Universität Mannheim
Fakultät für Sozialwissenschaften

Issues in Measuring Education in Cross-National and Migration Surveys

Inauguraldissertation zur Erlangung des akademischen Grades einer
Doktorin der Sozialwissenschaften der Universität Mannheim

vorgelegt von
Verena Ortmanns

Dekan der Fakultät für Sozialwissenschaften:	Prof. Dr. Michael Diehl
Erstbetreuerin:	Dr. Silke L. Schneider
Zweitbetreuer und Gutachter:	Prof. Dr. Christof Wolf
Gutachter:	dr. Ruud Luijkx
Vorsitzender der Prüfungskommission:	Prof. Dr. Michael Braun
Tag der Disputation:	27. Juli 2020

Acknowledgements

During my studies, I was interested in the topics of education and survey methods and I did not know what an excellent combination that is. In 2009, Professor Peter Hartmann recommended to me the book by Silke Schneider, in which she and her colleagues compared education systems of 15 different countries and conducted a criterion validation of the ISCED classification. This book changed my life. In my Masters thesis I focused on a related topic. Also during my work in the GESIS Data Archive in Cologne when documenting the EVS 2008 and later the Eurobarometer studies, I was confronted with the ISCED classification. In 2011, Silke started to work at GESIS in Mannheim and I first met her at the GESIS Summer School in 2012. A year later the CAMCES project started and I was so happy to work on this project and also on its extension within the SERISS project. For many years, I worked in depth on the topic of comparing education measures across countries in those projects and also in my thesis, and I enjoyed it so much. Nevertheless, it was a long road to finishing this thesis and I would not have succeeded without many people who have helped, advised, and encouraged me so much along the way.

Firstly, I thank Silke Schneider for her excellent supervision of my thesis and for generally being a great mentor. I am grateful for your encouragement, for your patience, for providing important guidance and for taking so much time for brainstorming, continuous discussions and keeping things together as well as for your constant feedback. I really enjoyed working together and I very much liked our expert talks. I also want to thank my second advisor Christof Wolf, for his support, his helpful comments and advice and for reviewing my thesis. I am also very thankful to Ruud Luijkx who also reviewed my thesis. I would like to thank Michael Braun for his helpful recommendations and for agreeing to chair my dissertation committee. A special thanks to Jürgen Hoffmeyer-Zlotnik for his support and feedback especially in the early years of my dissertation project. I greatly appreciate the time and effort that you all gave me.

This dissertation would have been impossible without support from several former and current colleagues at GESIS, especially of the ‘Survey Design and Methodology’ and ‘Data Archive for the Social Sciences’ departments. I am very thankful for the support I received from Jessica Herzing and Roberto Briceno-Rosas, with whom I worked in the CAMCES project. I enjoyed working, discussing, laughing and sometimes moaning with you and I am grateful for your feedback. I would also like to thank my colleagues in the 'Questionnaire Design and Evaluation' team, namely Natalja Menold, Cornelia Neuert, Ranjit Singh, Patricia Hadler, Cornelia Züll, Timo Lenzner, Angelika Stiegler and Niklas Reisepatt. I appreciate the valuable and inspiring discussions we had and I am grateful for your support. I am also thankful to Katharina Meitinger, Matthias Sand, Evi Scholz, Gabriele Wahlig, Carina Cornesse, Anna-Carolina Haensch, Isabella Minderop, Jessica Daikeler, Britta Gauly, Beatrice Rammstedt, Tobias Gummer, Joss Roßmann, Markus Quandt, Serap Firat, Ivet Solanes-Ros, Petra Brien, Insa Bechert, Irene Müller, Carina Behling and Nilgün Canan. Special thanks to Peter Hartmann, who brought the ISCED classification to my attention, and to Evelyn Brislinger and Meinhard Moschner, who challenged me with checking and documenting of the ISCED classification in the EVS and the Eurobarometer. Thank you so much for guiding me along the way and for removing my doubts about starting this project. I also thank Jane Roberts for proofreading this thesis.

While writing this dissertation, I have had the privilege of meeting many people at workshops, conferences and guest stays, who also provided helpful comments and support. I want to especially thank the colleagues working with me in the SERISS project, the colleagues of the Harmonisation project in Poland and my conference companion Peter Granda.

Finally, I am thankful to my family to whom I dedicate my thesis - for everything that cannot be measured. Thank you!

Content

List of Tables

List of Figures

1 Introduction and Theoretical Framework

1.1 Setting the Scene

Education is a central concept in social science research, and thus a key socio-demographic characteristic that is measured in almost every survey. Many researchers take the education variable for granted (Smith, 1995) and refer to it as “conventional and self-evident” (Braun & Müller, 1997, p. 164). However, the different meanings of the concept of education and its related measurement are often not as explicit and self-explanatory as many researchers might think (Braun & Mohler, 2003; Schneider, 2016).

Due to the importance and wide usage of education in empirical studies, I take a closer look at this concept and the related education variables. In this thesis, I assess the quality of education variables from the perspective of survey organisers producing the data and of researchers using these data. For a survey organiser, it takes some effort to define the aim and the underlying concept of the education variable, to implement a high-quality measurement instrument and to derive the respective variables. However, this investment is very likely to pay off because almost every researcher uses the education variable. Thus, a good education measure is a rational use of resources for the survey organiser and is also likely to enhance the survey’s reputation and legitimacy. This will encourage researchers to make greater use of the survey. From the perspective of researchers, the education variable (like all variables) should be of high quality because they often rely on it heavily in their data analysis. For instance, when the education variable is included in regression models, it is important that the effect of educational attainment is not over or underestimated, and that this variable does not bias the effects of other variables that correlate with education. Only then can researchers generate trustworthy results and draw appropriate conclusions. Moreover, it is important that researchers understand what has been measured with the education variable and for which purposes it can be used. This requires good documentation to be provided by the survey producer.

To assess the quality of the instruments used in surveys for measuring education and of the resulting variables, I consider the following quality criteria: objectivity, reliability and validity (Krebs & Menold, 2014; Rammstedt, 2010). Firstly, objectivity indicates that the information is measured and later also analysed independently of subjective influence. To achieve this, surveys use standardised questions when asking respondents their highest educational qualification, they provide interviewer instructions and conduct fieldwork monitoring ensuring that respondents' answers are not influenced by the interviewer or other external factors (Krebs & Menold, 2014; Rammstedt, 2010). Surveys in most countries rely on self-reports because they do not have adequate register data that could be linked with the survey data. Secondly, reliability refers to the consistency of the measurement. It shows the extent to which the measurement instrument for educational attainment indicates the same qualification when a respondent is questioned repeatedly using the same instrument (Schermelleh-Engel & Werner, 2012). Since surveys usually do not implement repeated education measurements, as second best we can assess aggregated reliability and regard education distributions of different surveys, which all have probability-based samples, as results of repeated measurements of the same population. The distributions of the education variable should be quite similar across surveys for the same country and year when looking at the same age groups. Lastly, validity indicates that the measurement instrument actually measures respondents' educational attainment and not, for example, the subject of the education or the subjective social status. If several education variables are available and we need to choose one, we can conduct a construct validation analysis estimating and comparing the predictive power of the different variables (Hartig, Frey, & Jude, 2012). The three quality criteria relate to each other hierarchically. At the top is validity, which is the goal that good measurement strives for. Objectivity and reliability meanwhile are necessary but not sufficient preconditions of validity. The extent of objectivity and reliability determines the maximum possible validity of a measurement instrument. However, even if an instrument is objective and reliable, it may still not be valid (Hartig et al., 2012; Rammstedt, 2010).

I also consider the quality criterion of comparability, because this is central to analysing data of cross-national and migration surveys, as is done in this thesis. Comparability indicates that the instruments measuring education, and their design and implementation, are comparable across countries, regions and cultures and/ or over time as well as across surveys (Harkness et al., 2010; Harkness, van de Vijver, & Mohler, 2003; Przeworski & Teune, 1970). To achieve a comparable measure of education across countries, first of all the concept being measured should be identical, namely educational attainment, and defined and understood in the same way. However, for measuring this, the instruments need to be country-specific in order to capture the differences in the education systems, their idiosyncratic institutions, and their related qualifications that cannot be translated, in the answer categories of the education question (Braun & Mohler, 2003; Schneider, Joye, & Wolf, 2016; Smith, 1995). To generate a comparable variable, the categories of the country-specific qualifications are assigned to a standard classification, such as the International Standard Classification of Education (ISCED) (UNESCO-UIS, 2006, 2012). In this context, we have to be aware that comparability strongly relates to reliability, meaning that if we do not find the education measure to be reliable we also cannot compare it, for example across surveys. Overall, the practical application in surveys of the four quality criteria, among others, is specified in guidelines, for instance, of national initiatives, such as the German Data Forum (RatSWD - Quality Standard Working Group, 2015), or of international stakeholders like Eurostat (2018).

Previous studies looking closely at the harmonised ISCED variable across surveys for countries and years have revealed inconsistencies in the education distributions (Kieffer, 2010; Schneider, 2009). However, the full extent of the problem is unclear because most of these studies captured only a small number of surveys and countries. These studies tried to identify the error sources by using qualitative investigations, and identified various errors in the measurement of educational attainment; however, this has not been done in a systematic, quantitative fashion. In this thesis, firstly I contribute to this research by analysing more recent data as well as a larger number of countries, namely 31, from ten cross-national surveys. For these, I assess the aggregated reliability and thus the implied comparability of the harmonised ISCED variable. Also, for the first time, I examine quantitatively survey characteristics as potential causes for these inconsistencies. This allows me to quantify the extent of such inconsistencies and to identify the error sources in cross-national surveys more thoroughly than previous research has done. This knowledge will be useful for survey organisers, who can reduce these errors in a targeted way and thereby enhance the aggregated reliability and comparability of their education measures in the future.

Secondly, I consider the perspective of data users when exploring a timely use case study in the context of migration research. In this study, I analyse another kind of cross-cultural data, namely of a national migration survey. This faces similar challenges to those in cross-national surveys when measuring education achieved by immigrants in their countries of origin. Migration surveys often contain different education variables derived from different measurement instruments. Therefore, it is important to identify the most valid education variable to be used in substantive analyses. To assess this, I estimate a construct validation of different education variables. With this analysis, I aim to increase researchers' awareness of different education measures in migration research and the related migration surveys. Previous studies, for instance of Kerckhoff, Ezell and Brown (2002), Kerckhoff and Dylan (1999), and in particular of Braun and Müller (1997) have conducted similar validation analyses. They all observed differences in the predictive power of the education variables. These studies did not focus on migration surveys, and this is another contribution added by this thesis. Moreover, with the use case study focusing on substantive research questions, I want to give an example of how the results of the construct validation can be useful for substantive research.

This introductory chapter aims to illustrate the importance of the concept of educational attainment in social science research and its measurement in surveys. In the next section, I introduce different concepts of education. Then in section three, I expose the relevance of the concept of educational attainment in social science research by looking at different theories, models and approaches of social stratification research in which education is the essential element. I will look at further research strands, in which the education variable is widely used but for different purposes. Related to these theories and models, I present different education variables and measurement instruments in section four. In section five, I summarise the four papers of my thesis, and in section six I discuss the main results, draw conclusions and develop ideas for enhancing the quality of measurement of educational attainment in cross-national and cross-cultural surveys. The four papers of this dissertation follow this introductory chapter.

1.2 The Concept of Education in Social Science Research

Education can be defined in various ways; this applies especially to the German term 'Bildung' that has a broader connotation than the English one. Therefore, I will now introduce the main concepts of education. The broadest and most comprehensive concept follows the ideal of Humboldt and regards education as a multidimensional and inclusive concept of acquiring knowledge and skills. It covers formal and informal education, courses such as adult education courses and specified training courses, cognitive and non-cognitive skills, and personal characteristics such as self-responsibility and conscientiousness (Raithel, Dollinger, & Hörmann, 2009; Schaub &

Zenke, 2007). This overarching concept is hard to measure in empirical studies and therefore I distinguish between two more explicit concepts that indicate the outcomes of learning, namely competence and educational attainment.

The first concept, cognitive competence, indicates that individuals successfully learn and acquire knowledge, cognitive skills and abilities and thus develop expertise (Raithel et al., 2009). They acquire competencies through ‘learning by doing’ when they are confronted with themselves, their environment, cultures and the social values of society. Cognitive competencies are acquired through formal education (e.g. in schools or universities), non-formal education (e.g. on the job training or adult education courses), and through informal learning. The latter refers to learning that is motivated by personal interests (intentional), and it can also occur incidentally in a family or job-related context (unintentional) (Kleinert & Matthes, 2009; Rüber & Bol, 2017). Thus, formal institutions are not solely responsible for transferring skills and competencies (Schaub & Zenke, 2007). This concept also emphasises the idea of lifelong learning (Raithel et al., 2009; Schaub & Zenke, 2007). It also corresponds to Bourdieu’s description of 'incorporated cultural capital' (Bourdieu, 1983). This covers skills and competencies, such as linguistic skills, awareness of cultural peculiarities as well as specific knowledge, attitudes and behaviours that are important for being successful in the education system, the labour market and life. These competencies are often absorbed through individuals’ socialisation with their family and friends.

The second concept focuses on formal educational attainment. Students receive systematic teaching in institutions to acquire knowledge and skills, which are reflected in the educational qualifications they have achieved (LEXICO, 2020; Raithel et al., 2009). These are important because they allow students to advance to the next educational level and proceed through their educational career. In contrast to the first concept, formal education highlights the importance of educational institutions and formalisation. Moreover, this concept considers the utility of educational qualifications for indicating individuals’ efficiency and performance for the labour market (Raithel et al., 2009). This concept is reflected by Bourdieu (1983) in the term 'institutionalised cultural capital'. Students’ competencies and skills in their incorporated cultural capital become objectified and institutionalised through the awarding of qualifications. Thus, the education system and its structure and institutions strongly determine students’ educational attainment.

Most studies in the social sciences refer to the concept of educational attainment[1] because of the importance of formal education in an individual's life course. Educational attainment strongly determines whether individuals have a chance of achieving a higher occupational position and thus a higher socio-economic status. From a macro perspective, the chances of achieving high educational attainment are often not equally distributed in society, so educational attainment is a central factor in social stratification research. Moreover, studies favour the concept of educational attainment over the concept of competencies, because educational attainment has higher objectivity and it is demonstrable for employers as well as in surveys. For measuring educational attainment in a survey, the interviewer can directly question respondents on their highest educational qualification. On the other hand, to measure cognitive competencies adequately, surveys need to conduct extensive tests, commonly on respondents' literacy and numeracy skills. Such tests considerably increase the costs and the duration of surveys, and also increase the response burden for participants. Apart from specialised surveys such as the German National Educational Panel Study (NEPS) or international studies like Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), Programme for International Student Assessment (PISA) or Programme for the International Assessment of Adult Competencies (PIAAC), this is rarely done in social science surveys.

1.3 The Relevance of Education in Social Science Research

In this section, I demonstrate the relevance of the concept of educational attainment in the social sciences. Firstly, I introduce major theories and models of social stratification research in which education in the formal sense (see section 1.2) is central. Secondly, I present further research in which educational attainment is also central but for different purposes. I refer to the usage of the education variable as a proxy for another concept, such as competencies or social status, or when implemented as a background or control variable, or used for designing survey weights.

[1] Concerning the terminology, I use educational attainment and education interchangeably in this thesis to enhance readability.

1.3.1 Education in Social Stratification Research

Social stratification researchers analyse the social structures within society. A central model of this research strand is the OED triangle (origin-education-destination) developed by Erikson and Goldthorpe (1992), which indicates the paths linking individuals' social origin with their current socio-economic position. This relationship (path C in Figure 1.1) is strongly mediated by individuals' education (paths A and B). To properly estimate the social mobility regardless of education, a good measure of educational attainment is required. Relationship A shows the connection between the social class of origin and education, indicating educational inequality, and relationship B refers to the link between education and destination class, known as the returns to education. To determine the effect of education in relationships A and B we also need a good measure. In this section, I look in more detail at relationships A and B because these directly involve education.

