlands and Australia, however, also provide empirical evidence for attrition in the syntactical structure of the language used by Turkish immigrants which also affects the passive use of language and has thus also implications for the understanding of questionnaires. By means of a comparison of the language usage of Turkish immigrants and Turks residing in Turkey, this syntactical attrition became evident in a lower complexity of sentences, a lower linguistic test performance, a lower self-reported fluency in the Turkish language, and in a higher self-reported difficulty to understand jokes and ironic comments (Huls & van de Mond, 1992; Yagmur, de Bot, & Korzilius, 1999; Gürel, 2004). The studies also showed that both the lexical and syntactical attrition of the Turkish language spoken by Turkish immigrants was positively correlated with the length of stay.

As most languages, also Turkish dynamically develops over time and is therefore continuously changing (Can & Patton, 2010; Dogancay-Aktuna, 1995; Brendemoen, 1998). In the case of Turkish, this development is also characterized by a substantial increase in neologisms which have entered the Turkish language in the 20th and 21st century. In part, this is attributable to an ongoing and government-supported language reform, which aims to modernize the Turkish language. As a consequence, words of Arabic or Persian origin have been replaced by Turkish equivalents (Tekinay, 1983a; Lewis, 1999). Amongst others, this included back-formation from the old Turkish language (e.g., *sınamak* is nowadays more common than *imtihan etmek* ['to try', 'to examine']), loan words such as *ruhbilim* instead of *psikoloji* ('psychology'), and new formations such as *seçenek* instead of *alternatif* ('alternative') (Dogancay-Aktuna, 1995; Brendemoen, 1998). Since many Turkish immigrants, who today constitute large communities throughout Europe, had emigrated in the 1950s through 1970s and had almost no access to Turkish-language media until the mid-1990s, they were less exposed to the development of the Turkish language that was taking

place in Turkey during that time (Sirim, 2009). This may result in the use of many Ottoman Turkish language terms, such as *münazara yapmak* instead of *tartışmak* ('to discuss'), that may be less known to new generations as empirical evidence suggests (Özata, 1993).

3 The Construct of Illness Perceptions and its Measurement

Illness perceptions refer to beliefs patients develop when they fall ill. Illness perceptions are usually conceptualized by means of the Self-Regulatory Model of Illness (SRM) (Leventhal, Brissette, & Leventhal, 2003), which considers health and illness behavior to be influenced by previous experiences and subjective theories of illness. Patients' perceptions develop around the answers they find to certain questions emerging in the case of illness. They can be clustered into different dimensions: How are my symptoms related to my illness (Identity)? How long will the illness last (Timeline)? What has caused my condition (Cause)? How will it affect me (Consequences)? Can it be controlled by personal actions and/or treatment (Control)? In addition to these cognitive beliefs, patients develop emotional representations such as fear of possible disease consequences.

Illness perceptions are associated with physical functioning, psychological distress, psychological well-being, role functioning, social functioning, vitality, and disease outcomes (Jones, Smith, & Llewellyn, 2016; Hagger & Orbell, 2003). Congruent illness beliefs between patients and doctors are important for optimal disease management and coping with illness. Likewise, knowledge about patients' illness perceptions is essential to establish an effective patient-provider relationship. This makes the valid assessment of illness perceptions in research and practice highly relevant.

Illness perceptions are commonly assessed by means of the Revised Illness Perception Questionnaire (IPQ-R) (Moss-Morris et al., 2002). The IPQ-R consists of three parts. The first assesses illness identity through 14 binary items. The second part measures the dimensions of timeline (consisting of the sub-dimensions acute/chronic timeline and cyclical timeline), consequences, control (distinguishing between beliefs related to the controllability by means of treatment and by means of personal actions), coherence, and emotional representations by means of 38 items. The items are rated on a five-point response scale from "1=strongly agree" to "5=strongly disagree" (see Appendix, Tab. 1 for an overview of exemplary items). The third part, using the same response format, assesses beliefs patients hold about potential causes of their illness.

The validity of the IPQ-R has been confirmed for different diseases such as diabetes (Searle, Norman, Thompson, & Vedhara, 2007; Abubakari et al., 2013), cardiac diseases (Cooper, Lloyd, Weinman, & Jackson, 1999; Petrie, Cameron, Ellis, Buick, & Weinman, 2002), and cancer (Hagger & Orbell, 2005). The questionnaire was also translated into different languages. While in many of these non-English versions only exploratory examinations have been conducted, confirmatory factor analyses (CFA) exist for only a few languages, for instance, Chinese (Chen, Tsai, & Lee, 2008) and Swedish (Brink, Alsén, & Cliffordson, 2011). Although many of these studies identified minor sources of poor model fit, the general measurement model of the questionnaire has been confirmed in most cases.

