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

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

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

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.

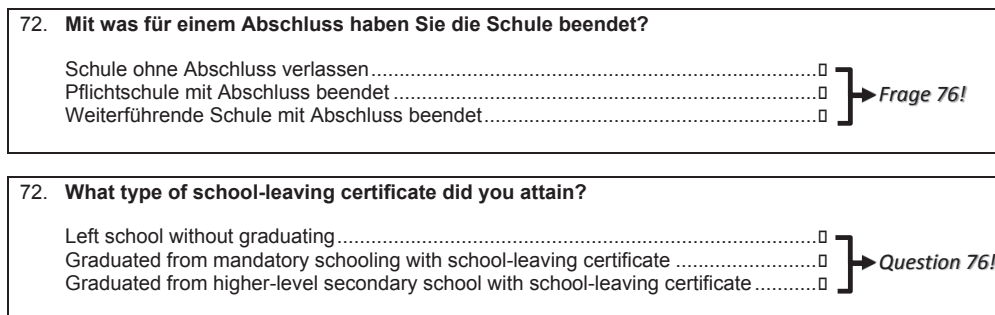# 2 Measurement of Foreign Educational Qualifications

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.

---

72. **Mit was für einem Abschluss haben Sie die Schule beendet?**

Schule ohne Abschluss verlassen...☐ ⎫
Pflichtschule mit Abschluss beendet ....☐ ⎬ ➔ *Frage 76!*
Weiterführende Schule mit Abschluss beendet...☐ ⎭

---

72. **What type of school-leaving certificate did you attain?**

Left school without graduating...☐ ⎫
Graduated from mandatory schooling with school-leaving certificate ....☐ ⎬ ➔ *Question 76!*
Graduated from higher-level secondary school with school-leaving certificate ....☐ ⎭

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 angelernt	☐ ⇨	☐ ☐ ☐ ☐
Ich habe in einem Betrieb eine längere Ausbildung gemacht	☐ ⇨	☐ ☐ ☐ ☐
Ich habe eine berufsbildende Schule besucht	☐ ⇨	☐ ☐ ☐ ☐
Ich habe eine Hochschule / Universität mit eher praktischer Ausrichtung besucht	☐ ⇨	☐ ☐ ☐ ☐
Ich habe eine Hochschule / Universität mit eher theoretischer Ausrichtung besucht	☐ ⇨	☐ ☐ ☐ ☐
Ich habe ein Promotionsstudium absolviert	☐ ⇨	☐ ☐ ☐ ☐
Sonstige Ausbildung	☐ ⇨	☐ ☐ ☐ ☐

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	☐ ⇨	☐ ☐ ☐ ☐
I completed an extended apprenticeship at a company	☐ ⇨	☐ ☐ ☐ ☐
I attended a vocational school	☐ ⇨	☐ ☐ ☐ ☐
I attended a university / college with a more practical orientation	☐ ⇨	☐ ☐ ☐ ☐
I attended a university / college with a more theoretical orientation	☐ ⇨	☐ ☐ ☐ ☐
I completed doctoral studies	☐ ⇨	☐ ☐ ☐ ☐
Other	☐ ⇨	☐ ☐ ☐ ☐

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.
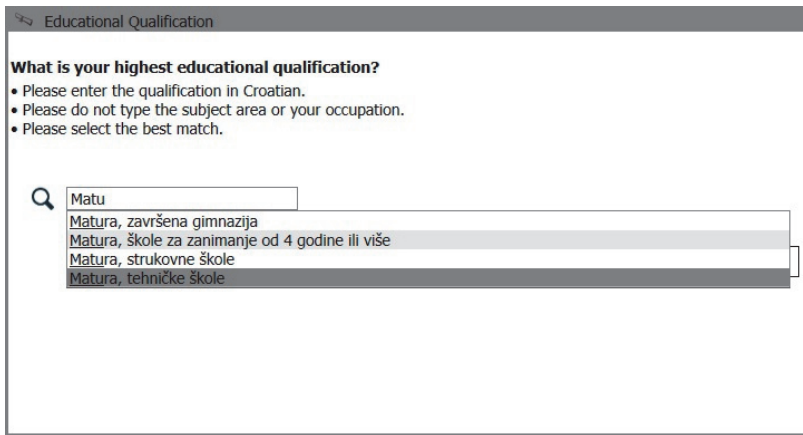