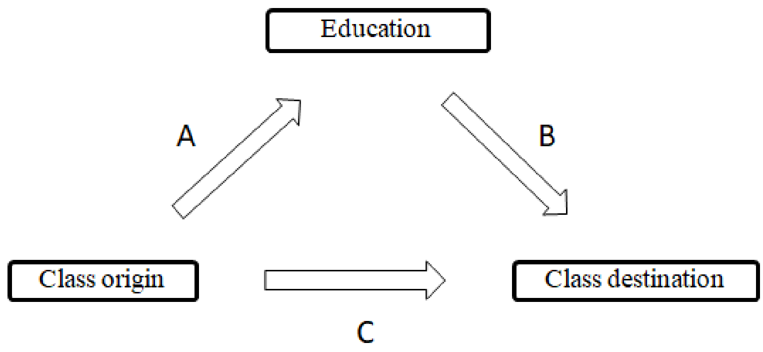


Figure 1.1 The OED triangle indicating the relationships of origin, education and destination, related to Erikson and Goldthorpe (1992)

Education as the Dependent Variable: Analysing Educational Inequalities

Since the mid 20^{th} century, numerous studies have analysed the impact of social class of origin (and later also gender, migration background or other 'ascribed' characteristics) on individuals' educational attainment. Most studies focus on industrial societies that have experienced large social and economic changes in the 20^{th} century, primarily due to industrialisation and major advances in technology (Blau & Duncan, 1967; Blossfeld & Shavit, 1993; Breen & Jonsson, 2000; Mare, 1981). Industrialisation changed the structure of the social classes by increasing the proportion of people who are in paid employment. For achieving a higher occupational position a formal educational qualification was required. Thus, the educational aspirations of the population and the participation rates in educational programmes increased, as well as political decisions leading to the educational expansion in the 1950s. Individuals' education strongly determines individuals' social status and thus societies have become more meritocratic. Although the educational levels in all social strata increased through this educational expansion, many studies observe the so-called education paradox. This describes how individuals of a higher social class still acquire a higher level of education than those from a lower class. Thus, in spite of this expansion educational inequality between different strata was preserved; according to some studies, it was even reinforced, which was surprising (Bell, 1999; Blossfeld & Shavit, 1993; Boudon, 1974). To explain this paradox, researchers developed different theories and models. I will now describe the main ones.

One of the first studies analysing social mobility (path A in Figure 1.1) between two generations is by Blau and Duncan (1967). They use path-analytical stratification models capturing the educational level and the occupational position of fathers and sons. Educational attainment was implemented as a dependent and an independent variable, mediating the relationship between fathers' education and occupation and sons' occupation. For measuring educational attainment, Blau and Duncan use years spent in school. Thus, education is regarded as a linear accumulation of knowledge and skills. This view of education is also considered in human capital theory (Becker, 1993; Mincer, 1974). However, to better explain the relationship between social background and education, Sewell, Haller and Portes (1969) extend the Blau-Duncan model and add further mediating socio-psychological factors, such as individuals' educational and occupational aspirations. In the following years, simple multiple linear regression models using the years of schooling variable were estimated when analysing educational inequality (Hauser & Featherman, 1976).

In the 1970s criticism of the linear models grew. One point of criticism was that these models did not mirror the cumulative extension of education through educational decisions. Moreover, these models do not reflect the family's effects on educational decisions. Boudon (1974) covers this and distinguishes primary and secondary effects. Primary effects describe parents' impact on children's educational achievement, largely because of economic and socio-cultural resources as well as genetic factors. Secondary effects indicate parents' impact on students' educational decisions, and thus students' educational transitions controlling for their prior achievement. Secondary effects are a result of families' economic and socio-cultural resources that influences the costs of achieving a higher educational level as well as its expected usefulness (Boudon, 1974). Most empirical studies focus on analysing the impact of secondary effects when explaining educational inequality (Shavit & Blossfeld, 1993 Blau & Duncan, 1967)

Another point of criticism of the linear models is that they do not adequately consider the effects of the educational expansion. The linear models cannot distinguish changes in educational inequality that are structurally determined (by changing marginal distributions due to expansion) from 'net' changes. Mare (1980, 1981) better reflects this in his model that analyses students' educational decisions and transitions as a function of their social background. The 'Mare model' divides students' educational careers into a set of sequential educational decisions. For each transition, the continuation probability is calculated through binary logistic regressions and the resulting odds ratios indicate the chances of remaining in the education system. The education variable for this model focuses on students' educational transitions. This information is derived from the years of schooling variable when analysing the American education system, which is rather linear. In contrast, for the education systems of European countries, which are not linear and offer parallel educational tracks at the same level, the information is derived from the variable of the highest educational qualification. The 'Mare model' has been widely used in empirical studies (Cobalti, 1990; Hout, 1989; Müller & Karle, 1993; Shavit & Blossfeld, 1993) and for trying to disentangle empirically the primary and secondary effects of social background (Jackson, Erikson, Goldthorpe, & Yaish, 2007; Nash, 2006; Neugebauer, 2010).

Since the mid-1990s, the 'Mare model' was extended and more models were developed explaining educational inequality in the framework of rational choice theory (Breen & Goldthorpe, 1997; Erikson & Jonsson, 1996; Gambetta, 1987; Hillmert & Jacob, 2005). Breen and Goldthorpe (1997), for instance, developed a model that to a large extent considers direct and indirect costs of education, and thus reflects individuals' risk aversion for failing or successfully completing an educational programme. Risk aversion is relative to individuals' social origin and also depends on the motivation to maintain families' social status. Thus, educational aspirations and the utility of education are higher for children from a family having higher social status and who are often ambitious to maintain this status and avoid social decline for themselves.

In 2000, Breen and Jonsson extended the 'Mare model' and developed a multinomial transition model. This enables estimating educational inequality more accurately by assessing students' pathways through the education system (Breen & Jonsson, 2000). In contrast to the 'Mare model', the multinomial transition model reflects the structure of the education system, such as the level of tracking or the distinguishing of general or vocational programmes, which are especially important in non-linear education systems. The information on the single transition is again derived from the variable on the highest educational qualification.

At the same time, Lucas (2001) in his hypothesis on effectively maintained inequalities (EMI) combines theoretical and empirical perspectives of research related to stratification in education systems and educational transitions. He assumes that students are part of stratified programmes before and after each transition. This stratification and the related educational inequality can be vertical (relating to educational level) or horizontal (relating to quality at the same level, e.g. through different institutes), depending on the context and the structure of the education system. Lucas applies an ordered probit model, and for the education measure he uses information on the years spent in the education system as well as a small set of variables on single courses and the achieved levels in these. Thereby he better reflects additional aspects of education, such as the type of qualification, the related field of study, as well as the institution attended. All these are important characteristics of the education system that should be considered in the context of educational inequality (Lucas, 2001).

All these studies are engaged in the development and extension of theory and statistical modelling of the relationship between social origin and educational attainment. Marks (2011, 2014) extends this relationship by introducing cognitive ability and competencies as mediating factors. Analysing educational inequality, Marks shows that by considering individuals' cognitive ability the effect of social background decreases. Accordingly "the association between socioeconomic background and ability is much weaker than that between parent's and child's abilities" (Marks, 2011, p. 58). Thus, Marks focuses on measuring competencies and therefore uses IQ scores or PISA test scores (Marks, 2011).

Although these theories and models differ widely and focus on different aspects when explaining educational inequality, educational attainment is the dependent variable in these studies and thus the main focus. As seen, most of these studies focus their analysis on a single point, often a central educational transition. However, we have to bear in mind that a transition is only an extract of an individuals' educational career, and not represent the complete pathway within the education system. For measurement, most studies, especially those in European countries, use the variable indicating individuals' highest completed educational level, and from this they derive the information they are interested in, for example on a specific transition. Thus, the measurement quality of the educational attainment variable needs to be high for it to be used successfully in deriving the information required for these models.

Education as the Main Independent Variable: Analysing Returns to Education

A good education measure is also important when assessing returns to education (path B in Figure 1.1), which indicates the effect of individuals' educational attainment on their social status or related socio-economic outcomes, such as earnings. This relationship is often analysed from an economic point of view, for example in human capital theory (Becker, 1993; Mincer, 1974), where more highly educated individuals are regarded as more productive than less educated individuals and consequently receive higher wages. Therefore, individuals will invest in their education (Becker, 1993; Mincer, 1974). In this theory, educational attainment is typically measured by years spent in the education system to assess the effect/ the return of each year.

Building on human capital theory, Thurow (1975) highlights the importance of educational attainment in the job application process. From the perspective of applicants, investing in education and achieving a high qualification is advantageous because this signals their productivity to the employer. From the perspective of the employer, applicants' educational attainment is helpful as a screening device to screen out and/ or rank them and to assess how well they match with the job (Spence, 1973; Thurow, 1975). In the context of the application process, researchers also refer to educational attainment as a positional good (Hirsch, 1976; Ultee, 1980) because they estimate "the value of educational credentials as attributable, in part, to their relative scarcity in the population" (Shavit & Park, 2016, p. 1). Consequently, the value of educational qualifications and their utility, e.g. for the labour market, differs across cohorts and time.

In the context of educational expansion, qualifications are devaluated because an increasing number of people have such credentials. Investment in education may not be fully rewarded and some people may, for instance, be overqualified for their job (Burris, 1983; Clogg & Shockey, 1984). This development is critically reflected in the credentialist theory (Brown, 1995; Collins, 1971, 1979; Kerckhoff, 1976). Employers' decisions when hiring staff are no longer based only on applicants' educational qualifications. Instead, the association with a specific culture or social elite expressed by a certain qualification and shared societal assumptions on the link between education and occupation become more important (Brown, 1995; Collins, 1971). Thus, applicants' social background becomes more relevant. This is another research strand that requires a relative measure of educational attainment that considers the distribution of each qualification in society for certain points in time.

Moreover, returns to education also depend on the structure of the national education system and the labour market. Both differ greatly across countries and this needs to be considered in comparative research on this topic. Allmendinger (1989) and more recently Bol and colleagues (Bol & van de Werfhorst, 2013, 2016; Di Stasio, Bol, & van de Werfhorst, 2016) emphasise the different structures of education systems, looking at the level of standardisation, tracking and vocational orientation. Concerning the labour market, Marsden (Marsden, 1990; Marsden & Germe, 1991) distinguishes 'occupational' labour markets (OLM), which strongly rely on formal educational qualifications and 'internal' labour markets (ILM) in which formal qualifications are less important. Further context factors, such as the national economic situation, unemployment rates and national regulations for employment protection should also be considered (Bol & van de Werfhorst, 2013; Gangl, 2002, 2003). A large study analysing both the effects of structures of the education system and the labour market was conducted by Shavit and Müller (1998), who analysed the school-to-work transition across 13 countries. To consider adequately the effects of the education system and the labour market and to avoid over or underestimation of these context effects, education needs to be measured accurately.

In all these studies that analyse the returns to education, educational attainment is the key independent variable and should be of high quality. In this research strand, the years of schooling variable or a categorical education variable are quite commonly used. For a relative measure, we also need to consider the distribution of the qualifications to the respective point in time. The comparative studies require an education measure that is internationally comparable but also considers the positions of the qualifications within the different national education systems. Overall, in social stratification research, education is a core variable of theoretical and empirical importance, and the variable and its related measurement instrument should fulfil the quality criteria.

1.3.2 Education Beyond Social Stratification Research

As we have seen, in social stratification research educational attainment is the central concept of interest and the related education variable is either implemented as the dependent variable when analysing educational inequality or as independent variable of primary interest when predicting returns to education. Beyond social stratification research, educational attainment is also an important concept but we observe different relationships between the concept and the indicator. In one case, researchers use the education variable as a proxy for another concept, such as competencies or social status, in which they are interested. Other studies implement the education variable as a background or control variable, but education is not the main focus of their interest. The education variable is also frequently considered when designing survey weights. In this section, I will discuss the usage of the education variable in these contexts and thereby underline its importance.

Education as a Proxy Variable

Researchers who are interested in concepts related to education, such as competencies, social status, cultural capital or socialisation, frequently use the individual's educational attainment variable as a proxy in their analysis. They will apply this workaround because a measure of their concept of interest is not available. For instance, epidemiological and health studies estimating the risk of disease for people of different social strata use the education variable as a proxy for social status (Liberatos, Link, & Kelsey, 1988; Link, Northridge, Phelan, & Ganz, 1998; Miech & Hauser, 2001; Ross & Wu, 1995). In social science studies, many researchers use this variable as a proxy for competencies, when analysing the effect of education on social and political attitudes or behaviours (Bekhuis, Lubbers, & Verkuyten, 2014; Hyman & Wright, 1979; Weakliem, 2002). This also applies to migration research examining the effect of education on immigrants' second language proficiency (Chiswick & Miller, 2001; Espenshade & Fu, 1997; Esser, 2006; van Tubergen, 2010) or on their earnings (Chiswick & Miller, 1995; Dustmann, 1994; Dustmann & van Soest, 2002; Friedberg, 2000; Weins, 2010).

This approach of using the variable as a proxy for another concept has attracted criticism. When using a proxy we have to be careful with the interpretation of its effects. Therefore, Kingston and colleagues (2003) recommend not using the education variable as proxy for social status, although social status explains a considerable part of the educational effect. Using educational attainment as a proxy for cognitive skills is also not ideal because attainment includes the effects of school and the related socialisation (Kingston et al., 2003). Also, educational attainment does not properly cover competencies of students who drop out of the education system, but who nevertheless have acquired competencies that are useful, for example for finding a job (Hübler, 1984). This also applies for competencies acquired outside the formal education system, in particular vocational skills gained from an internship or trainee programme (Schneider, 2016). Massing and Schneider (2017) analysed the relationship between educational qualifications and competencies, looking at PIAAC data of 21 countries, and found that educational attainment on average explains only 26% of individuals' literacy skills. Thus, educational attainment is not an adequate proxy for competencies, and we underestimate the effect of competencies when using educational attainment as proxy.

Concerning operationalisation, measuring respondents' social status, social background or competencies require greater effort than 'just' measuring educational attainment. Alongside educational attainment, we need to measure respondents' occupation and income for measuring their social status. For measuring social background we also need to consider parents' occupation. To measure competencies and skills directly, surveys need to include comprehensive and detailed assessments. Asking a batch of additional questions or conducting a competence test in a survey is often not possible due to restrictions in time and budget, and so researchers have to use the education variable, which is included in virtually every survey. When using the education variable as a proxy, it is important that the variance of the actual concept is reflected as well as possible in the education variable. Otherwise, the education variable is not a good enough proxy for that concept.

Education as a Background or Control Variable

When using educational attainment as a background or control variable, researchers do not establish hypotheses involving education and the variable is included 'only' as a control. This is done to identify whether the inclusion of educational attainment in the model changes the relationship between the dependent variable and the main independent variables of interest, especially if these variables correlate with education. Educational attainment is, for instance, considered as a control when analysing social and political attitudes, such as voting behaviours (Almond & Verba, 1963; Bekhuis et al., 2014; Weakliem, 2002), attitudes towards minorities and immigrants (Coenders & Scheepers, 2003; Hyman & Wright, 1979; Semyonov, Raijman, & Gorodzeisky, 2008), gender role attitudes (Bolzendahl & Myers, 2004; Harris & Fireston, 1998), media consumption and information retrieval (Pardos-Prado & Cano, 2012), as well as health status and health literacy (Glanz, Rimer, & Viswanath, 2008; Ross & Wu, 1995). In most studies, controlling for education (as well as for gender and age) is taken for granted because education correlates with many other variables, and therefore, researchers often do not explicate the necessity of controlling for education.

Although educational attainment is 'just' implemented as a control variable, its quality and the applied measurement criteria should be the same as in studies where educational attainment is the main variable of interest or regarded as a proxy. A poorly conceptualised and measured education variable increases the risk that "the effects of other variables will include the effects of unmeasured differences in education" (Schneider, 2016, p. 3).

Using the Education Variable for Designing Survey Weights

Together with the variables of sex and age, education is central to analysing the composition of a sample. Studies indicate that many surveys, especially web surveys, face an education bias: the better educated are over-represented whereas the less well educated are under-represented, so that the sample is not representative of the general population (Abraham, Maitland, & Bianchi, 2006; Billiet, Philippens, Fitzgerald, & Stoop, 2007; Dillman et al., 2009; Groves & Couper, 1998; Lugtig, Lensvelt-Mulders, Frerichs, & Greven, 2011). In general, the less well educated more often refuse to participate, due to lack of interest or the sensitivity of the topic of the survey (Groves, Presser, & Dipko, 2004; Rogelberg & Luong, 1998) or because they fear they might not respond 'properly' to the questions. The latter, in particular, applies for education studies, such as PISA or PIAAC, which conduct performance tests (Helmschrott & Martin, 2014).

To correct the education bias, data organisers design post-stratification weights that adjust nonresponse differences across educational groups. In a first step, they compare the distribution of the education variable of their survey to the distribution in another data source, the benchmark. This would ideally be data from official statistics. Thereby survey organisers can judge how similar or different their education distribution is. In a second step, they adapt the survey data to that of the benchmark by calculating weights (Koch, Halbherr, Stoop, & Kappelhof, 2014; Peytcheva & Groves, 2009). Thus, a high quality education measure is essential for comparing data across surveys and designing weights. In this context, categorical education measures are usually preferred due to their flexibility in generating comparability across surveys.

Overall, this section illustrates the widespread usage of the education variable in substantive research and for methodological purposes. In social stratification research educational attainment is the key concept. Beyond this research strand, the education variable has often been used as a proxy for another concept, or as a background or control variable in studies on other topics, or when designing survey weights. Thus, the education variable is clearly important, and its related measurement instrument should therefore fulfil the quality criteria of objectivity, reliability, validity and comparability.