The Turkish version of the questionnaire was translated by Brzoska et al. (2012). Its validity was examined by means of a survey of 302 patients with diabetes mellitus type 2 and/or cardiovascular disease in out-patient clinics in Turkey. Its factorial validity was examined by means of CFA. Similar to other language validations, the questionnaire exhibited minor sources of poor model fit, resulting in the deletion of four of the 38 items of the original measurement model. The remaining 34 items loading on seven dimensions (Fig. 1) showed a satisfactory model fit.

4 Functional Equivalence Testing in Turkish Migrants and Turks Residing in Turkey

Immigrants are often surveyed by means of questionnaires that have been translated and/or adapted for the populations of the countries they originate from. The performance of these questionnaires has been reported to be limited among immigrants (see for example Brzoska & Razum, 2010; Mewes, Christ, & Rief, 2009). In order to examine reasons for this limited performance, Brzoska (2014) conducted a mixed-method study in Turks residing in Turkey and Turkish migrants residing in Germany based on a comprehensive analytical framework of equivalence covering conceptual, semantic, and measurement characteristics of the IPQ-R. To examine differences in the psychometric characteristics of the questionnaire between the two populations and to identify sources of poor model fit, the measurement invariance of the questionnaire was tested by means of multi-group confirmatory factor analysis (MGCFA) and multiple indicators, multiple causes (MIMIC) models using data from 601 patients of Turkish origin in Turkey and Germany.

Several areas of poor model fit in the Turkish IPQ-R were identified in the sample of Turkish immigrants. Amongst others, five of the 34 items of the Turkish IPQ-R had to be deleted because of low factor loadings resulting in a solution which comprised six of the original seven IPQ-R factors and which thus only fulfilled partial configural invariance. Only 19 of the remaining 29 items exhibited scalar invariance. This means that Turkish immigrants and Turks in Turkey provided different responses to ten items despite having the same position on the respective latent factor. The findings were not affected by sociodemographic/-economic factors or disease characteristics as MIMIC analyses revealed.

A limited equivalence of measurement instruments administered to Turks in Turkey and Turkish migrants in Germany were also observed by Mewes et al. (2009), who studied differential item functioning in both populations with respect to the Screening for Somatoform Disorders-2 (SOMS-2) inventory.

As for the IPQ-R, the observed measurement non-invariance between both population groups cannot be explained by a construct bias resulting from differences in the conceptualization of illness perceptions in both populations as a systematic literature review and qualitative studies show (Yilmaz-Aslan, Brzoska, Bluhm, Aslan, & Razum, 2014; Brzoska, 2014). Cognitive interviews conducted among 58 patients of Turkish origin who were asked to reflect on the perceived difficulty and usability of the IPQ-R show that several of the non-equivalent IPQ-R items were misunderstood by Turkish immigrants because

of complex item wording and unknown or ambiguous words and expressions used in the questionnaire (Brzoska & Razum, 2010). In order to study potential differences in the understanding of the questionnaire items and to examine the appropriateness of the Turkish language and style, Brzoska (2014) also conducted expert interviews with seven individuals experienced in the language usage of Turkish immigrants. In the majority of items, all respondents identified words and expressions they considered as difficult to understand for this population because of formal speech, complex wording or terms seldom used by this population group. This, for instance, comprised formal terms such as *muhtemelen* ('probably') and *gayet net* ('very clear'), for which also less formal equivalents would exist (*galiba* and *çok açık*, respectively). Complex wordings, for instance, were expressions such as *olumsuz etkileri* ('negative influence') and *farklılık gösteriyor* ('show differences'), which could also be replaced by easier to understand alternatives such as *zarar* ('consequences') and *değişiyor* ('change'), respectively. Items which are difficult to understand or which are prone to misunderstanding may contribute to an item bias resulting in measurement non-equivalence (Byrne et al., 2009; van de Vijver & Poortinga, 2005) and may indicate that both population groups differ in their usage of the Turkish language.