Figure 3   Interface "combination box" of the CAMCES tool version 1.0 (example shows qualifications for Croatia in an English-language survey environment)

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.
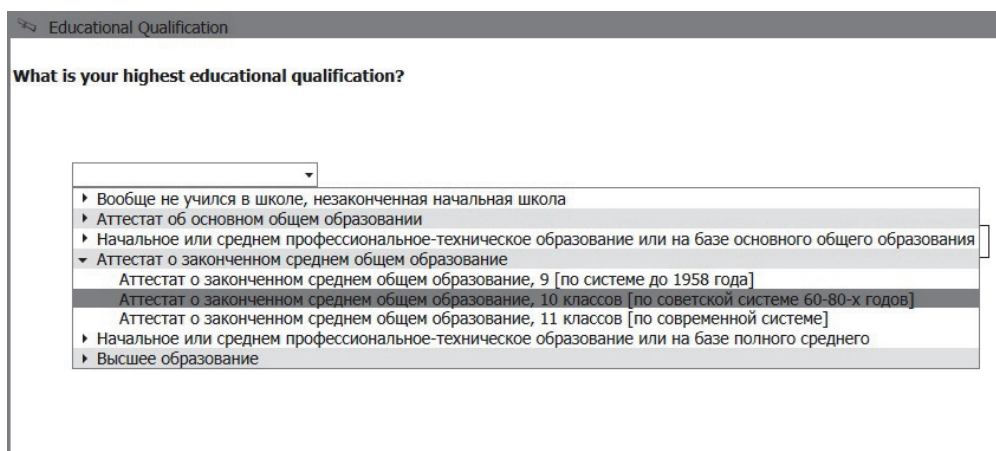

Figure 4   Interface "search tree" of the CAMCES tool version 1.0 (example shows qualifications for the Russian Federation in an English-language survey environment)

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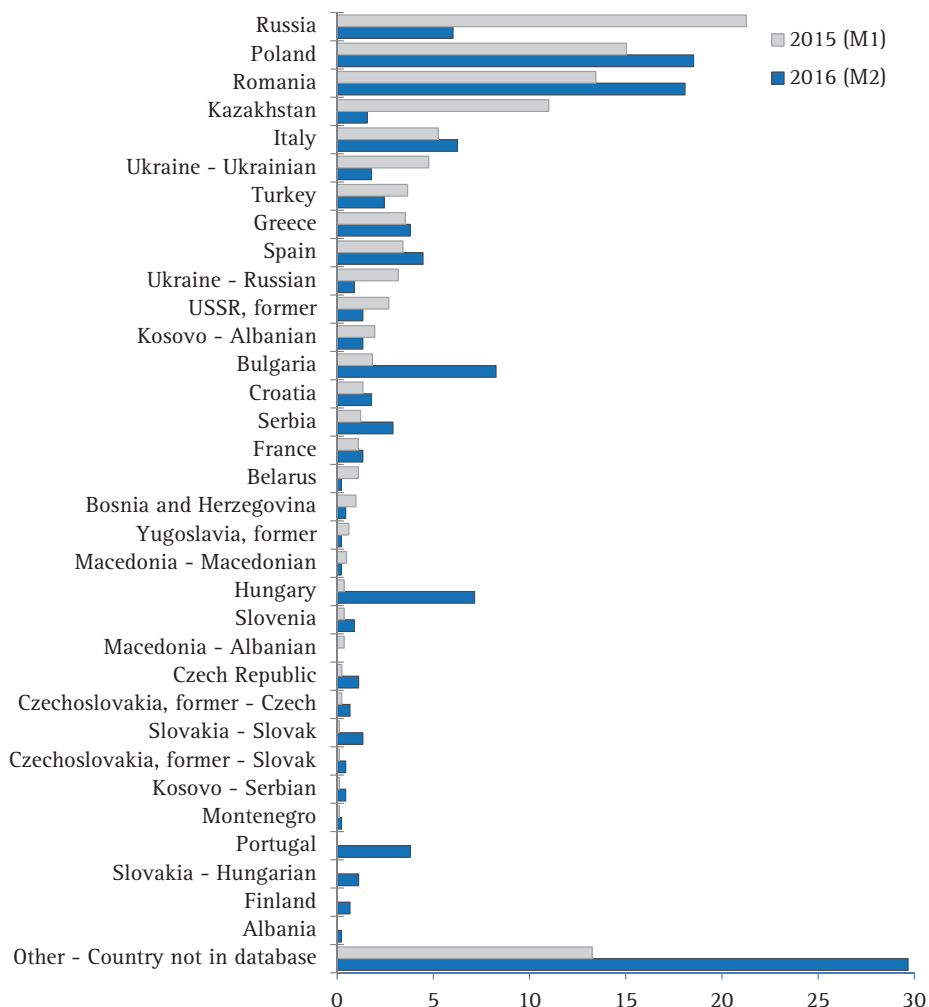
Figure 5 Countries in which respondents obtained their highest educational qualification in the 2015 and 2016 Migration Samples (%)