1.4 Measuring Education

In this section, I will look at the different education variables that are commonly used in survey research and have been mentioned in the previous sections. These are the years spent in the education system, education scores, and those referring to educational standard classifications. These variables, in turn, are based on the survey instruments measuring respondents' education. Therefore, I will next introduce those instruments and discuss the impact a national, cross-national and migration survey might have on the instruments. Moreover, I will assess the quality of the different variables and instruments, considering their reliability, validity and comparability. Regarding objectivity, survey organisers try hard to ensure it, through standardisation of interview situations and question text, but a certain degree of subjectivity can never be ruled out when conducting a survey. I assume that objectivity is of an equally standard for all measurement instruments.

Before introducing the different education variables in detail, I want to emphasise that measuring educational attainment is challenging, particularly in cross-national and cross-cultural surveys like migration surveys. Education systems differ greatly across countries and each system has its own structure, its idiosyncratic institutions, and awards qualifications that cannot be translated (Braun & Mohler, 2003; Schneider et al., 2016; Smith, 1995). Thus, for properly measuring respondents' highest educational qualification, survey organisers need to apply country-specific instruments. These instruments differ especially with regard to the answer categories that indicate specific educational qualifications or a certain level of education.

These country-specific measurement instruments cannot be compared across countries. To derive a comparable education variable from the country-specific instruments, ex-ante output or ex-post harmonisation is commonly applied. Following an ex-ante output harmonisation approach, the survey organisers and country teams agree *before* data collection on the concept, usually educational attainment, the permitted indicator(s) and on the strategy with which they will generate a comparable variable and often also on the educational classification. With this in mind, the country teams develop country-specific instruments. When harmonising the data using an educational classification, these often ask respondents their highest educational attainment, and respondents will indicate their highest qualification from a country-specific list. When developing these categories the country teams have to ensure that these can be assigned to the broader educational classification (Ehling, 2003; Granda, Wolf, & Hadorn, 2010). Most cross-national surveys, such as the European Union Labour Force Survey (EU-LFS), the European Social Survey (ESS), the European Values Study (EVS) or the International Social Survey Programme (ISSP) apply such ex-ante output harmonisation. If this is not possible, survey organisers and more often data users apply ex-post harmonisation. In these cases, each country implements its own educational measure without agreeing on a common standard beforehand and the data are harmonised *after* data collection (Ehling, 2003). Identifying the greatest common denominators to achieve a comparable education variable can be quite demanding. This approach is frequently used to compare data of different survey rounds not using the same measures, or over time or across surveys.

1.4.1 Education Variables

Now, we will look at the education variables most commonly used in survey research, namely years spent in education, education scores and educational classification.

Years of Education

The variable indicating the years spent in the education system focuses on the quantitative aspect of education. Due to the metric level of measurement, this variable is typically used by researchers who want to identify the effect or the returns of each year that individuals have invested in their education. Thus, this variable is often considered in empirical studies related to the context of human capital theory and returns to education (see section 1.3.1).

This variable refers to the question of how many years respondents have spent in the education system. Questions asking respondents the number of years they spent in school or the age when they completed education all follow a similar purpose (Hoffmeyer-Zlotnik & Warner, 2007; Schneider, 2016). Such questions texts in principle can be translated and therefore implemented in national, cross-national or migration surveys (Braun & Müller, 1997; Schneider, 2016). However, the question needs to be worded very carefully if it is to be understood similarly across countries. The term 'school' is often used in the question but it refers to various elements of the education system, and in some languages 'school' in a narrow sense can also include university. It should also be clear which years will count and which not, especially regarding early childhood education, repeated grades or doctoral research, and how to deal with dropouts, in order to improve comparability. To better capture these, additional instructions are needed; but specifying all exceptions increases the number of instructions, and they become more unwieldy.

The main shortcoming of this instrument is that indicting the number of years spent in education can be challenging for the respondents. They have to remember the different educational episodes in their life and calculate the length of the educational programmes they attended, which is likely to be prone to errors, thus reducing the reliability of the measure (Braun & Müller, 1997; Schneider, 2016). Furthermore, we have to keep in mind that the same number of years of education has a different meaning across countries due to differences in the education systems (Braun & Müller, 1997). The question on the age when full-time education was completed has an additional disadvantage; people may have interrupted their education for a spell in work, and thus graduated at a higher age. The frequency of such interruptions also depends on the national education systems and labour markets, in particular on the prevailing requirements for career development and the tuition fees of college and university programmes. Overall, the vagueness of this instrument as well as the measurement errors it causes, reduce its reliability and validity and also its comparability (Helberger, 1988; Schneider, 2010, 2016).

To avoid this calculation task for respondents and to enhance the reliability and validity of this variable, researchers can derive so-called ‘hypothetical years of education’ from respondents’ highest educational qualification (Helberger, 1988). This variable uses the metric scale but it is more precise and less burdensome for respondents. We have to bear in mind that this variable depends upon the quality of the country-specific instrument measuring respondents’ educational attainment in categories. For example, if the categories are broad and incorporate qualifications with varying durations, the derived variable will not accurately represent the years required to obtain the qualifications in that category.

However, both kinds of years of education variables do not consider institutional differences. The differences of parallel tracks, which are on the same educational level and have the same duration, are not reflected in the total years of education. This is not problematic for countries having a linear system, like the U.S., but many European countries have more complex education systems (Braun & Müller, 1997; Hoffmeyer-Zlotnik & Warner, 2014).

Scoring of Education

Another kind of education variable is education scores, which for instance, are often used when referring to education as a positional good (see section 1.3.1). Such research requires a relative measure of education to consider the position of the educational qualification, and of the individuals having this qualification, within the education distribution of a certain time and cohort (Shavit & Park, 2016). Usually, this variable of education scores has a metric level of measurement, like the years of education variable. The scores are often derived from the variable indicating years spent in education or the highest educational qualification and its respective distribution. For generating education scores, different approaches exist, here we distinguish between univariate or bivariate/ multivariate techniques.

In univariate scoring approaches, each educational qualification or year spent in education is considered in its relative position. The easiest approach is to rank respondents according to their educational qualifications. Related to this idea, Bukodi and Goldthorpe (2016) generated an ordinal variable with five categories by changing the definitions of the categories in a more detailed variable and aggregating those categories. This approach allows them to assess changes in the education distribution and in the resulting value of qualifications across time and cohorts. More complex methods have converted the measures of respondents' highest qualification or the years they spent in education into a proportional score or a percentile position (Bol, 2015; Wolbers, de Graaf, & Ultee, 2001). This applies, for instance, to the Positional Status Index (PSI) (Tam, 2016) that indicates individuals' relative position in the education distribution and its distance to the next level, based on the average number of competitors the individual has to beat for reaching that level. Referring to the PSI, Triventi and colleagues (2016) developed the Educational Competitive Advantage Scores (ECAS) that are assigned to each educational qualification and indicate its competitive advantage.

Bivariate or multivariate scoring approaches consider further criteria, in addition to the education measure, such as occupational or social status or income (Shavit & Park, 2016). A relatively simple approach is the Hoffmeyer-Zlotnik/ Warner matrix of education that combines two different dimensions of education – general school education and vocational education (including tertiary education) (Hoffmeyer-Zlotnik & Warner, 2007, 2014). This matrix does not use a criterion variable; instead ten ordinal ranks are manually derived related to the major groups of the International Standard Classification of Occupation (ISCO) (International Labour Office, 2012), without conducting further statistical analysis. Usually, education scores are calculated through regression analysis or log-linear modelling that maximise the correlation of the education measure to other criteria. The resulting scores "can be used to judge the "value" of the original education categories with respect to the criterion" (Braun & Müller, 1997, p. 172). Examples of these approaches can be found in Treimann and Terrell (1975), who provide education scores based on individuals' occupation, or the adapted measures used by Fujihara and Ishida (2016) or Triventi and colleagues (2016).

Rotman and colleagues (2016) use the income distribution as a criterion and the final education scores indicate the mean income for each educational level. These approaches refer to outcome criteria co-determined by education. In contrast, Smith and Garnier (1986) use an input criterion that influences individuals' education, namely fathers' occupation. Based on these approaches, Schröder and Ganzeboom (2014) developed the International Standard Level of Education (ISLED) that combines input (parents' education and occupation) and output criteria (individuals' occupation and partners' education) for calculating the scores of each country-specific educational category. In their approach, Schröder and Ganzeboom use the categorical education variable, the years of education variable for calibration.

Inspired by these scoring techniques, especially by the approach of Schröder and Ganzeboom (2014), I developed an index for measuring immigrants' homeland education in paper IV of this thesis. This index combines a categorical variable on immigrants' highest educational qualification and the years they spent in the education system abroad. Through conducting a non-linear principal component analysis (PCA), I can score variables that have different levels of measurement (Linting, Meulman, Groenen, & van der Kooij, 2007; Meulman, van der Kooij, & Heiser, 2004). This index allows me to generate a comprehensive measure. Another benefit of the index is its metric level of measurement.

The quality of the education scores and of the index I generated strongly depends on the underlying education variables, i.e. those indicating the years spent in education or the highest educational qualification, or both. The quality of the variable on the years spent in education has been discussed above. The measure of the highest educational qualification often has higher reliability and validity than the years spent in education because reporting a qualification is easier for most respondents. They usually remember having successfully completed an educational programme because it is an important event in their lives (Schneider, 2016). Concerning the comparability of the education scores, this should be adequate for univariate approaches if all countries implement a high-quality national education measure. In contrast, for bivariate and multivariate approaches Braun and Müller (1997) reflect that the comparability of the education measure additionally depends on the measurement quality of the criterion variable. This also must be measured in a comparable way across countries, or the education scores will not be comparable.

Educational Classifications

Applying an educational classification is another option for achieving a comparable education variable based on country-specific instruments measuring respondents' highest educational qualification. Education classifications are widely implemented in surveys and used by researchers, for example when analysing educational inequality (see section 1.3.1), or when using education as a proxy or control variable, and also when designing weights for cross-national surveys (see section 1.3.2). Commonly, classifications consist of a set of categories and each of them simultaneously covers different aspects, such as the level (e.g. low or higher education), the length, or the orientation (general vs. vocational) of a programme (Braun & Müller, 1997; Schneider, 2016). Therefore, the level of measurement of educational classifications often is not explicit: the levels are hierarchically ordered and thus ordinal, whereas programme orientation has a nominal level of measurement. The CASMIN and the ISCED classifications are commonly used.

In the 1970s researchers developed the CASMIN classification for ex-post harmonisation of education data within the project 'Comparative Analysis of Social Mobility in Industrial Nations' (König, Lüttinger, & Müller, 1988). Most studies analysing social mobility across cohorts and countries (see section 1.3.1) rely on national data. To use these data for conducting international comparisons the CASMIN classification can be applied. This classification distinguishes eight categories from elementary to higher education, and also differentiates general and vocational education within secondary education. In 2003 the CASMIN classification was updated, and categories added to reflect the extension of vocational training to different levels, and to better mirror the institutional diversification of higher general education. (Brauns, Scherer, & Steinmann, 2003). This classification has been implemented in the Scientific Use Files of the German Microcensus from 1976 to 2004, as well as in the German population census of 1970 and its supplementary 1971 survey (Lechert, Schroedter, & Lüttinger, 2006). However, to the best of my knowledge, until now no cross-national survey has implemented the CASMIN variable.

In contrast, the International Standard Classification of Education (ISCED) is widely used in surveys for ex-ante output harmonisation of educational attainment across countries. ISCED is an official classification developed by UNESCO to enable comparisons of country-specific education programmes for producing international education statistics (UNESCO-UIS, 2006). The classification was developed in the 1970s, and was updated in 1997 and 2011. I focus on the 1997 version of ISCED in this introduction and the papers of this dissertation because most of the datasets at time of conducting the analyses did not offer ISCED 2011.

ISCED 1997 distinguishes seven main educational levels:

- ISCED 0: Pre-primary education (or not completed primary education)
- ISCED 1: Primary education or first stage of basic education
- ISCED 2: Lower secondary or second stage of basic education
- ISCED 3: Upper secondary education
- ISCED 4: Post-secondary non-tertiary education
- ISCED 5: First stage of tertiary education
- ISCED 6: Second stage of tertiary education

These categories mirror the vertical (‘ladder’) aspect through the different educational levels, which also corresponds with higher cumulative duration in the education system. These hierarchical levels determine the first digit of the classification, which is most often used in surveys. In addition, the ISCED classification distinguishes up to three complementary dimensions at specific levels, namely the single duration of an educational programme, its orientation (general, pre-vocational or vocational) and whether it provides access to the next educational level or to the labour market (OECD, 1999; UNESCO-UIS, 2006). These dimensions are often neglected, especially in the ISCED version of 1997, which does not have a numeric code for these. This has been changed in the 2011 version of ISCED that offers a numeric coding scheme using three digits. From this detailed variable, researchers can derive their own education variable that captures the aspects of education they want to focus on (e.g. programme orientation or whether a programmes gives access to a higher level). Unfortunately, most surveys only consider the main educational level shown by the first digit. The new ISCED version distinguishes nine main education levels to better reflect the differences between short cycle, Bachelors and Masters degrees or equivalent qualifications at the tertiary level. More information on ISCED 2011 can be found in Schneider (2013), UNESCO-UIS (2012), OECD, European Union and UNESCO (2015).

In practical usage, the ISCED classification is applied to country-specific measurement instruments by assigning country-specific educational qualifications to the

ISCED classification. Official ISCED mappings can be found on the websites of UNESCO (http://uis.unesco.org/en/isced-mappings) and the European Commission's Communication and Information Resource Centre for Administrations, Businesses and Citizens (CIRCABC; https://circabc.europa.eu/w/browse/c2dc65ad-5163-4935-b0c2-e5ea1f44929b). The latter source provides annual information for European countries, while UNESCO only provides one ISCED mapping per country.

The ISCED classification is applied in official data such as EU-LFS and the European Union Statistics on Income and Living Conditions (EU-SILC), as well as in politically driven surveys such as the European Working Conditions Survey (EWCS) or PIAAC. Academic surveys, whether national (such as the German National Educational Panel Study (NEPS) or the German Socio-Economic Panel (SOEP)) or international (such as ESS, EVS, ISSP), are not obliged to use ISCED but they do often use it, or a closely related adaptation. ISCED is also quite often used in studies focusing on epidemiological and health issues, such as the National Health Interview Survey (NIS), the European Health Literacy Survey (HLS-EU), and the Survey of Health, Ageing and Retirement in Europe (SHARE).

For assessing the quality of the CASMIN and ISCED classifications, we must bear in mind that their reliability and validity depend to a large extent on the country-specific instruments measuring respondents' highest educational qualification. The validity of the education classifications also depends on the level of aggregation of the categories. As indicated, with the ISCED classification survey organisers can provide a detailed three-digit or a less detailed one-digit version. A further aggregation combines ISCED 1997 levels 0 and 1, and 5 with 6, where low numbers of people are assigned to the marginal levels. Quite often a categorical variable of just three categories is derived distinguishing low (ISCED levels 0-2), medium (ISCED levels 3-4) and high (ISCED levels 5-6) education. Aggregating categories of the ISCED classification affects the validity of the variable, its predictive power, and its comparability across countries (Schneider, 2010, 2018b). Concerning comparability, previous studies identified that the CASMIN classification better reflects the structure of European education systems and their qualifications, but for the U.S. education system the ISCED classification matches better than CASMIN does (Braun & Müller, 1997; Kerckhoff & Dylan, 1999; Kerckhoff et al., 2002).

As we have seen, different education variables exist and all of them are used in the research community. Most of these variables are derived from measurement instruments upon which their quality depends. Therefore, we next look at the measurement instruments in more detail.

1.4.2 Survey Instruments for Measuring Education in National, Cross-National and Migration Surveys

In addition to data processing, the quality of the education variable selected depends upon the measurement instruments, and how they are implemented, in different kinds of surveys.

National survey organisers are quite flexible when questioning respondents on their highest educational qualification and they often implement the measurement instrument that fits best with the purpose of their study. The survey organiser decides on the design of the instrument, the number of questions and the routing between them, question wording, interviewers' instructions as well as the answer categories (Granda et al., 2010). Although in most countries no standard instrument for measuring educational attainment exists, the instruments are often similar across surveys for the same country. Usually, these national instruments do not adequately address how to deal with foreign qualifications that have been introduced in the population, by immigrants and students for instance, who have been educated abroad. Incorporating these requires certain adaptations to the measurement instrument (Schneider, 2018b). Moreover, achieving comparability with other data sources or surveys also requires adherence to certain standards, for instance when implementing an international classification such as ISCED (Granda et al., 2010; Hoffmeyer-Zlotnik & Warner, 2014).