5 Conclusions

Immigrants who have a limited proficiency in the main language(s) of the host country are usually surveyed by means of questionnaires in their first language ('mother tongue'). Many standard questionnaires are available in several language versions, usually validated for the population of the countries where immigrants originate from. However, questionnaires developed for native populations may perform differently when administered to the respective immigrant communities. Differences in the performance can result from a differential development of language characteristics in both populations over time, including language attrition among the emigrated population. This differential variation of language usage among immigrants and their population of origin may introduce different forms of method and item bias, which reduce the structural validity of "native" questionnaires when administered to immigrants. It therefore must be considered through thorough testing and re-adaptation when research instruments are to be used across both populations. An analytical framework of questionnaire equivalence, such as that proposed by Brzoska (2014), can guide the process of testing and re-adaptation.

In cross-group comparative research, for instance, on differences in attitudes, values, and beliefs, good psychometric properties of a questionnaire in each of the separate groups to be compared do not ensure a valid cross-group comparison. Unless equivalence is established, it remains unclear whether differences observed between groups reflect true differences or are the result of a systematic measurement error. Potential sources of non-equivalence, therefore, must be taken into account in different research phases, for instance, by field testing and the use of latent variable modelling in order to ensure comparability of measurements. Not only researchers must be aware of the differences that may exist between source and immigrant populations and that need to be considered when question-

naires are to be transferred between both groups. Also funding bodies should acknowledge the need for a formal validation process prior to substantive research and provide appropriate resources accordingly.

The differences in language usage illustrated are also relevant for other fields of social research and practice. Comparable to questionnaires, also written and oral material which is used across populations, e.g., for purposes of education or information, must be evaluated in terms of its transferability. Similarly, in translations of documents which are meant to be distributed to immigrant communities, the special (linguistic) characteristics in this population must be considered. This can be facilitated by including individuals into the translation process who are familiar with the language usage of immigrants.

References

- Abubakari, A.R., Joney, M. C., Lauder, W., Kirk, A., Anderson, J., Devendra, D., & Naderali, E. K. (2013). Ethnic differences and socio-demographic predictors of illness perceptions, self-management, and metabolic control of type 2 diabetes. *International Journal of General Medicine*, 6, 617–628.
- Anastasi, A. (1972). *Psychological testing*. London: Collier-Macmillan.
- Backus, A. (2006). Turkish as an immigrant language in Europe. In T.K. Bhatia & W.C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 689–724). Balden: Blackwell.
- Backus, A. (1992). *Patterns of language mixing: A study in Turkish-Dutch bilingualism*. Wiesbaden: Harrassowitz.
- Boeschoten, H., & Verhoeven, L. (1985). Integration niederländischer lexikalischer Elemente ins Türkische. Sprachmischung bei Immigranten der ersten und zweiten Generation. *Linguistische Berichte*, 98, 437–464.
- Boeschoten, H., & Verhoeven, L. (1987). Language-mixing in children's speech: Dutch language use in Turkish discourse. *Language Learning*, 37, 191–215.
- Brendemoen, B. (1998). The Turkish Language Reform. In L. Johanson & E.A. Csato (Eds.): *The Turkic Languages* (pp. 242–247). New York: Routledge.
- Brink, E., Alsén, P., & Cliffordson, C. (2011). Validation of the Revised Illness Perception Questionnaire (IPQ-R) in a sample of persons recovering from myocardial infarction – the Swedish version. *Scandinavian Journal of Psychology*, 52, 573–579.
- Brzoska, P. (2014). *Psychometrically relevant differences between source and migrant populations*. Frankfurt am Main: Peter Lang.
- Brzoska, P., & Razum, O. (2010). *Validity issues in quantitative migrant health research. The example of illness perceptions*. Frankfurt am Main: Peter Lang.
- Brzoska, P., Yilmaz-Aslan, Y., Sultanoglu, E., Sultanoglu, B., & Razum, O. (2012). The factor structure of the Turkish version of the Revised Illness Perception Questionnaire (IPQ-R) in patients with diabetes and cardiovascular disease. *BMC Public Health*, 12, 852.
- Byrne, B.M., Oakland, T., Leong, F. T. L., Vijver, F. J. R. v. d., Hambleton, R. K., & Cheung, F. M. (2009). A critical analysis of cross-cultural research and testing practices: Impli-