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.
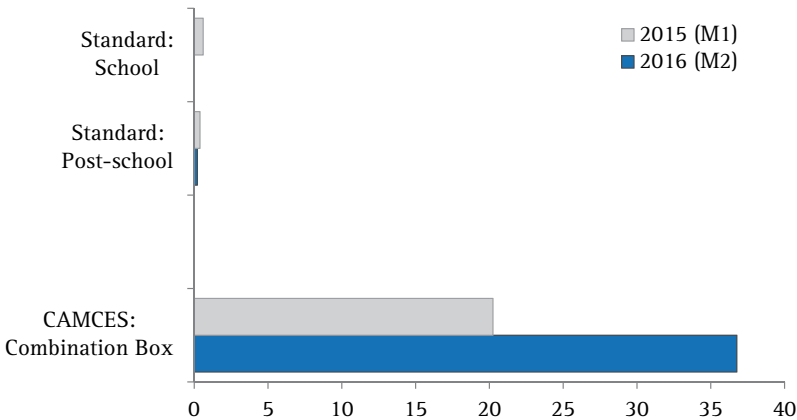

Figure 6   Item nonresponse for the standard measures and the CAMCES combination box in %

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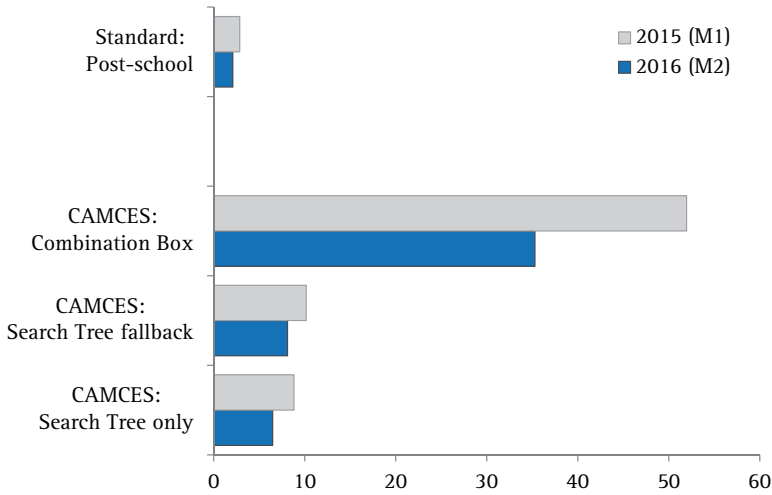
Figure 7   Responses not automatically coded or "other" answers per interface in comparison to the standard item in %

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.