In contrast, national teams working for cross-national surveys usually have less freedom and flexibility when developing their survey instruments. The added requirement of achieving comparability across countries increases the surveys' complexity. As mentioned, the different education systems and the fact that qualifications cannot be translated require the harmonisation of respondents' educational attainment (Braun & Mohler, 2003; Schneider et al., 2016). Therefore, most surveys apply ex-ante output harmonisation, where they design country-specific measurement instruments from which they derive a comparable education variable, commonly the ISCED classification.

Lastly, national surveys on migration face somewhat similar challenges to cross-national surveys. These surveys often cover immigrants from various countries of origin, with different education systems and qualifications. However, many migration surveys lack the expertise for measuring and harmonising educational qualifications in a cross-cultural context and cannot afford the additional effort and costs. Three kinds of survey instruments are frequently implemented for measuring immigrants' educational attainment (Schneider, 2018a). Firstly, migration surveys can use the same instrument as national surveys, in which case, for instance, immigrants living in Germany would indicate the German qualification that best corresponds to their foreign one. Using a national instrument is not very costly but responding may be challenging for immigrants, especially when they are not (yet) familiar with the education system of the destination country. This approach, for instance, is used in the survey 'Experiences of Discrimination in Germany' (Beigang, Fetz, Foroutan, Kalkum, & Otto, 2016), which however draws the sample from the total population.

Another approach is used in the migration and refugee surveys of the German Socio-Economic Panel (SOEP) (Brücker et al., 2014; Kroh et al., 2016; Kühne & Kroh, 2017). The measurement instrument is more general and describes different educational levels by using generic terms that work for almost all education systems. However, in the SOEP the instrument is strongly inspired by the typical German instrument that consists of two questions (one on school-based education and another one on vocational and higher education). It also refers to German qualifications without explicitly naming them by their technical term, such as 'extended apprenticeship at a company', or attended a university/ college with either 'a more practical' or 'a more theoretical orientation'. The vagueness of these answer categories might also confuse respondents or lead to misinterpretation (Schneider, Briceno-Rosas, Ortmanns, & Herzing, 2018).

Finally, an approach that better reflects the differences in education systems is to offer immigrants country-specific lists of qualifications for the country in which they completed their education. This approach is used for instance in the German National Educational Panel Study (NEPS), starting cohort 6, round 2, in 2010/ 11 for the two largest groups of immigrants coming from Turkey or the former Soviet Union (FDZ-LIFBi, 2018; FDZ-LIFBi & infas, 2018a, 2018b). The SCIP project ('Socio-cultural integration processes among New Immigrants in Europe'), which studies integration trajectories of immigrants in four European countries, also offers country-specific educational categories for the largest immigrant groups in their survey, coming from Poland, Turkey, Morocco and Pakistan (Diehl, Lubbers, Mühlau, & Platt, 2016; Gresser & Schacht, 2015).

The idea of providing country-specific lists of educational qualifications is central to the recently developed CAMCES tool (see: https://www.surveycodings.org/levels-education) of the related project on 'Computer-Assisted Measurement and Coding of Educational Qualifications in Surveys'. The country-specific lists of educational qualifications can be regarded as standardised answer categories in the question on respondents' highest foreign educational attainment. For all these qualifications, the ISCED code as well as codes of related classifications, such as the 'edulvlb' classification of the ESS (ESS, 2010), are stored in the CAMCES database. To access the database in a survey, a questionnaire module has been developed that asks respondents, for instance, the country where they received their education before asking about their qualifications (Schneider & Ortmanns, 2019). The CAMCES tool can be used in cross-national and migration surveys. Until now, it has been implemented in the SOEP migration and refugee surveys (Briceno-Rosas, Liebau, Ortmanns, Pagel, & Schneider, 2018; Schneider et al., 2018). The ReGES project on 'Refugees in the German Education system' of the LIFBi (Gentile, Heinritz, & Will, 2019; Will, Gentile, Heinritz, & von Maurice, 2018) uses country-specific lists of educational qualifications, which are derived from the CAMCES database.

This section described the different education variables and their related instruments and evaluated their quality from a conceptual point of view, to prepare for the presentation of the four empirical papers forming this dissertation.

1.5 The Papers of the Dissertation

In my dissertation, I analyse the quality of educational attainment measures. In papers I, II and III I evaluate the quality of the widely used ISCED variable by assessing the reliability and comparability of this measure in ten cross-national surveys. These papers have a methodological focus, largely from the perspective of survey organisers who produce data. In contrast, paper IV has a substantive focus and uses the education variable as a proxy for competencies. In addition to this paper, I conducted a construct validation of different education measures to decide which variable to include in the main analysis of paper IV. Thereby I take on the perspective of a data user.

Papers I and II, which are closely related, assess the aggregate reliability and comparability of the ISCED variable in a wide range of cross-national surveys. I compare the education distribution using the ISCED variable for the same countries and the same surveys across years (paper I) and the same countries and years across surveys (papers I and II). To compare the distributions, I calculate Duncan's Dissimilarity Index (Duncan & Duncan, 1955) that indicates what percentage would need to change to another ISCED category to achieve an equal education distribution across years or surveys. Thus in these papers, I use repeated measures of different years or surveys and thereby evaluate the aggregated reliability of the ISCED variable. Although the surveys cover different respondents, they all use randomised probability-based samples so that the education distribution should be similar when analysing the same age groups for the same country across surveys. The distribution should also be similar when comparing it for the same country and survey across consecutive years, as the actual distribution in the population changes only slowly through cohort replacement. Thus, with the dissimilarity index I assess the aggregated reliability of the ISCED variable. If the data for the same country and year are not reliable at the aggregated level they are also not comparable across countries and years. Thus, I also simultaneously assess the comparability of the ISCED variable.

These papers update and extend earlier studies of Schneider (2008, 2009) and Kieffer (2010) by using survey data from 2008 to 2012 and by including many more surveys that have not been analysed before in this context. Paper I compares the education distributions of four social science surveys, namely the Eurobarometer, the European Social Survey (ESS), the European Values Study (EVS) and the International Social Survey Programme (ISSP). Paper II extends the number of surveys by adding official data of the European Union Labour Force Survey (EU-LFS), the European Survey of Income and Living Conditions (EU-SILC) as well as OECD data of the Programme for the International Assessment of Adult Competencies (PIAAC). Overall, the results show an almost stable education distribution when comparing the ISCED variable for the same country and survey across years, at least as long as the measurement instrument and the applied coding procedure are consistent. Comparing the education distributions across surveys within countries and years, I partly identify very large inconsistencies that cannot reflect actual differences. These inconsistencies indicate a severe problem in the aggregated reliability as well as in the comparability across surveys. This worrisome finding is in line with previous studies (Kieffer, 2010;

Schneider, 2009). Therefore, in paper II, I conduct an exploratory analysis to identify the reasons for the largest inconsistencies. In doing so, I refer to the Total Survey Error (TSE) framework (Groves et al., 2009; Groves & Lyberg, 2010) and distinguish between errors related to the measurement dimension of the TSE and errors related to the representation of the population. I find many errors related to measurement, in particular the categories of the education question, which can use ambiguous terms or descriptions of qualifications, and in the assignment of the country-specific educational qualification into ISCED. For a few cases, I can exclude measurement errors and thus I suspect that errors related to the representation of the population cause these inconsistencies. However, this analysis consists only of qualitative analyses of the measurement instruments and the process of the ISCED coding.

A more advanced and systematic analysis to explain these inconsistencies is conducted in paper III, where I analyse the impact of 15 survey characteristics using quantitative methods. Such a comprehensive analysis to quantitatively examine these inconsistencies has not to my knowledge been conducted before. This paper also includes data from the Adult Education Survey (AES), the European Quality of Life Survey (EQLS), and the European Working Conditions Survey (EWCS). Thus, I compare the education distribution and the survey characteristics of nine surveys to benchmark data from the EU-LFS for 31 European countries. In this paper, I refer to the TSE in more detail and examine survey characteristics relating to different kinds of errors. Concerning errors related to the representation of the population, I consider the sampling design, final sampling unit, sample size, response rate, whether survey participation is mandatory, fieldwork duration, and also an index to validate probability sampling and an index of the age and gender distribution. On the measurement dimension, I look at the response categories of the education question, and consider whether proxy-reporting is allowed, whether the education information is taken from a register, whether the official ISCED mappings are applied, and I also assess the degree of centralisation when applying ISCED. I also consider the mode of data collection and the fieldwork agency, which may affect both measurement and representation. I estimate regression models to analyse the effect of the survey characteristics on the inconsistencies in the education distribution across surveys. The results support the expectations of previous studies (Kieffer, 2010; Schneider, 2008, 2009) as well as of papers I and II and highlight a predominant effect of measurement errors. In particular, deviations in the application of the official ISCED mappings, as well as differences in the education categories cause large inconsistencies in the education distribution across surveys when they should theoretically be the same. Concerning the survey characteristics related to the representation of the population, I find that apart from the sampling design, the other survey characteristics are not systematically related to inconsistencies.

Overall, the results of papers I, II and III illustrate a lack of aggregated reliability and thus also lack of comparability of the harmonised ISCED for the same country and year across surveys. Thus, at the moment *even at the national level* the education distribution is not comparable across surveys. While this is the case, there is no need to assess comparability *across countries.*

In paper IV, I change perspective and look at the education variable from that of a data user. I conduct a use case study that examines the effects of three mechanisms on immigrants' German language proficiency – language exposure, incentives and efficiency, the latter being operationalised by educational attainment (among other indicators). To reflect that the process of learning a language is not linear (Esser, 2006; Stevens, 1999) and so the effects of the mechanisms may vary over time, this study looks at two different groups of immigrants, namely established and recently arrived immigrants. These groups differ in their length of stay in Germany and in their German language proficiency. By using panel data from the SOEP for both groups, this study additionally considers effects through intra-individual changes within the mechanisms. This is the major contribution of this study, which allows me to consider almost the whole process of learning the German language. The results indicate that language exposure, efficiency and incentives all enhance immigrants' German language skills, which is in line with previous research (Chiswick & Miller, 1995, 2001; Dustmann, 1994; Espenshade & Fu, 1997; Esser, 2006; van Tubergen, 2010). Concerning the effects of intra-individual changes, these affect the mechanism of language exposure for recently arrived immigrants, and the mechanism of incentives for the established immigrants.

A central component of the mechanism of efficiency is immigrants' cognitive ability. This is not measured directly in the SOEP surveys and therefore I use the education variable as a proxy for cognitive ability (see section 1.3.2), which is frequently done in this field of research (Esser, 2006). In this introductory chapter, I will now reflect on the measurement of education in more detail than in paper IV itself, to better establish the links with the overall dissertation. Most studies use the education variable indicating the years spent in the education system (Chiswick & Miller, 1995, 2001; Dustmann, 1994; Espenshade & Fu, 1997; Mesch, 2003; Stevens, 1999). Only a few studies use other variables, such as hypothetical years of schooling derived from the ISCED variable (van Tubergen & Wierenga, 2011), a categorical variable related to the distinction of low, medium and high education (van Tubergen, 2010), or a more detailed variable that considers educational institutions (Dustmann, 1997). The rationale for the selection and coding of the education variable is often underdeveloped, which is especially problematic if a survey offers more than one education measures. This also applies to SOEP and therefore I conducted an additional analysis to identify the best education measure for this paper. This analysis is not part of the paper itself and can be found in section 6 of this thesis.

In this construct validation exercise, I compare the quality of a handful of education variables derived from different instruments all measuring immigrants' highest educational qualification awarded in their country of origin. I analysed the predictive power of the variables when estimating the impact of educational attainment on immigrants' German language proficiency. The results illustrate that the adjusted R^2 is quite similar across the different measures. I decided to implement the education index that combines two direct education measures, namely the metric variable indicating the years spent in school and the categorical variable on educational attainment. To generate this index, I apply a scoring approach (see section 1.4.1). The benefit of this education index is that it covers as much information on immigrants' education as possible from two independent measurement instruments. In this study, I am interested neither in the signalling effects of the qualifications, nor the impact of each additional year spent in education. Thus, combining these two variables to generate a more comprehensive education variable seems reasonable. Moreover, I use the education variable as a proxy for cognitive skills and competencies and therefore favour an education variable with a metric level of measurement. The education index meets these criteria and also has higher predictive power than the metric years of schooling variable alone. In the context of this analysis, unfortunately I cannot assess the reliability of the index as well as of the other education variables because I do not have repeated measures with the same instrument. I also do not assess the comparability of the different education variables by running separate analyses for groups of immigrants according to their country of origin because the immigrants come from many different countries and the numbers of cases from each country are too small.

When predicting immigrants' German language proficiency in paper IV, I implemented the education index without problems. In line with the results of previous studies (Chiswick & Miller, 1995, 2001; Dustmann, 1994; Espenshade & Fu, 1997; Esser, 2006; Mesch, 2003; van Tubergen, 2010), I identify a positive and statistically significant effect of immigrants' homeland education on their German language proficiency.

1.6 Conclusion and Discussion

As seen in the detailed descriptions of the papers making up this dissertation, the quality of education variables in cross-national and migration surveys is evaluated. In doing so, it considers the perspectives of survey organisers as data producers and of researchers as data users.

The methodological papers I, II and III of this thesis emphasise the perspective of survey organisers providing the data. The surveys used in these papers measure respondents' educational attainment with country-specific instruments and derive from these a comparable education variable using the ISCED classification. The papers highlight severe problems in the data quality of the resulting harmonised education variable. Major inconsistencies in the education distributions within countries and years indicate a lack of reliability at the aggregated level, which implies difficulties with data comparability, especially across surveys. Alarmingly, education distributions of different surveys for the same country and year are not comparable. While this is the case, we also must be careful when comparing education distributions across countries. Not every survey seems to be suitable for making country comparisons. The EU-LFS, EU-SILC and PIAAC, for instance, performed quite well in the analysis and did not show large inconsistencies, so these surveys can sensibly be used for country comparisons. In contrast, I would not recommend using Eurobarometer or ISSP data for country comparisons owing to the large inconsistencies in them.

To enhance the quality of the education variable, survey organisers should firstly improve the coding of the country-specific education categories to the ISCED classification, and especially avoid accidental misclassifications by applying accurately the official ISCED mappings. Secondly, for the country-specific answer categories of the education question, vague or ambiguous terms and descriptions of qualifications should be avoided. Instead, explicitly naming the qualifications and using common terminology that is understandable to all respondents is desirable. I have discussed in depth the results and their consequences within the respective sections of the papers I, II and III. Therefore, here I will discuss only the central aspects of these findings and put them into a larger context.

To assess these results and recommendations from papers I, II and III, we should be aware of some peculiarities of the ISCED classification and its quality assurance. Firstly, some ISCED criteria are not as explicit as they could be, resulting in different interpretations across countries and thus to variations when assigning ISCED codes to national qualifications. Secondly, the ISCED classification itself is vulnerable to political influence. National statistical offices or ministries of education determine how the ISCED codes are mapped to their national educational qualifications. However, they seem not always to act independently of political interests or target agreements of national or international institutions. Thirdly, the quality assurance by UNESCO, the custodian of ISCED, seems to be insufficient. As indicated, there are some qualifications for which it might be ambiguous which ISCED code to assign or where the assigned ISCED codes do not seem to follow the ISCED criteria. In these cases, it is not documented why a particular code was assigned. There are no reports of discussions, and the differing points of view on such cases are not transparent. Finally, with ISCED 1997, Eurostat did not publish the country-specific education variables or any other documentation on the coding of the country-specific qualifications to the ISCED variable included in the EU-LFS and EU-SILC. Fortunately this has changed, and since 2013 the official ISCED mappings for all European countries, including the codings used in the EU-LFS and EU-SILC, are published annually. However, these documents are still hard to find on the Eurostat website, and the country-specific education variables are still not included in the datasets. As a consequence, there is a lack of knowledge the quality of the classification and its implementation, and lack of trust in its quality; this contributes to the difficulties for survey organisers in properly measuring educational attainment and coding this into ISCED.

With this in mind, survey organisers have two options to improve the quality of the ISCED variable. The first option is that survey organisers and country teams use the publically available ISCED 2011 mappings to check the coding of the country-specific educational qualification into ISCED. Thereby the country teams identify and then can correct accidental coding mistakes. However, this approach does not increase trust in the ISCED classification itself and in UNESCO's quality assurance. The second option is more radical: organisers of academic surveys, which are not obliged to follow the official ISCED, can choose to deviate from the official standard or develop their own harmonised education variable. This is already done in the ESS that since 2010 uses the 'edulvlb' variable (ESS, 2010). This variable is closely linked to ISCED 2011 but contains intentional deviations from the official ISCED mappings for specific qualifications of single countries, in order to improve cross-country comparability. The 'edulvlb' variable has also been applied in the 2017 wave of the EVS (EVS, 2019; Losi, Maineri, Luijkx, Schneider, & Ortmanns, 2019). Both surveys have documented these deviations, allowing data users to recode the 'edulvlb' variables into the official ISCED if they wish so.