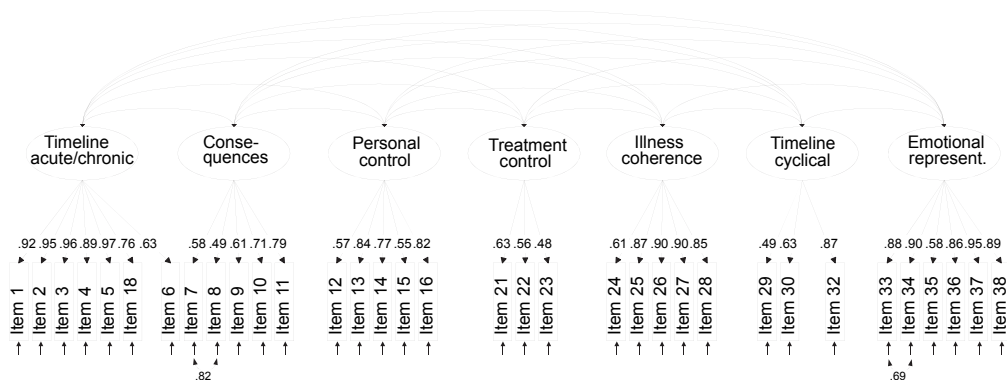
- cations for improved education and training in psychology. *Training and Education in Professional Psychology*, 3, 94–105.
- Can, F., & Patton, J. M. (2010). Change of word characteristics in 20th-Century Turkish literature: A statistical analysis. *Journal of quantitative linguistics*, 17, 167–190.
- Chen, S.L., Tsai, J. C., & Lee, W. L. (2008). Psychometric validation of the Chinese version of the Illness Perception Questionnaire-Revised for patients with hypertension. *Journal of Advanced Nursing*, 64, 524–534.
- Cooper, A., Lloyd, G., Weinman, J., & Jackson, G. (1999). Why patients do not attend cardiac rehabilitation: role of intentions and illness beliefs. *Heart*, 82, 234–236.
- Dogancay-Aktuna, S. (1995). An evaluation of the Turkish language reform after 60 years. *Language Problems and Language Planning*, 19, 221–249.
- Gürel, A. (2004). Attrition in L1 competence. The case of Turkish. In M. Schmid, B. Köpke, M. Keijzer & L. Weilemar (Eds.), *First language attrition. Interdisciplinary perspectives on methodological issues* (pp. 225–242). Amsterdam: John Benjamins Publishing Company.
- Hagger, M.S., & Orbell, S. (2003). A meta-analytic review of the common-sense model of illness representations. *Psychology and Health*, 18, 141–184.
- Hagger, M.S., & Orbell, S. (2005). A confirmatory factor analysis of the revised illness perception questionnaire (IPQ-R) in a cervical screening context. *Psychology and Health*, 20, 161–173.
- Hanna, L., Hunt, S., & Bhopal, R. (2008). Insights from research on cross-cultural validation of health-related questionnaires. *Current Sociology*, 56, 115–131.
- Harkness, J. (1998). *Cross-cultural survey equivalence*. Mannheim: ZUMA.
- Harris, A., & Zu, Z. (2006). Diachronic morphological typology. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 509–515). Oxford: Elsevier.
- Huls, E., & van de Mond, A. (1992). Some aspects of language attrition in Turkish families in the Netherlands. In W. Fase, K. Jaspaert & S. Kroon (Eds.), *Maintenance and loss of minority languages* (pp. 99–116). Amsterdam: John Benjamins.
- Jones, C.J., Smith, H. E., & Llewellyn, C. D. (2016). A systematic review of the effectiveness of interventions using the Common Sense Self-Regulatory Model to improve adherence behaviours. *Journal of Health Psychology*, 21, 2709–2724.
- Lambert, R. D., & Freed, B. F. (1982). *The loss of language skills*. Rowley: Newbury House.
- Leventhal, H., Brissette, L., & Leventhal, E. A. (2003). The common-sense model of self-regulation of health and illness. In L.D. Cameron & H. Leventhal (Eds.), *The self-regulation of health and illness behaviour* (pp. 42–65). New York: Routledge.
- Lewis, G. (1999). *The Turkish Language Reform: A Catastrophic Success*. Oxford: Oxford University Press.
- Mewes, R., Christ, O., & Rief, W. (2009). Ein Vorgehen zur Überprüfung psychometrischer Kennwerte in der Migrationsforschung. *Klinische Diagnostik und Evaluation*, 2, 106–118.
- Moss-Morris, R., Weinman, I., Petrie, K. J., Horne, R., Cameron, L. D. & Buick, D. (2002). The Revised Illness Perception Questionnaire (IPQ-R). *Psychology and Health*, 17, 1–16.