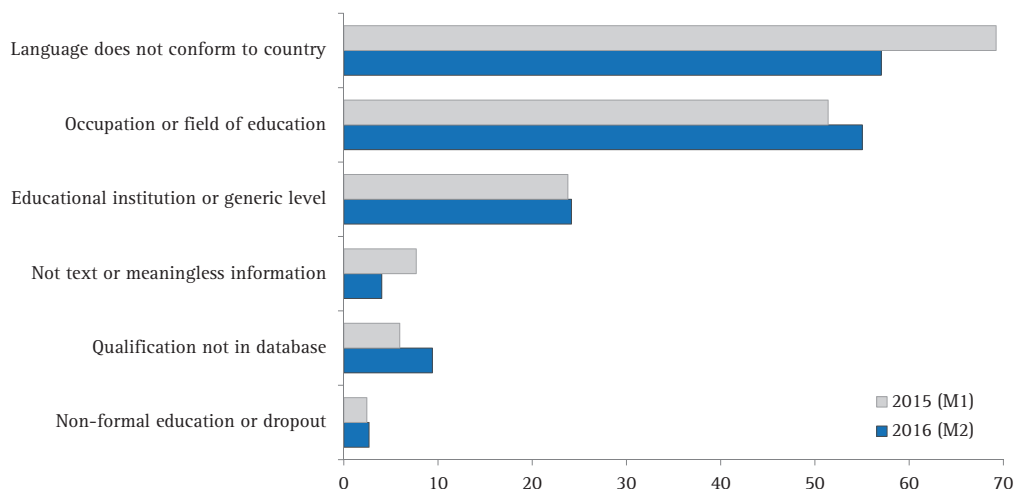*Figure 8*   Reasons for non-coded responses of the combination box in %

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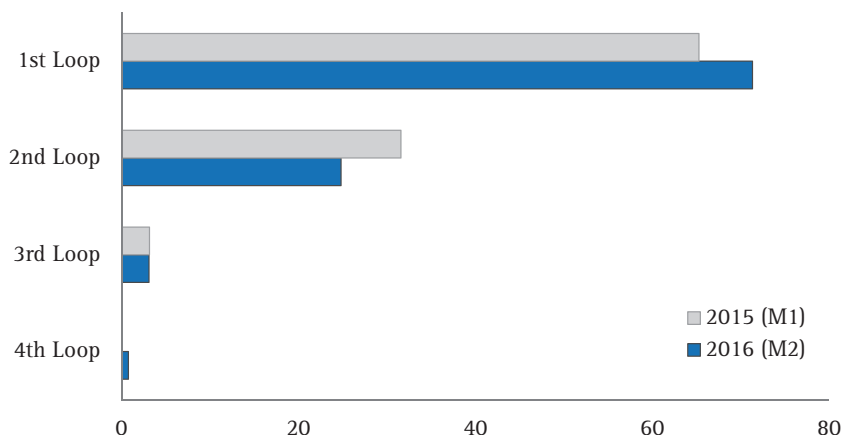
Figure 9   Loop in which the actually highest educational qualification was reported, in %

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			5	0.4	
ISCED 1	20	1.8	10	0.9	1.3
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								
		0-1	2	3 vocational	3 general	4	5	6-7	8	Total
ISCED derived from CAMCES	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 CAM-CES 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 version[22] 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

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. Vesion: 1.0. GESIS - Pretestlabor. Retrieved from http://doi.org/10.17173/pretest25

OECD. (1999). *Classifiying 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: Universtiy of Ljulbjana, 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-pulished 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_ssp0219.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

ISCED categories	IAB–SOEP instrument 2013 / 2014	
	School-leaving certificate (question 72)	Vocational training and higher education (question 83 and 84)
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 General upper secondary education	graduated from higher-level secondary school	no answer / other / in-house training at a company
ISCED 3 vocational Vocational upper secondary 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