As a next step for further research, it would be worth replicating these methodological studies using more recent data in which ISCED 2011 is implemented. We could check whether the survey organisers, in particular those of surveys showing large inconsistencies, have used the new ISCED mappings for checking their coding into ISCED, or for improving their measurement instruments. For survey organisers and data users, it would be important to know if the inconsistencies in the education distributions across surveys still exist or have been reduced with the implementation of ISCED 2011 and the availability of the ISCED mappings. Moreover, it would be helpful to examine empirically what consequences the non-comparable education distributions have for analysis of statistical correlations (including regression analysis), in substantive research. To estimate this, we have to implement the education variables of the different surveys into the otherwise constant model. Thereby we can assess if the different education distributions change the results of substantive analyses and thus how meaningful these are.

Paper IV of this thesis and its related additional analysis takes over the perspective of data users. The results of the construct validation I conducted of different education variables show only small differences in the predictive power of these variables when assessing immigrants' German language proficiency. Even when running the main analysis in paper IV using different education variables, the results do not change significantly. This might initially be quite surprising because the education variables focus on different aspects of education (time spend in education and educational qualifications), or they measure the same aspect but differ in their degree of detail. However, it seems that an aspect that is central to acquiring German language proficiency is captured in every education variable. It is part of the variance in each variable and is what the variables have in common, making their predictive power almost the same. For the analysis in paper IV, I decided to use the education index that combines the information of two independent measurement instruments (years spent in school and educational qualifications) and has a metric level of measurement. This index performs quite well in the analysis and the observed effect of this is in line with previous studies assessing immigrants' second language skills (Carliner, 2000; Chiswick & Miller, 2001; Espenshade & Fu, 1997; Esser, 2006; van Tubergen, 2010).

As indicated in section 1.4.2 and the additional analysis to paper IV, currently a wide range of measurement instruments is used for migration surveys and all have their merits and disadvantages. In a first step, it would be worth evaluating the different instruments more systematically to better assess their validity and reliability. Therefore, it would be advisable to conduct more analysis like the construct validation in this thesis. Also, conducting more pretests of these instruments is desirable to assess how well immigrants cope with them, and whether these instruments are intuitive and have sufficient quality of measurement. Based on such pretests and further analyses, survey organisers could improve the measurement instruments for migrant surveys and again evaluate these.

As mentioned, in the context of paper IV, I developed a new education index for measuring immigrants' homeland education. In future research, it would be good to validate this index further and extend its use to different research areas. To ensure that the index is good enough to use it as a proxy for competencies, it should be validated with data that contain a direct measure of respondents' competencies, such as that of PIAAC. Unfortunately, PIAAC does not contain a direct measure of the number of years respondents spent in the education system, which is integral in the index. Instead, we could use the derived hypothetical years of education variable for this, which might not be ideal but is still appropriate.

To better validate the operationalisation of the index concerning its combined measure of years spent in education and educational attainment, we can use data of the ESS or EVS, which employ two separate instruments for these variables. Using these surveys, we can compare and validate each education measure separately as well as the index. Additionally, with these data, we can put the results into a larger context by repeating this analysis and using different dependent variables, such as occupation, social status or attitudes and behaviours. Moreover, through running analyses with the cross-national data of PIAAC, the ESS or the EVS, it is also possible to check whether the new education index performs similarly across countries or whether there are differences. Additionally, we can test which variable for measuring educational attainment fits best in the education index – the country-specific education measure or the harmonised ISCED variable. This might be likely also to vary across countries or surveys.

To sum up, this thesis contributes to research into the quality of the education variable in large cross-national and recent migration surveys. The results of the methodological papers I, II and III, in which I analysed data for 31 countries for the years 2008 to 2012 for ten cross-national surveys, clearly illustrate quality concerns with the harmonised ISCED variable. The inconsistencies identified in the education distribution across surveys signal that reliability of the data at the aggregated level is not a given. Moreover, the lack of reliability also raises reasonable doubts about the comparability of the education variable across countries. These findings are in line with previous studies, which years ago raised similar quality concerns for a smaller number of countries and surveys in older data (Kieffer, 2010; Schneider, 2008, 2009, 2010). The quality criterion of objectivity has not been tested in this thesis because I assume that objectivity of the education measure is of an equal standard across surveys. Overall, this research indicates that these quality concerns are not limited to single countries, years or surveys. Unfortunately, many of the surveys analysed in this thesis are affected by these problems, and therefore quality concerns for the education variable are still acute. Consequently, survey organisers should take these concerns seriously and improve the quality of the education variable for their survey, in particular concerning the coding of the country-specific education categories to the ISCED classification, and by enhancing the measurement instrument itself. Data user should also be aware of these problems, especially when conducting comparative research using data of different surveys. Regarding the validity of different education variables, I looked at migration surveys that often include different instruments for measuring immigrants' homeland education, and therefore are a good vehicle for running construct validation analysis. I observe that although the variables measure different aspects of education and are operationalised differently, their predictive power is similar when estimating migrants' German language skills. This is good news for researchers because their results will probably change only marginally when using another education variable in their substantive analysis. Nevertheless, the researcher should reflect on the different purposes of the education variables and the different measures, depending on the reasons for using the education variable in their analyses.

1.7 References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does it Matter? *Public Opinion Quarterly*, *70*(5), 676–703. doi:10.1093/poq/nfl037

Allmendinger, J. (1989). Educational Systems and Labor Market Outcomes. *European Sociological Review*, *5*(3), 231–250. doi:10.1093/oxfordjournals.esr.a036524

Almond, G. A., & Verba, S. (1963). *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton University Press.

Becker, G. S. (1993). *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education.* (3rd ed.). Chicago: The University of Chicago Press.

Beigang, S., Fetz, K., Foroutan, N., Kalkum, D., & Otto, M. (2016). *Experiences of Discrimination in Germany. Initial Results of a Representative Survey and a Survey of the People Affected* (Federal Anti-Discrimination Agency, Ed.). Retrieved from: http://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/Umfragen/Handout_en_Umfrage_Diskriminierung_in_Dtschl_2015.pdf?__blob=publicationFile&v=3 (accessed April 2020)

Bekhuis, H., Lubbers, M., & Verkuyten, M. (2014). How Education Moderates the Relation Between Globalization and Nationalist Attitudes. *International Journal of Public Opinion Research*, *26*(4), 487–500. doi:10.1093/ijpor/edt037

Bell, D. (1999). *The Coming of Post-Industrial Society. A Venture in Social Forecasting* (3rd ed.). New York: Basic Books.

Billiet, J., Philippens, M., Fitzgerald, R., & Stoop, I. (2007). Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents. *Journal of Official Statistics*, *23*(2), 135–162.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: John Wiley & Sons.

Blossfeld, H.-P., & Shavit, Y. (1993). Persisting Barriers. Changes in Educational Opportunities in Thirteen Countries. In Y. Shavit & H.-P. Blossfeld (Eds.), *Persistent Inequality. Changing Educational Attainment in Thirteen Countries* (pp. 1–24). Boulder: Westview Press.

Bol, T. (2015). Has Education Become More Positional? Educational Expansion and Labour Market Outcomes, 1985–2007. *Acta Sociologica*, *58*(2), 105–120. doi:10.1177/0001699315570918

Bol, T., & van de Werfhorst, H. G. (2013). Educational Systems and the Trade-Off Between Labor Market Allocation and Equality of Educational Opportunity. *Comparative Education Review*, *57*(2), 285–308.

Bol, T., & van de Werfhorst, H. G. (2016). Measuring Educational Institutional Diversity: Tracking, Vocational Orientation and Standardisation. In A. Hadjar & C. Gross (Eds.), *Education Systems and Inequalities: International Comparisons* (pp. 73–94). Bristol: Policy Press.

Bolzendahl, C. I., & Myers, D. J. (2004). Feminist Attitudes and Support for Gender Equality: Opinion Change in Women and Men, 1974-1998. *Social Forces*, *83*(2), 759–789. doi:10.1353/sof.2005.0005

Boudon, R. (1974). *Education, Opportunity, and Social Inequality. Changing Prospects in Western Society*. New York: John Wiley & Sons.

Bourdieu, P. (1983). Ökonimisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Ed.), *Soziale Ungleichheiten (Soziale Welt Sonderband 2)* (pp. 183–198). Göttingen.

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.

Brauns, H., Scherer, S., & Steinmann, S. (2003). The CASMIN Educational Classification in International Comparative Research. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book of Demographic and Socio-Economic Variables* (pp. 221–244). New York: Kluwer Academic/ Plenum Publishers.

Breen, R., & Goldthorpe, J. H. (1997). Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society*, *9*(3), 275–305. doi:10.1177/104346397009003002

Breen, R., & Jonsson, J. O. (2000). Analyzing Educational Careers: A Multinomial Transition Model. *American Sociological Review*, *65*(5), 754–772. doi:10.2307/2657545

Briceno-Rosas, R., Liebau, E., Ortmanns, V., Pagel, L., & Schneider, S. L. (2018). *Documentation on ISCED Generation Using the CAMCES Tool in the IAB-SOEP Migration Samples M1/M2.* SOEP Survey Papers 579: Series D. Berlin: DIW/SOEP. Retrieved from: https://www.diw.de/documents/publikationen/73/diw_01.c.608161.de/diw_ssp0579.pdf (accessed April 2020)

Brown, D. K. (1995). *Degree of Control: A Sociology of Educational Expansion and Occupational Credentialism*. New York: Teacher College Press.

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 216: Series C. Berlin: DIW/SOEP. Retrieved from: http://doku.iab.de/fdz/iab_soep/surveypaper_en.pdf (accessed April 2020)

Bukodi, E., & Goldthorpe, J. H. (2016). Educational Attainment - Relative or Absolute - As a Mediator of Intergenerational Class Mobility in Britain. *Research in Social Stratification and Mobility*, *43,* 5–15. doi:/10.1016/j.rssm.2015.01.003

Burris, V. (1983). The Social and Political Consequences of Overeducation. *American Sociological Review*, *48*(4), 454–467. doi:10.2307/2117714

Carliner, G. (2000). The Language Ability of U.S. Immigrants : Assimilation and Cohort Effects. *International Migration Review, 34*(1), 158–182. doi:10.1177/019791830003400107

Chiswick, B. R., & Miller, P. W. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics*, *13*(2), 246–288.

doi:10.1086/298374

Chiswick, B. R., & Miller, P. W. (2001). A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada. *Demography*, *38*(3), 391–409. doi:10.2307/3088354

Clogg, C. C., & Shockey, J. W. (1984). Mismatch Between Occupation and Schooling: A Prevalence Measure, Recent Trends and Demographic Analysis. *Demography*, *21*, 235–257.

Cobalti, A. (1990). Schooling Inequalities in Italy : Trends Over Time. *European Sociological Review, 6*(3), 199–214. doi:10.1093/oxfordjournals.esr.a036562

Coenders, M., & Scheepers, P. (2003). The Effect of Education on Nationalism and Ethnic Exclusionism : An International Comparison. *Political Psychology*, *24*(2), 313–343.

Collins, R. (1971). Functional and Conflict Theories of Educational Stratification. *American Sociological Review*, *36*(6), 1002–1019. doi:10.2307/2093761

Collins, R. (1979). *The Credential Society: An Historical Sociology of Education and Stratification*. New York: Academic Press.

Di Stasio, V., Bol, T., & van de Werfhorst, H. G. (2016). What Makes Education Positional? Institutions, Overeducation and the Competition for Jobs. *Research in Social Stratification and Mobility*, *43*, 53–63. doi:10.1016/j.rssm.2015.09.005

Diehl, C., Lubbers, M., Mühlau, P., & Platt, L. (2016). Starting Out: New Migrants' Socio-Cultural Integration Trajectories in Four European Destinations. *Ethnicities*, *16*(2), 157–179. doi:10.1177/1468796815616158

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone , Interactive Voice Response (IVR) and the Internet. *Social Science Research*, *38*(1), 1–18. doi:10.1016/j.ssresearch.2008.03.007

Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, *20*(2), 210–217. doi:10.2307/2088328

Dustmann, C. (1994). Speaking Fluency, Writing Fluency and Earnings of Migrants. *Journal of Population Economics*, *7*, 133–156. doi:10.1007/BF00173616

Dustmann, C. (1997). The Effects of Education, Parental Background and Ethnic Concentration on Language. *The Quarterly Review of Economics and Finances*, *37*(1), 245–262. doi:10.1016/S1062-9769(97)90068-5

Dustmann, C., & van Soest, A. (2002). Language and the Earnings of Immigrants. *Industrial and Labor Relations Revie*, *55*(3), 473–492. doi:10.2307/2696052

Ehling, M. (2003). Harmonising Data in Official Statistics: Development, Procedures, and Data Quality. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book of Demographic and Socio-Economic Variables* (pp. 17–31). New York: Kluwer Academic/ Plenum Publishers.

Erikson, R., & Goldthorpe, J. H. (1992). *The Constant Flux. A Study of Class Mobility in Industrial Societies.* Oxford: Clarendon Press.

Erikson, R., & Jonsson, J. O. (1996). *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder: Westview Press.

Espenshade, T. J., & Fu, H. (1997). An Analysis of English-Language Proficiency among U.S. Immigrants. *American Sociological Review, 62*(2), 288–305. doi:10.2307/2657305

ESS. (2010). *European Social Survey - 2010 Documentation Report, Appendix A1 Education* (4.2). Retrieved from: http://www.europeansocialsurvey.org/docs/round5/survey/ESS5_appendix_a1_e04_2.pdf (accessed April 2020)

Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt: Campus Verlag.

Eurostat. (2018). *European Statistics Code of Practice - For the National Statistical Authorities and Eurostat (EU Statistical Authority).* Luxembourg: Publication Office of the European Union. doi:10.2785/798269

EVS. (2019). *EVS 2017 - Variable Report, Appendix 3: Educational Level EVS 2017.* GESIS Variable Reports 2019/12. Documentation of the Second Pre-Release 2019/07/15. Cologne: GESIS Data Archive. doi:10.4232/1.13314

FDZ-LIFBi. (2018). *Codebook. NEPS Starting Cohort 6 - Adults. Adult Education and Lifelong Learning.* Related to Scientific Use File Version 9.0.1. Bamberg: Leibniz Institute for Educational Trajectories (version: 9). Retrieved from: https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC6/9-0-1/SC6_9-0-1_Codebook_en.pdf (accessed April 2020)

FDZ-LIFBi, & infas. (2018a). *Codebook. NEPS Starting Cohort 6 - Adults. Adult Education and Lifelong Learning.* Related to Scientific Use File Version 9.0.1. Side Sheet: Educational Degrees of the Former Soviet Union. Bamberg: Leibniz Institute for Educational Trajectories.

FDZ-LIFBi, & infas. (2018b). *Codebook. NEPS Starting Cohort 6 - Adults. Adult Education and Lifelong Learning.* Related to Scientific Use File Version 9.0.1. Side Sheet: Educational Degrees of Turkey. Bamberg: Leibniz Institute for Educational Trajectories.

Friedberg, R. M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics, 18*(2), 221–251. doi: 10.1086/209957

Fujihara, S., & Ishida, H. (2016). The Absolute and Relative Values of Education and the Inequality of Educational Opportunity: Trends in Access to Education in Postwar Japan. *Research in Social Stratification and Mobility*, *43*, 25–37. doi:10.1016/j.rssm.2016.03.001

Gambetta, D. (1987). *Were They Pushed or Did They Jump? Individual Decision Mechanims in Education*. Cambridge: Cambridge University Press.

Gangl, M. (2002). Changing Labour Markets and Early Career Outcomes: Labour Market Entry in Europe Over the Past Decade. *Work, Employment and Society, 16*(1), 67–90. doi:10.1177/09500170222119254

Gangl, M. (2003). The Only Way is Up ? Employment Protection and Job Mobility among Recent Entrants to European Labour Markets. *European Sociological Review, 19*(5), 429–449. doi:10.1093/esr/19.5.429

Gentile, R., Heinritz, F., & Will, G. (2019). *Übersetzung von Instrumenten für die Befragung von Neuzugewanderten und Implementation einer audio-basierten Interviewdurchführung.* LIfBi Working Paper No. 86. Bamberg: Leibniz Institute for Edcuational Trajectories. Retrieved from: https://www.lifbi.de/Portals/13/LIfBi Working Papers/WP_LXXXVI.pdf (accessed April 2020)

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). *Health Behavior and Health Education: Theory, Research and Practice* (4th ed.). San Francisco: Jossey-Bass.

Granda, P., Wolf, C., & Hadorn, R. (2010). Harmonizing Survey Data. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, … T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 315–332). Hoboken: John Wiley & Sons.