- Özata, M. (1993). Die türkische Sprache in Berlin. In Berliner Geschichtswerkstatt e.V. (Ed.), „...da sind wie keine Ausländer mehr“. *Eingewanderte ArbeiterInnen in Berlin 1961-1993* (pp. 100–102). Berlin: Berliner Geschichtswerkstatt e.V.
- Pavlenko, A. (2004). L2 influence and L1 attrition in adult bilingualism. In M.S. Schmid, B. Kopke, M. Keijler & L. Weilemar (Eds.), *Interdisciplinary perspectives on methodological issues* (pp. 47–59). Amsterdam: John Benjamins.
- Petrie, K.J., Cameron, L. D., Ellis, C. J., Buick, D., & Weinman, J. (2002). Changing illness perceptions after myocardial infarction: An early intervention randomized controlled trial. *Psychosomatic Medicine*, 64, 580–586.
- Porter, A. C., & Gamoran, A. (2002). *Methodological advances in cross-national surveys of educational achievement*. Washington: National Academic Press.
- Rehbein, J. (2001). Turkish in European Societies. *Lingua e Stile*, 36, 317–334.
- Schroeder, C. (2003). Der Türkischunterricht in Deutschland und seine Sprache(n). *Zeitschrift für Fremdsprachenforschung*, 14, 23–39.
- Searle, A., Norman, P., Thompson, R., & Vedhara, K. (2007). Illness representations among patients with type 2 diabetes and their partners: Relationships with self-management behaviors. *Journal of Psychosomatic Research*, 63, 175–184.
- Seliger, H. W., & Vago, R. M. (1991). The study of first language attrition: An overview. In H.W. Seliger & R.M. Vago (Eds.), *First language attrition* (pp. 3–16). Melbourne: Cambridge University Press.
- Sirim, E. (2009). *Das Türkische einer deutsch-türkischen Migrantinnengruppe*. Institut für Deutsche Sprache.
- Spielberg, C. D., Moscoso, M. S., & Brunner, T. M. (2005). Cross-cultural assessment of emotional states and personality traits. In R.K. Hambleton, P.F. Merenda & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 343–367). Mahwah: Lawrence Erlbaum Associates.
- Tekinay, A. (1983a) Neuschöpfungen im Türkischen. *Muttersprache: Zeitschrift zur Pflege und Erforschung der Deutschen Sprache*, 83, 171–180.
- Tekinay, A. (1983b). Wie eine „Mischsprache“ entsteht. *Muttersprache: Zeitschrift zur Pflege und Erforschung der Deutschen Sprache* 93, 396–403.
- Tekinay, A. (1987). Deutsche Einflüsse im Türkischen von Arbeitsmigranten. In A. Tekinay (Ed.), *Sprachvergleich Deutsch-Türkisch: Grenzen und Möglichkeiten einer kontrastiven Analyse* (pp. 96–103). Wiesbaden: Reichert.
- van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R.K. Hambleton, P. F. Merenda & C. D. Spielberg (Eds.), *Adapting educational tests and psychological tests for cross-cultural assessment* (pp. 39–64). Mahwah: Lawrence Erlbaum.
- Yagmur, K., de Bot, K., & Korzilius, H. (1999). Language attrition, language shift and ethno-linguistic vitality of Turkish in Australia. *Journal of Multilingual and Multicultural Development*, 20, 51–69.
- Yilmaz-Aslan, Y., Brzoska, P., Bluhm, M., Aslan, A., & Razum, O. (2014). Illness perceptions in Turkish migrants with diabetes: A qualitative study. *Chronic Illness*, 10, 107–121.

Appendix

Table 1 Dimensions and sample items of the IPQ-R part II (Moss-Morris et al., 2002)

IPQ-R part II dimension	Number of items	Sample items
Acute/chronic timeline	6	“My illness will last a short time” “I expect to have this illness for the rest of my life”
Cyclical timeline	4	“My illness is very unpredictable” “The symptoms of my illness change from day to day”
Consequences	6	“My illness is a serious condition” “My illness causes difficulties for those who are close to me”
Personal control	6	“There is a lot I can do to control my illness” “What I do can determine whether my illness gets better or worse”
Treatment control	5	“Treatment will be effective in treating my illness” “There is nothing that can help my illness”
Illness coherence	5	“The symptoms of my illness are puzzling to me” “I have a clear picture or understanding of my illness”
Emotional representations	6	“My illness makes me feel afraid” “I get depressed when I think about my illness”



Note. The coefficients displayed on the arrows signify completely standardized factor loadings (straight arrows) or error covariances (curved arrows below the items). All coefficients are significant at $p < 0.05$. Coefficients for factor covariances are not displayed.