Gresser, A., & Schacht, D. (2015). *SCIP Survey. Causes and Consequences of Socio-Cultural Integration Processes among New Immigrants in Europe. Methodological Report*. Retrieved from: www.scip-info.org (accessed April 2020)

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken: John Wiley & Sons.

Groves, R. M., & Lyberg, L. E. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, *74*(5), 849–879. doi:10.1093/poq/nfq065

Groves, R. M., Presser, S., & Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, *68*(1), 2–31. doi:10.1093/poq/nfh002

Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., Mohler, P. P., Smith, T. W. (2010). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Hoboken: John Wiley & Sons.

Harkness, J. A. , van de Vijver, F. J. R., & Mohler, P. P. (2003). *Cross-Cultural Survey Methods*. Hoboken: John Wiley & Sons.

Harris, R. J., & Fireston, J. M. (1998). Changes in Predictors of Gender Role Ideologies among Women : A Multivariate Analysis. *Sex Roles*, *38*(3/4), 238–252. doi:10.1023/A:1018785100469

Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (2nd ed., pp. 143–171). doi:10.1007/978-3-642-20072-4

Hauser, R. M., & Featherman, D. L. (1976). Equality of Schooling : Trends and Prospects. *Sociology of Education*, *49*(2), 99–120. doi:10.2307/2112516

Helberger, C. (1988). Eine Überprüfung der Linearitätsannahme der Humankapitaltheorie. In H.-J. Bodenhöfer (Ed.), *Bildung, Beruf, Arbeitsmarkt* (pp. 151–170). Berlin: Duncker & Humbolt.

Helmschrott, S., & Martin, S. (2014). Nonresponse in PIAAC Germany. *Methods, Data, Analyses*, *8*(2), 243–266. doi:10.12758/mda.2014.010

Hillmert, S., & Jacob, M. (2005). Institutionelle Strukturierung und inter-individuelle Variation. Zur Entwicklung herkunftsbezogener Ungleichheiten im Bildungsverlauf. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *57*(3), 414–442. doi:10.1007/s11577-005-0183-8

Hirsch, F. (1976). *Social Limits to Growth*. Cambridge: Harvard University Press.

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2007). How to Survey Education for Cross-National Comparisons: The Hoffmeyer-Zlotnik/Warner-Matrix of Education. *Metodološki Zvezki*, *4*(2), 117–148.

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research*. Dordrecht: Springer Science+Business Media.

Hout, M. (1989). *Following in Father's Footsteps. Social Mobility in Ireland*. Cambridge, Mass.: Harvard University Press.

Hübler, O. (1984). Zur empirischen Überprüfung alternativer Theorien der Verteilung von Arbeitseinkommen - Ökonometrische Ein- und Mehrgleichungsmodelle. In L. Bellmann, K. Gerlach, & O. Hübler (Eds.), *Lohnstruktur in der Bundesrepublik Deutschland. Zur Theorie und Empirie der Arbeitseinkommen* (pp. 17–189). Frankfurt am Main: Campus Verlag.

Hyman, H. H., & Wright, C. R. (1979). *Education's Lasting Influence on Values*. Chicago: University of Chicago Press.

International Labour Office. (2012). *International Standard Classification of Occupations (ISCO-08). Volume 1: Structure, Group Definitions and Correspondence Tables*. Retrieved from: https://www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf (accessed April 2020)

Jackson, M., Erikson, R., Goldthorpe, J. H., & Yaish, M. (2007). Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-Level Courses in England and Wales. *Acta Sociological*, *50*(3), 211–229. doi:10.1177/0001699307080926

Kerckhoff, A. C. (1976). The Status Attainment Process: Socialization or Allocation? *Social Forces*, *55*(2), 368–381. doi:10.1093/sf/55.2.368

Kerckhoff, A. C., & Dylan, M. (1999). Problems with International Measures of Education. *Journal of Socio-Economics*, *28*(6), 759–775. doi:10.1016/S1053-5357(99)00056-6

Kerckhoff, A. C., Ezell, E. D., & Brown, J. S. (2002). Toward an Improved Measure of Educational Attainment in Social Stratification Research. *Social Science Research*, *31*(1), 99–123. doi:10.1006/ssre.2001.0720

Kieffer, A. (2010). Measuring and Comparing Levels of Education: Methodological Problems in the Classification of Educational Levels in the European Social Surveys and the French Labor Force Surveys. *Bulletin de Méthodologie Sociologique*, *107*(1), 49–73. doi:10.1177/0759106310369974

Kingston, P. W., Hubbard, R., Lapp, B., Schroeder, P., & Wilson, J. (2003). Why Education Matters. *Sociology of Education*, *76*(1), 53–70. doi:10.2307/3090261

Kleinert, C., & Matthes, B. (2009). *Data in the Field of Adult Education and Lifelong Learning: Present Situation, Improvements and Challenges*. RatSWD Working Paper Series No. 91. Berlin: Council for Social and Economic Data (RatSWD). Retrieved from: https://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_91.pdf (accessed April 2020)

Koch, A., Halbherr, V., Stoop, I. A. L., & Kappelhof, J. W. S. (2014). *Assessing ESS Sample Quality by Using External and Internal Criteria.* Mannheim: European Social Survey, GESIS. Retrieved from: https://www.europeansocialsurvey.org/docs/round5/methods/ESS5_sample_composition_assessment.pdf (accessed Arpil 2020)

König, W., Lüttinger, P., & Müller, W. (1988). *A Comparative Analysis of the Development and Structure of Educational Systems. Methodological Foundations and the Construction of a Comparative Educational Scale.* CASMIN-Working Paper No. 12. Mannheim: Univesität Mannheim. Retrieved from: https://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/mikrodaten_tools/CASMIN/Koenig_Casmin.pdf (accessed April 2020)

Krebs, D., & Menold, N. (2014). Gütekriterien quantitativer Sozialforschung. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 425–438). Wiesbaden: Springer Fachmedien.

Kroh, M., Brücker, H., Kühne, S., Liebau, E., Schupp, J., Siegert, M., & Trübswetter, P. (2016). Das Studiendesign der IAB-BAMF-SOEP-Befragung von Geflüchteten. SOEP Survey Papers 365: Series C. Berlin: DIW/SOEP. Retrieved from: http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0365.pdf (accessed April 2020)

Kühne, S., & Kroh, M. (2017). *The 2015 IAB-SOEP Migration Study M2: Sampling Design, Nonresponse, and Weighting Adjustment.* Survey Papers 473: Series C. Berlin: DIW/SOEP. Retrieved from: https://www.diw.de/documents/publikationen/73/diw_01.c.571195.de/diw_ssp0473.pdf (accessed April 2020)

Lechert, Y., Schroedter, J. H., & Lüttinger, P. (2006). Die Umsetzung der Bildungsklassifikation CASMIN für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971 und die Mikrozensen 1976-2004. ZUMA-Methodenbericht. Mannheim: Zentrum Für Umfragen, Methoden Und Analysen (ZUMA)., 12. Retrieved from: https://www.ssoar.info/ssoar/handle/document/26235 (accessed April 2020)

LEXICO. (2020). *Definition of Education in English.* Oxford Online Dictionary. Retrieved from https://www.lexico.com/en/definition/education (accessed April 2020)

Liberatos, P., Link, B. G., & Kelsey, J. L. (1988). The Measurement of Social Class in Epidemiology. *Epidemiologic Reviews, 10*(1), 87–121. doi:10.1093/oxfordjournals.epirev.a036030

Link, B. G., Northridge, M. E., Phelan, J. C., & Ganz, M. L. (1998). Social Epidemiology and the Fundamental Cause Concept: On the Structuring of Effective Cancer Screens by Socioeconomic Status. *The Milbank Quarterly, 76*(3), 375–402. doi:10.1111/1468-0009.00096

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Nonlinear Principal Components Analysis: Introduction and Application. *Psychological Methods, 12*(3), 336–358. doi:10.1037/1082-989X.12.3.336

Losi, L., Maineri, A., Luijkx, R., Schneider, S. L., & Ortmanns, V. (2019). *Measuring Education in the ESS and EVS.* ESS Blog. Retrieved from: https://www.europeansocialsurvey.org/findings/singleblog.html?a=/findings/blog/essblog0013.html (accessed April 2020)

Lucas, S. R. (2001). Effectively Maintained Inequality : Education Transitions , Track Mobility , and Social Background Effects. *American Journal of Sociology*, *106*(6), 1642–1690. doi:10.1086/321300

Lugtig, P., Lensvelt-Mulders, G. J. L. M., Frerichs, R., & Greven, A. (2011). Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey. *International Journal of Market Research*, *53*(5), 669–686. doi:10.2501/IJMR-53-5-669-686

Mare, R. D. (1980). Social Background and School Continuation Decisions. *Journal of the American Statistical Association*, *75*(370), 295–305. doi:10.2307/2287448

Mare, R. D. (1981). Change and Stability in Educational Stratification. *American Sociological Review*, *46*(1), 72–87. doi:10.2307/2095027

Marks, G. N. (2011). Issues in the Conceptualisation and Measurement of Socioeconomic Background: Do Different Measures Generate Different Conclusions? *Social Indicators Research*, *104*, 225–251. doi:10.1007/s11205-010-9741-1

Marks, G. N. (2014). *Education, Social Background and Cognitive Ability. The Deciline of the Social.* Oxfordshire: Routledge.

Marsden, D. (1990). Institutions and Labour Mobility: Occupational and Internal Labour Markets in Britain, France, Italy and West Germany. In R. Brunetta & C. Dell'Aringa (Eds.), *Labour Relations and Economic Performance.* (pp. 414–438). London: Palgrave Macmillan.

Marsden, D., & Germe, J.-F. (1991). Young People and Entry Paths to Long-term Jobs in France and Great Britain. In P. Ryan, P. Garonna, & R. C. Edwards (Eds.), *The Problem of Youth. The Regulation of Youth Employment and Training in Advanced Economies* (Vol. 179–199). Houndmills, U.K.: Macmillan Academic and Professional Ltd.

Massing, N., & Schneider, S. L. (2017). Degrees of Competency: The Relationship Between Educational Qualifications and Adult Skills Across Countries. *Large-Scale Assessments in Education*, *5*(6), 1–34. doi:10.1186/s40536-017-0041-y

Mesch, G. S. (2003). Language Proficiency among New Immigrants: The Role of Human Capital and Societal Conditions. The Case of Immigrants from the FSU in Israel. *Sociological Perspectives*, *46*(1), 41–58. doi:10.1525/sop.2003.46.1.41

Meulman, J. J., van der Kooij, A. J., & Heiser, W. J. (2004). Principal Components Analysis With Nonlinear Optimal Scaling Transformations for Ordinal and Nominal Data. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 49–70). Thousand Oaks: SAGE Publications.

Miech, R. A., & Hauser, R. M. (2001). Socioeconomic Status and Health at Midlife. A Comparison of Educational Attainment with Occupation-Based Indicators. *Annals of Epidemiology*, *11*(2), 75–84. doi:10.1016/s1047-2797(00)00079-x

Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. Cambridge: National Bureau of Economic Research.

Müller, W., & Karle, W. (1993). Social Selection in Educational Systems in Europe. *European Sociological Review*, *9*(1), 1–23. doi:10.1093/oxfordjournals.esr.a036652

Nash, R. (2006). Controlling for "Ability": A Conceptual and Empirical Study of

Primary and Secondary Effects. *British Journal of Sociology of Education*, *27*(2), 157–172. doi:10.1080/01425690600556149

Neugebauer, M. (2010). Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. *Zeitschrift für Soziologie*, *39*(3), 202–214. doi:10.1515/zfsoz-2010-0303

OECD. (1999). *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries - 1999 Edition.* Paris: OECD Publication Service. Retrieved from: http://www.oecd.org/education/skills-beyond-school/1962350.pdf (accessed April 2020)

OECD, European Union, & UNESCO-UIS. (2015). *ISCED 2011 Operational Manual. Guidelines for Classifying National Education Programmes and Related Qualifications*. Retrieved from: http://uis.unesco.org/sites/default/files/documents/isced-2011-operational-manual-guidelines-for-classifying-national-education-programmes-and-related-qualifications-2015-en_1.pdf (accessed April 2020)

Pardos-Prado, S., & Cano, F. (2012). Education and Media Exposure Across Systemic Levels of Conflict. *International Journal of Public Opinion Research*, *24*(2), 121–140. doi:10.1093/ijpor/edr005

Peytcheva, E., & Groves, R. M. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. *Journal of Official Statistics*, *25*(2), 193–201.

Przeworski, A., & Teune, H. (1970). *Logic of Comparative Social Inquiry*. *Comparative Studies in Behavioral Science*. New York: John Wiley & Sons.

Raithel, J., Dollinger, B., & Hörmann, G. (2009). *Einführung Pädagogik. Begriffe, Strömungen, Klassiker, Fachrichtungen.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Rammstedt, B. (2010). Reliabilität, Validität, Objektivität. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 239–258). Wiesbaden: VS Verlag für Sozialwissenschaften, Springer Fachmedien.

RatSWD - Quality Standard Working Group. (2015). *Quality Standards for the Development, Application, and Evaluation of Measurement Instruments in Social Science Survey Research.* RatSWD Workin Paper No. 245. Berlin: RatSWD. Retrieved from: https://www.ratswd.de/dl/RatSWD_WP_245.pdf (accessed April 2020)

Rogelberg, S. G., & Luong, A. (1998). Nonresponse to Mailed Surveys. A Review and Guide. *Current Directions in Psychological Science*, *7*(2), 60–65. doi:10.1111/1467-8721.ep13175675

Ross, C. E., & Wu, C. (1995). The Links Between Education and Health. *American Sociological Review*, *60*(5), 719–745. doi:10.2307/2096319

Rotman, A., Shavit, Y., & Shalev, M. (2016). Nominal and Positional Perspectives on Educational Stratification in Israel. *Research in Social Stratification and Mobility*, *43*, 17–24. doi:10.1016/j.rssm.2015.06.001

Rüber, I. E., & Bol, T. (2017). Informal Learning and Labour Market Returns. Evidence from German Panel Data. *European Sociological Review*, *33*(6), 765–778. doi:10.1093/esr/jcx075

Schaub, H., & Zenke, K. G. (2007). *Wörterbuch Pädagogik* (6th ed.). Munich: Deutscher Taschenbuch Verlag.

Schermelleh-Engel, K., & Werner, C. S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 119–141). Berlin, Heidelberg: Springer-Verlag.

Schneider, Briceno-Rosas, R., Ortmanns, V., & Herzing, J. M. E. (2018). Measuring Migrants' Educational Attainment: The CAMCES Tool in the IAB-SOEP Migration Samples. In D. Behr (Ed.), *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues* (pp. 50–74). Cologne: GESIS - Leibniz-Institute for the Social Sciences, GESIS-Schriftenreihe, 19. Retrieved from: https://www.ssoar.info/ssoar/handle/document/58074 (accessed April 2020).

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). Oxford: University of Oxford, Nuffield College. Retrieved from: http://ora.ouls.ox.ac.uk/objects/uuid:15c39d54-f896-425b-aaa8-93ba5bf03529 (accessed April 2020).

Schneider, S. L. (2010). Nominal Comparability is not Enough: (In-)equivalence of Construct Validity of Cross-National Measures of Educational Attainment in the European Social Survey. *Research in Social Stratification and Mobility*, *28*(3), 343–357. doi:10.1016/j.rssm.2010.03.001

Schneider, S. L. (2013). The International Standard Classification of Education 2011. In G. E. Birkelund (Ed.), *Class and Stratification Analysis. Comparative Social Research.* (pp. 365–379). Bingley: Emerald Group Publishing.

Schneider, S. L. (2016). *The Conceptualisation, Measurement, and Coding of Education in German and Cross-National Surveys*. GESIS Survey Guidelines. Mannheim: GESIS-Leibniz Institute for the Social Sciences. doi:10.15465/gesis-sg_en_020

Schneider, S. L. (2018a). Das Bildungsniveau von Migrantinnen und Migranten: Herausforderung in Erfassung und Vergleich. In D. B. Maehler, A. Shajek, & H. U. Brinkmann (Eds.), *Diagnostik bei Migrantinnen und Migranten - Ein Handbuch* (pp. 47–56). Göttingen: Hogrefe Verlag.

Schneider, S. L. (2018b). Education in OECD's PIAAC Study: How Well do Different Harmonized Measures Predict Skills? *Methods, Data, Analysis*, *12*(1), 151–176. doi:10.12758/mda.2017.15

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. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (pp. 288–307). London: Sage Publications.

Schneider, S. L., & Ortmanns, V. (2019). *Database of Educational Attainment, With Explanatory Note*. Deliverable 8.8 of the SERISS project funded under the European Union's Horizon 2020 research and innovation programme GA No:

654221. Retrieved from: www.seriss.eu/resources/deliverables (accessed April 2020).

Schröder, H., & Ganzeboom, H. B. G. (2014). Measuring and Modelling Level of Education in European Societies. *European Sociological Review*, *30*(1), 119–136. doi:10.1093/esr/jct026

Semyonov, M., Raijman, R., & Gorodzeisky, A. (2008). Foreigners ’ Impact on European Societies: Public Views and Perceptions in a Cross-National Comparative Perspective. *International Journal of Comparative Sociology*, *49*(1), 5–29. doi:10.1177/0020715207088585

Sewell, W. H., Haller, A. O., & Portes, A. (1969). The Educational and Early Occupational Attainment Process. *American Sociological Review*, *34*(1), 82–92. doi:10.2307/2092789

Shavit, Y., & Blossfeld, H.-P. (1993). *Persistent Inequality. Changing Educational Attainment in Thirteen Countries*. Boulder: Westview Press.

Shavit, Y., & Müller, W. (1998). *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*. Oxford: Oxford University Press.

Shavit, Y., & Park, H. (2016). Introduction to the Special Issue: Education as a Positional Good. *Research in Social Stratification and Mobility*, *43*, 1–3. doi:10.1016/j.rssm.2016.03.003

Smith, H. L., & Garnier, M. A. (1986). Scaling via Models for the Analysis of Association: Social Background and Educational Careers in France. *Sociological Methodology*, *17*, 205–245. doi:10.2307/271034

Smith, T. W. (1995). Some Aspects of Measuring Education. *Social Science Research*, *24*(3), 215–242. doi:10.1006/ssre.1995.1008

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374. doi:10.2307/1882010

Stevens, G. (1999). Age at Immigration and Second Language Proficiency among Foreign-Born Adults. *Language in Society*, *28*(4), 555–578. doi:10.1017/S0047404599004030

Tam, T. (2016). Academic Achievement as Status Competition: Intergenerational Transmission of Positional Advantage Among Taiwanese and American Students. *Chinese Journal of Sociology*, *2*(2), 171–193. doi:10.1177/2057150X16638600

Thurow, L. C. (1975). *Generating Inequality. Mechanisms of Distribution in the U.S. Economy*. London: Macmillan Press.

Treiman, D. J., & Terrell, K. (1975). The Process of Status Attainment in the United States and Great Britain. *American Journal of Sociology*, *81*(3), 563–583.

Triventi, M., Panichella, N., Ballarino, G., Barone, C., & Bernardi, F. (2016). Education as a Positional Good: Implications for Social Inequalities in Educational Attainment in Italy. *Research in Social Stratification and Mobility*, *43*, 39–52. doi:10.1016/j.rssm.2015.04.002

Ultee, W. C. (1980). Is Education a Positional Good? An Empirical Examination of Alternative Hypotheses on the Connection Between Education and Occupational Level. *The Netherlands’ Journal of Sociology*, *16*, 135–153.

UNESCO-UIS. (2006). *International Standard Classification of Education ISCED 1997*. Montreal: UNESCO-UIS. Retrieved from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf (accessed April 2020)

UNESCO-UIS. (2012). *International Standard Classification of Education 2011*. Montreal: UNESCO-UIS. Retrieved from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf (accessed April 2020)

van Tubergen, F. (2010). Determinants of Second Language Proficiency among Refugees in the Netherlands. *Social Forces*, *89*(2), 515–534. doi:10.1353/sof.2010.0092

van Tubergen, F., & Wierenga, M. (2011). The Language Acquisition of Male Immigrants in a Multilingual Destination: Turks and Moroccans in Belgium. *Journal of Ethnic and Migration Studies*, *37*(7), 1039–1057. doi:10.1080/1369183X.2011.572476

Weakliem, D. L. (2002). The Effects of Education on Political Opinions: An International Study. *International Journal of Public Opinion Research*, *14*(2), 141–157. doi:10.1093/ijpor/14.2.141

Weins, C. (2010). Kompetenzen Oder Zertifikate? Die Entwertung Ausländischer Bildungsabschlüsse auf dem Schweizer Arbeitsmarkt. *Zeitschrift Fur Soziologie*, *39*(2), 124–139. doi:10.1515/zfsoz-2010-0203

Will, G., Gentile, R., Heinritz, F., & von Maurice, J. (2018). *REGES - Refugees in the German Educational System: Überblick über Forschungsdesing, Strichprobenziehung und Ausschöpfung der ersten Welle.* LIfBI Working Paper No. 75. Bamberg: Leibniz Institute for Edcuational Trajectories. Retrieved from: https://www.neps-data.de/Portals/0/Working Papers/WP_LXXV.pdf (accessed April 2020)

Wolbers, M. H. J., de Graaf, P. M., & Ultee, W. C. (2001). Trends in the Occupational Returns to Educational Credentials in the Dutch Labor Market: Changes in Structures and in the Association? *Acta Sociologica*, *44*(1), 5–19. doi:10.1177/000169930104400102

Paper I

Harmonization Still Failing? Inconsistency of Education Variables in Cross-National Public Opinion Surveys

2 Harmonization Still Failing? Inconsistency of Education Variables in Cross-National Public Opinion Surveys[2]

2.1 Introduction

During recent decades, cross-national comparative research in public opinion has grown tremendously, both in quantity and quality. Through the increased availability of various types of international public opinion survey data, many research questions can today be tackled from a comparative point of view. This allows researchers to test the generality of hypotheses, as well as contextual effects that may explain why countries differ the way they do (Przeworski & Teune, 1970).

The credibility of comparative studies, however, hinges on the cross-national comparability of the data they are based on. This is a matter of continuous debate among comparative survey researchers and methodologists (e.g., Heath, Martin, & Spreckelsen, 2009; Hoffmeyer-Zlotnik & Warner, 2014). Consistency across data sources is a necessary condition of comparability. If data are not coded consistently across time points and surveys, they are not comparable across countries. Consistency across data sources is important because it allows researchers to compare results from different studies and pool different data sources for analysis.

The aim of this study is to evaluate the consistency of four cross-national public opinion survey data sets for a large number of European countries. We chose the variable highest level of education for this purpose. This measure is of special interest for two reasons: Firstly, it is one of the most widely used variables in public opinion research (see Smith, 1995) because it is a 'core' variable reflecting socialization, social stratification, and individual life chances. From numerous studies we know that, across countries, educational attainment substantially correlates with attitudes, beliefs, values, and behaviors (e.g., Bekhuis, Lubbers, & Verkuyten, 2014; Kalmijn, 2003; Weakliem, 2002). Secondly, educational attainment is one of the most difficult background variables to measure and code in a comparable and coherent way across countries and

[2] This paper is published as Research Note: Ortmanns, V. & Schneider, S. L. (2016). Harmonization Still Failing ? Inconsistency of Education Variables in Cross-National Public Opinion Surveys. *International Journal of Public Opinion Research,* 28(4), 562–82. doi:10.1093/ijpor/edv025

studies (see Braun & Müller, 1997; Kerckhoff & Dylan, 1999; Schneider, 2009; Schröder & Ganzeboom, 2013).

Until recently, little research involving only small numbers of countries and surveys was available with respect to the comparable measurement of educational attainment (e.g., Braun & Müller, 1997; Kerckhoff, Ezell, & Brown, 2002; Smith, 1995). Regarding the more specific issue of incoherence of educational attainment data across data sources, the distribution of education coded using the International Standard Classification of Education (ISCED) was shown to be inconsistent across time and surveys. Hoffmeyer-Zlotnik (2008) compares the distributions of ISCED in the European Social Survey (ESS) 2002 with (unspecified) Eurostat data for Austria, Denmark, France, and Spain. He finds discrepancies, and he explains them by inconsistent coding of country-specific education categories into ISCED across data sources. Kieffer (2010) compares data from the French ESS 2002–2008 with the French Labour Force Survey from corresponding years, confirming the results of Hoffmeyer-Zlotnik, while adding detailed explanations for these inconsistencies for France. Schneider (2009) compares official data from the European Labour Force Survey and the European Survey of Income and Living Conditions with data from the ESS for the years 2002–2006 for 26 European countries. In this most comprehensive analysis of education data inconsistencies to date, she reveals inconsistencies over time, surveys, and countries.

The research question of this study is: Do we find the same inconsistencies of educational attainment data across time and surveys for widely used, recent, cross-national public opinion surveys that have not yet been evaluated in this respect? This study thus adds to existing research by analyzing the ESS, the International Social Survey Programme (ISSP), the European Values Study (EVS), and the Eurobarometer. We compare data from 2008 to 2012.

In the next part of the article, the methodological background is summarized by reviewing the process of ex-ante output harmonization and the ISCED, which is widely used for the harmonization of educational attainment data. In the third section of the article, the data sources, analysis strategy, education coding, and the indicator for coherence across data sets, are introduced. The results are then presented, followed by a discussion of the findings and conclusions.

2.2 Methodological Background

2.2.1 Ex-Ante Harmonization of Education Categories in Cross-National Surveys

Educational systems, with their idiosyncratic institutions and certificates, differ substantially across countries. Response categories for the question ''what is your highest level of education?'' include many proper names rather than generic descriptions that are universally understood. Thus, they cannot be translated. Educational attainment can only be measured using country-specific response categories in survey questionnaires. Respondents choose their attainment from a list of country-specific educational qualifications or levels. For cross-national comparisons, these qualifications have to be coded into an internationally comparable education variable, a process called output harmonization.

Comparative survey designers normally plan how to make variables comparable across countries in advance of the survey. This process is called ex-ante output harmonization (Ehling, 2003). It requires specifying a cross-national coding framework and the mapping of country-specific survey responses to this framework in the survey design phase. The country-specific data collection instrument has to be developed with the cross-national coding framework in mind because the latter implies the kinds of distinctions required for cross-national comparison.

While most comparative surveys aim at ex-ante output harmonization for the education variable, they do not coordinate this process with each other. Country-specific questionnaire items and cross-national coding frameworks, as well as the relationship between the two and their documentation, differ across surveys.

2.2.2 The ISCED 1997

The most commonly used cross-national coding framework for ex-ante harmonization of education data in surveys is the ISCED. ISCED was designed by UNESCO in the 1970s and revised in 1997 and 2011. The aim of ISCED is ''to serve as an instrument suitable for assembling, compiling and presenting comparable indicators and statistics of education both within individual countries and internationally'' (UNESCO-UIS, 1997 [2006], p. 7). It defines comparable education categories applicable around the world. Because the newest ISCED version has not yet been widely implemented, this study refers to ISCED 1997. ISCED 1997 consists of seven main levels:

- ISCED 0: Pre-primary education (or not completed primary education)
- ISCED 1: Primary education or first stage of basic education
- ISCED 2: Lower secondary or second stage of basic education
- ISCED 3: Upper secondary education
- ISCED 4: Post-secondary non-tertiary education
- ISCED 5: First stage of tertiary education
- ISCED 6: Second stage of tertiary education

Dimensions of differentiation within education levels such as vocational, pre-vocational, and general education, as well as whether a qualification allows access to a higher level of education, are not implemented in most surveys and thus are not used in this study, despite their known importance for predicting, for example, labor market outcomes (e.g., Schneider, 2010).

2.3 Data and Method

2.3.1 Comparative Survey Data

We analyze data from four cross-national public opinion surveys that have not yet been examined with respect to education variable consistency. In the early 1970s, the Eurobarometer program was launched by the European Commission. ISCED 1997 main levels were implemented in three Standard and Special Eurobarometer studies in 2010 and 2011. From the late 1970s onward, the academically driven public opinion research programs EVS and ISSP were established. The EVS 2008 contains three ISCED variables. One reflects the seven main ISCED 1997 levels and is used in this study. The ISSP changed its harmonized education variable in 2011 from a nonstandard education scheme to a scheme closely related to but not identical with ISCED 1997. The ESS was launched in the early 2000s with an ISCED main level variable that was later corrected to a five-level version and introduced a detailed cross-national education variable closely related to ISCED 2011 in 2010. Respondents aged 25–64 are selected so that samples are as comparable as possible.

2.3.2 Analysis Strategy

We check the consistency of European public opinion data with respect to two dimensions: consistency over time and consistency across surveys. Firstly, we check whether education distributions differ within surveys and countries over time. This can be done for ESS, ISSP, and Eurobarometer data because, for these studies at least two rounds are available for this time period. Especially with respect to the education measurement changes in ESS 2010 and ISSP 2011, a detailed look at those data over time is useful for identifying and evaluating the effects of those changes. For comparison over time in the ESS, we also include ESS 2002, 2004, and 2006 because these data have been corrected since the inconsistencies were first presented (Hoffmeyer-Zlotnik, 2008; Kieffer, 2010; Schneider, 2009).

Secondly, we compare the distribution of the harmonized education variable within time points and countries across surveys. For these comparisons, we can also include the EVS. ESS data from 2010 or 2012 are used as a benchmark because, owing to the large-scale revision effort to improve cross-country consistency, as well as the validity of its education measurement, it can be expected to provide high-quality education variables. Because ISSP did not use ISCED for education coding before 2011, we only use years 2011 and 2012 for the comparison of ISSP and ESS.

2.3.3 Education Coding

The education variables of EVS, Eurobarometer, and ESS 2010–2012 are, or can be, coded into the seven main ISCED 1997 levels for comparisons over time and across surveys (see Table 2.1). The variables for ESS rounds 2002, 2004, 2006, and 2008 as well as the ISSP education variable, however, have to be treated separately. For the ESS, we aggregate ISCED categories 0 and 1 and categories 5 and 6 for the analysis over time. In the ISSP, we create a four-category measure to render the data before and after the changes in 2011 comparable (see Table 2.2). For the cross-survey comparison, the detailed education variable in ESS 2010 and 2012 is coded to fit the new ISSP education variable with seven categories.

Table 2.1 Categories and recodes of the education variables into ISCED 97 for cross-survey comparison

ISCED 97	Values		ESS since 2010		EVS		EB
0	Pre-primary education (or not completed primary education)	0	Not completed ISCED level 1	0	Pre-primary education or none education	0	Pre-primary education
1	Primary education or first stage of basic education	113	ISCED 1, completed primary education	1	Primary education or first stage of basic education	1	Primary education or first stage of basic education
		129	Vocational ISCED 2C < 2 years, no access ISCED 3				
2	Lower secondary or second stage of basic education	212	General/pre-vocational ISCED 2A/2B, access ISCED3 vocational	2	Lower secondary or second stage of basic education	2	Lower secondary or second stage of basic education
		213	General ISCED 2A, access ISCED 3A general/all 3				
		221	Vocational ISCED 2C >= 2 years, no access ISCED 3				
		222	Vocational ISCED 2A/2B, access ISCED 3 vocational				
		223	Vocational ISCED 2, access ISCED 3 general/all				
		229	Vocational ISCED 3C < 2 years, no access ISCED 5				
3	(Upper) Secondary education	311	General ISCED 3 >=2 years, no access ISCED 5	3	(Upper) secondary education	3	(Upper) secondary education
		321	Vocational ISCED 3C >= 2 years, no access ISCED 5				
		312	General ISCED 3A/3B, access ISCED 5B/lower tier 5A				
		322	Vocational ISCED 3A/3B, access 5B/lower tier 5A				
		313	General ISCED 3A, access upper tier ISCED 5A/all 5				
		323	Vocational ISCED 3A, access upper tier ISCED 5A/all 5				
4	Post-secondary non-tertiary education	421	ISCED 4 without access ISCED 5	4	Post-secondary non-tertiary education	4	Post-secondary, non-tertiary education
		412	General ISCED 4A/4B, access ISCED 5B/lower tertiary 5A				
		413	General ISCED 4A, access upper tier ISCED 5A/all 5				
		422	Vocational ISCED 4A/4B, access ISCED 5B/lower tertiary 5A				
		423	Vocational ISCED 4A, access upper tier ISCED 5A /all 5				

ISCED 97	Values		ESS since 2010		EVS		EB
5	First stage of tertiary education	510	ISCED 5A short, intermediate/academic/general tertiary below	5	First stage of tertiary education	5	First stage of tertiary education
		520	ISCED 5B short, advanced vocational qualifications				
		610	ISCED 5A medium, bachelor/equivalent from lower tertiary				
		620	ISCED 5A medium, bachelor/equivalent from upper/single tertiary				
		710	ISCED 5A long, master/equivalent from lower tertiary				
		720	ISCED 5A long, master/equivalent from upper/single tertiary				
6	Second stage of tertiary education	800	ISCED 6, doctoral degree	6	Second stage of tertiary education	6	Second stage of tertiary education

Data sources:
Eurobarometer 73.2 (February-March 2010) files from Eurostat, data file versions 2.0.1, variable v362;
Eurobarometer 73.3 (March-April 2010) files from Eurostat, data file versions 2.0.1, variable v362;
ESS 2010-2012, data file versions: 3.0 (2010), 2.0 (2012) variable edulvlb;
EVS 2008, data file version 3.0.0, variable v336

Table 2.2 Categories and relationship between the education variables in ISSP 2008-2012 and ESS 2010 for cross-time and cross-survey and time comparison