Figure 1 Measurement model of the Turkish IPQ-R part II (based on data from Brzoska et al., 2012)

Contributing Authors

Behr, Dorothee

GESIS-Leibniz Institute for the Social Sciences
Department of Survey Design and Methodology
Mannheim, Germany
dorothee.behr@gesis.org

Briceno-Rosas, Roberto

GESIS-Leibniz Institute for the Social Sciences
Department of Survey Design and Methodology
Mannheim, Germany
roberto.briceno-rosas@gesis.org

Brzoska, Patrick

Martin Luther University Halle-Wittenberg
Faculty of Medicine
Haale/Saale, Germany
patrick.brzoska@medizin.uni-halle.de

Comanaru, Ruxandra

NatCen Social Research
Questionnaire Development and Testing Hub
London, UK
Ruxandra.Comanaru@natcen.ac.uk

d'Ardenne, Jo

NatCen Social Research
Questionnaire Development and Testing Hub
London, UK
Joanna.D'Ardenne@natcen.ac.uk

García Trejo, Yazmin

U.S. Census Bureau
Center for Survey Measurement
Washington DC, US
yazmin.argen.garcia.trejo@census.gov

Goerman, Patricia

U.S. Census Bureau
Center for Survey Measurement
Washington DC, US
patricia.l.goerman@census.gov

Herzing, Jessica M. E.

University of Mannheim
Research Center 884 “Political Economy of Reforms”
Mannheim, Germany
and
University of Lausanne
FORS/LINES
Lausanne, Switzerland
jessica.herzing@unil.ch

Jacobsen, Jannes

Socio-economic Panel Study (SOEP) at
German Institute for Economic Research (DIW)
Berlin, Germany
JJacobsen@diw.de

Jesske, Birgit

infas Institute for Applied Social Sciences
Department of Data Collection and Survey Methodology
Bonn, Germany
B.Jesske@infas.de

Meyers, Mikelyn

U.S. Census Bureau
Center for Survey Measurement
Washington DC, US
mikelyn.v.meyers@census.gov

Ortmanns, Verena

GESIS-Leibniz Institute for the Social Sciences
Department of Survey Design and Methodology
Mannheim, Germany
verena.ortmanns@gesis.org

Schneider, Silke L.,

GESIS-Leibniz Institute for the Social Sciences
Department of Survey Design and Methodology
Mannheim, Germany
silke.schneider@gesis.org

Schoua-Glusberg, Alisú

Research Support Services, Inc
Evanston, US
alisu@researchsupportservices.com

Migration ist ein globales Phänomen. Die Zahl der Migranten und Flüchtlinge weltweit ist in den letzten Jahrzehnten – und insbesondere in den letzten Jahren – stark angestiegen. Mit dieser Entwicklung geht auch die Notwendigkeit einher, die Lebenssituation von Migranten und Flüchtlingen, ihre Bedürfnisse sowie die Prozesse der Integration besser zu verstehen, um so ein friedliches und nachhaltiges Zusammenwachsen in den Gesellschaften zu ermöglichen. Umfragen sind hier ein probates Mittel der Datenerhebung. Aus methodischer Sicht sind jedoch eine Reihe von Herausforderungen mit Umfragen unter Migranten und Flüchtlingen verbunden. Dieser Sammelband, der im Wesentlichen aus den Beiträgen eines internationalen Symposiums entstanden ist, fokussiert auf dem Fragebogen, seiner Übersetzung und Adaptation sowie auf kulturellen und sprachlichen Besonderheiten der Feldarbeit. Ziel dieser Publikation ist es, auf die methodischen Besonderheiten in der Migrations- und Flüchtlingsforschung aufmerksam zu machen, um die Erhebung von validen und reliablen Daten zu fördern.

Migration is a global phenomenon; the number of migrants and refugees worldwide has risen rapidly in recent decades – and especially in recent years. These developments come with the need to better understand the living conditions of migrants and refugees and the processes of integration in order to ensure peaceful and sustainable integration in societies. Surveys are an effective means of data collection in this regard. From a methodological point of view, however, a number of challenges are associated with surveys among migrants and refugees. These proceedings, which mainly draw on the contributions to an international symposium, focus on the survey questionnaire, its translation and adaptation as well as on cultural and linguistic challenges encountered during fieldwork. The goal of this publication is to draw attention to the methodological challenges in migration and refugee research in order to foster the collection of valid and reliable data.