	ISSP over time		ISSP until 2010		ISSP since 2011		ESS since 2010
1	Less than upper secondary qualification	0	No formal qualification	0	No formal education	0	Not completed ISCED level 1
				1	Primary school	113	ISCED 1, completed primary education
		1	Lowest formal qualification			129	Vocational ISCED 2C < 2 years, no access ISCED 3
				2	Lower secondary (secondary education completed that does not allow entry to university: end of obligatory school but also short programs (less than 2 years))	212	General/pre-vocational ISCED 2A/2B, access ISCED3 vocational
						213	General ISCED 2A, access ISCED 3A general/all 3
						221	Vocational ISCED 2C >= 2 years, no access ISCED 3
						222	Vocational ISCED 2A/2B, access ISCED 3 vocational
						223	Vocational ISCED 2, access ISCED 3 general/all
						229	Vocational ISCED 3C < 2 years, no access ISCED 5
2	University entrance qualification	3	Higher secondary completed	3	Upper secondary (programs that allow entry to university)	313	General ISCED 3A, access upper tier ISCED 5A/all 5
						323	Vocational ISCED 3A, access upper tier ISCED 5A/all 5
						413	General ISCED 4A, access upper tier ISCED 5A/all 5
						423	Vocational ISCED 4A, access upper tier ISCED 5A /all 5
3	Upper and post-secondary non-tertiary labour market preparatory qualification	2	Above lowest qualification	4	Post secondary, non-tertiary (other upper secondary programs toward the labour market or technical formation)	311	General ISCED 3 >=2 years, no access ISCED 5
						312	General ISCED 3A/3B, access ISCED 5B/lower tier 5A
						321	Vocational ISCED 3C >= 2 years, no access ISCED 5
						322	Vocational ISCED 3A/3B, access 5B/lower tier 5A
						412	General ISCED 4A/4B, access ISCED 5B/lower tier 5A
						421	ISCED 4 without access ISCED 5
						422	Vocational ISCED 4A/4B, access ISCED 5B/lower tertiary 5A

	ISSP over time		ISSP until 2010		ISSP since 2011		ESS since 2010
4	Tertiary qualification	4	Above higher secondary level, other qualification	5	Lower level tertiary, first stage (also technical schools at a tertiary level)	510	ISCED 5A short, intermediate/academic/general tertiary below Bachelor level
						520	ISCED 5B short, advanced vocational qualifications
		5	University degree completed			610	ISCED 5A medium, bachelor/equivalent from lower tier tertiary
						620	ISCED 5A medium, bachelor/equivalent from upper/single tier tertiary
				6	Upper level tertiary (Master, Dr.)	710	ISCED 5A long, master/equivalent from lower tier tertiary
						720	ISCED 5A long, master/equivalent from upper/single tier tertiary
						800	ISCED 6, doctoral degree

Data sources:
ESS 2010-2012, data file versions: 3.0. (2010), 2.0 (2012), variable edulvlb;
ISSP 2008-2012, data file versions: 2.2.0 (2008), 3.0.0 (2009), 2.0.0 (2010), 2.0.0 (2011), 2.0.0 (2012), variable DEGREE

2.3.4 Duncan's Dissimilarity Index

For comparing the education distributions and measuring inconsistencies, Duncan's dissimilarity index is used (Duncan & Duncan, 1955) as a summary measure consistent with Schneider (2009). The index was originally developed for measuring residential segregation, but it can be generalized to measure differences in the distributions of categorical variables. The index is rescaled to range from 0 to 100. The dissimilarity index can then be interpreted as the percentage of cases that would have to change categories to achieve equal distributions across two data sources. Formally, if x_i denotes the size of category *i* of *k* ISCED categories for country A in year B in survey S in percent and y_i denotes the same for country A in year B in survey T, the dissimilarity index is defined as D = ½ $\sum_{i=1}^{k} |x_i - y_i|$. The equivalent holds when survey rounds are compared rather than surveys.

An example of the calculation is given in Table 2.3: When comparing the distribution of the harmonized education variable for Spain in EVS 2008 and ESS 2010, we first calculate the absolute difference between data sources for each ISCED category. These differences are summed across ISCED categories and divided by two. The larger the differences between individual ISCED categories across surveys, the larger Duncan's index.

Table 2.3 Example of calculation Duncan's dissimilarity index for Spain

	EVS 2008	**ESS 2010**	**Absolute difference between EVS and ESS**
ISCED 0	9.7	6.1	3.60
ISCED 1	14.8	17.1	2.33
ISCED 2	16.7	26.2	9.43
ISCED 3	19.1	13.6	5.52
ISCED 4	17.4	6.3	11.03
ISCED 5	19.4	29.1	9.65
ISCED 6	3.0	1.7	1.26
Sum			42.83
Duncan's Index			**21.41**

Data sources: ESS 2010, data file version: 3.0, variable edulvlb, weighted using dweight. N=1276.
EVS 2008, data file version 3.0.0, variable v336. N=974.
Sample selection: Only respondents aged 25-64.

2.4 Results

2.4.1 Inconsistencies Over Time

Figure 2.1 shows Duncan's dissimilarity index for comparisons of education distributions within surveys and countries over time (successive years or survey rounds), averaged across those countries present in each round of the survey in question. Table 2.4 shows the index for all countries individually, including country means. In the three Eurobarometer surveys, the mean value of Duncan's dissimilarity index is around 5% for each of the three comparisons. Comparing the ISSP education data over time, the mean value of Duncan's index is as high as 10% when comparing data from 2008 to 2010, but it reaches even 26% between 2010 and 2011 when the comparative variable was changed. Comparing 2011 and 2012, Duncan's index decreases to 14%. With respect to the ESS, the mean values of Duncan's index are below 7% between 2002 and 2008, and with the measurement changes in 2010, it increases to 11%. In 2012, the mean inconsistency across countries nearly halves to 5%, the lowest value ever achieved for the ESS.

In summary, the distributions over time are fairly stable in the Eurobarometer and the ESS with <10% of cases having to move categories to achieve equal distributions across survey rounds. In both ESS and ISSP, deviations increase with measurement changes and decrease in the year following the measurement changes. However, they decrease more in the ESS than in the ISSP. The mean value across all comparisons over time is D=8.7%. Note that we compare a variable with seven categories for Eurobarometer, whereas in the ESS and ISSP, the variable consists of five and four categories, respectively. The larger the number of categories, the more classification errors can be made, so the stability in the Eurobarometer and the instability of the ISSP are even more astonishing. Note also however that the selection of countries is the same within, but not across, surveys.

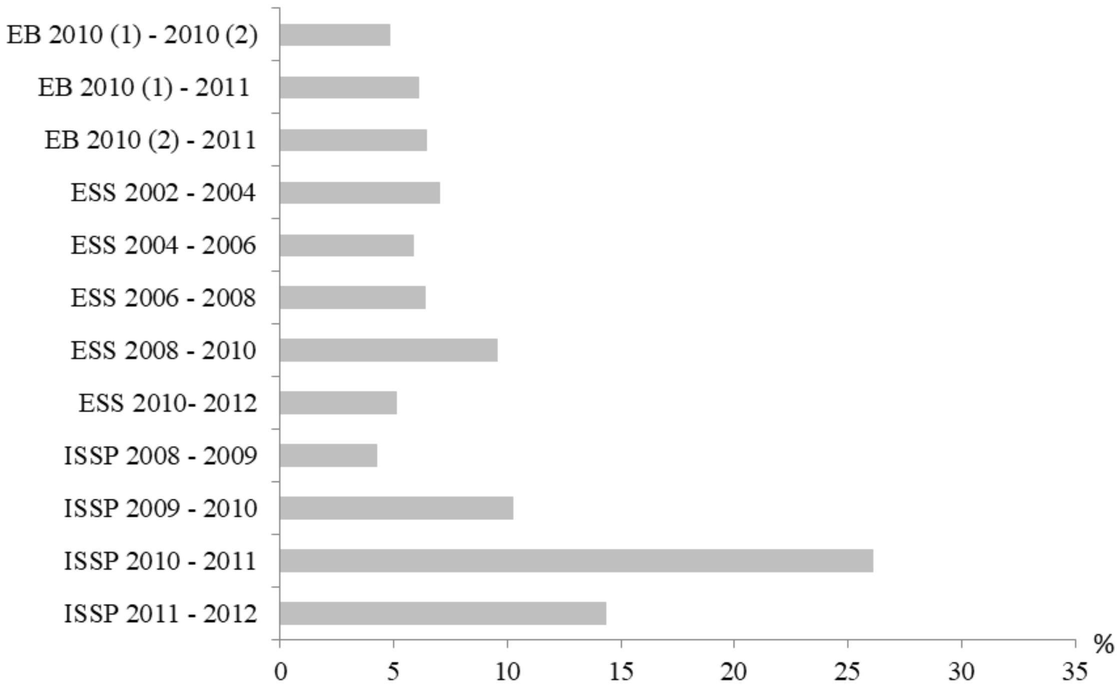


Figure 2.1 Duncan's dissimilarity index for educational attainment distributions over time within ESS, ISSP, and Eurobarometer data averaged across countries

Data sources:
ESS 2002-2008, data file versions: 6.2 (2002), 3.3 (2004), 3.4 (2006), 4.1 (2008), variable edulvla, weighted using dweight;
ESS 2010-2012, data file versions: 3.0. (2010), 2.0 (2012), variable edulvlb, weighted using dweight;
Eurobarometer 73.2 (February-March 2010) files from Eurostat, data file versions 2.0.1, variable v362 Eurobarometer 73.3 (March-April 2010) files from Eurostat, data file versions 2.0.1, variable v362 Eurobarometer 75.4 (2011) files from Eurostat, data file; version 3.0.1, variable v105 All Eurobarometer data weighted to correct regional oversampling for Germany and the UK;
ISSP 2008-2012, data file versions: 2.2.0 (2008), 3.0.0 (2009), 2.0.0 (2010), 2.0.0 (2011), 2.0.0 (2012), variable DEGREE, data weighted to correct regional oversampling for Germany;
Sample selection: Only respondents aged 25-64.
Countries included:
ESS: BE, BG, DE, DK, ES, FI, GB, IE, NL, NO, PL, PT, SE, SI;
Eurobarometer: all 27 EU-member states;
ISSP: CH, CZ, DE, DK, FI, FR, GB (excluding Northern Ireland), HR, NO, RU, SE, SI, SK, TR

Table 2.4 Duncan's dissimilarity index for educational attainment distributions over time for Eurobarometer, ISSP and ESS data

	Eurobarometer			ISSP				ESS					
	EB 2010 (1)-EB 2010 (2)	**EB 2010 (2)-EB 2011**	**EB 2010 (1)-EB 2011**	**08-09**	**09-10**	**10-11**	**11-12**	**02-04**	**04-06**	**06-08**	**08-10**	**10-12**	**mean per country**
AT	3.7	4.7	4.5	16.5				15.7	4.0				8.2
BE	4.0	13.5	11.7	2.4	4.1	13.4		2.2	2.0	3.8	9.6	8.5	6.8
BG	4.4	3.9	1.0		8.7	31.3				5.1	3.1	4.9	7.8
CH					5.8	7.6	4.2	0.8	7.4	3.3	8.2	1.5	4.8
CY	7.5	9.0	13.1	8.6						8.8	7.7	8.9	9.1
CZ	1.8	5.3	4.6	5.8	6.1	42.1	9.7	2.9			14.6	8.6	10.1
DE	10.0	10.8	1.9	2.2	1.9	27.3	4.0	1.8	5.1	8.2	2.7	5.9	6.8
DK	4.7	5.7	5.1	2.2	6.3	27.4		13.8	2.8	4.4	8.5	5.3	7.8
EE	2.8	8.8	6.3						6.8	6.6	11.5	1.7	6.4
ES	6.3	9.3	5.2	5.1	1.6			8.8	7.3	8.1	9.0	5.6	6.6
FI	3.6	3.3	2.4	3.5	27.2	29.6	5.1	6.2	1.9	3.4	13.4	4.5	8.7
FR	3.5	2.5	3.3	5.5	5.8	29.2	25.5	2.6	3.8	8.6	10.1		9.1
GB	6.0	5.3	9.4	2.2	3.9	23.7	22.1	14.0	23.2	3.7	20.0	2.8	11.4
GR	2.8	5.3	5.4					3.5			11.6		5.7
HR					6.4	20.7	2.6				4.1		8.5
HU	4.1	3.5	6.9	1.9				18.2	11.5	5.3	6.0		7.2
IE	3.3	6.7	4.4					4.4	12.6	4.0	15.6	4.3	6.9
IS	3.2												3.2
IT	2.8	3.1	3.6	17.7									6.8
LT	2.7	6.7	4.4			28.7	24.5						13.4
LU	7.9	5.1	9.5					6.2					7.2
LV	2.0	5.3	6.0		1.1								3.6
MT	8.9	7.0	3.7										6.5

	Eurobarometer			ISSP				ESS					
	EB 2010 (1)-EB 2010 (2)	**EB 2010 (2)-EB 2011**	**EB 2010 (1)-EB 2011**	**08-09**	**09-10**	**10-11**	**11-12**	**02-04**	**04-06**	**06-08**	**08-10**	**10-12**	**mean per country**
NL	2.7	6.5	8.6					4.8	3.5	5.5	11.0	5.7	6.0
NO	11.0			5.8	9.8	19.5	5.9	21.9	5.2	7.1	3.5	3.7	9.3
PL	4.9	6.0	3.3				30.1	4.5	2.5	8.2	29.6	2.6	10.2
PT	4.0	2.0	3.1					2.4	3.4	7.7	3.2	7.3	4.1
RO	1.7	7.0	7.5										5.4
RU				5.4	29.3	18.0	46.3			3.9	2.6	11.9	16.8
SE	3.9	3.5	6.0	4.6	1.5	29.7	4.6	1.5	5.0	6.8	15.7	3.3	7.2
SI	6.0	4.2	7.6		2.8	20.9	6.9	4.1	4.8	2.2	3.2	3.5	6.0
SK	1.7	1.0	2.4	3.0		33.5			1.7	8.9	3.8	1.3	6.3
TR				3.5	7.0	15.6	6.0						8.0
UA				2.1					2.5	17.6	8.6		7.7
mean value for all countries	**4.5**	**5.7**	**5.6**	**5.4**	**7.6**	**24.6**	**14.1**	**7.0**	**5.8**	**6.4**	**9.5**	**5.1**	**8.5**
mean value for all countries participating in all rounds per survey	**4.3**	**5.7**	**5.6**	**4.3**	**10.3**	**26.1**	**14.3**	**6.5**	**6.2**	**5.5**	**10.9**	**4.6**	**8.7**

Data sources:
ESS 2002-2008, data file versions: 6.2 (2002), 3.3 (2004), 3.4 (2006), 4.1 (2008), variable edulvla, weighted using dweight;
ESS 2010-2012, data file versions: 3.0 (2010), 2.0 (2012) variable edulvlb, weighted using dweight;
Eurobarometer 73.2 (February-March 2010) files from Eurostat, data file versions 2.0.1, variable v362;
Eurobarometer 73.3 (March-April 2010) files from Eurostat, data file versions 2.0.1, variable v362;
Eurobarometer 75.4 (2011) files from Eurostat, data file; version 3.0.1, variable v105 All Eurobarometer data weighted to correct regional oversampling for Germany and the UK;
ISSP 2008-2012, data file versions: 2.2.0 (2008), 3.0.0 (2009), 2.0.0 (2010), 2.0.0 (2011), 2.0.0 (2012), variable DEGREE, data weighted to correct regional oversampling for Germany; Wallonia excluded for BE, Northern Ireland is excluded for GB
Sample selection: Only respondents aged 25-64.

2.4.2 Inconsistencies Across Surveys

Figure 2.2 shows Duncan's dissimilarity index for comparisons within countries and time points across the four surveys, averaged across the 12 countries that participated in all surveys and waves. Table 2.5 shows the index for all countries individually, including country means. As described above, ESS data from 2010 and 2012 are used as a benchmark. With respect to the comparison of ESS 2010 and EVS 2008, the mean value of Duncan's index across countries is 17%. Comparing the ESS and the Eurobarometer, mean values of Duncan's index across countries similarly amount to around 18% in all three comparisons. Comparing the ESS and the ISSP, the resulting mean inconsistency of the education distributions, as indicated by Duncan's dissimilarity index, amounts to 27% in 2011 and 31% in 2012.

To summarize, all surveys produce somewhat different education distributions than the ESS. The lowest average dissimilarity can be observed between the ESS and EVS and between the ESS and Eurobarometer (D<20%). A high level of inconsistency is identified between ESS and ISSP, despite the fact that the coding between the two education variables was carefully adjusted for this study. The mean value across the 12 countries is D=22%. Discrepancies across surveys are thus more than twice as high as inconsistencies over time (this is the case also when only looking at the six countries included in all comparisons).

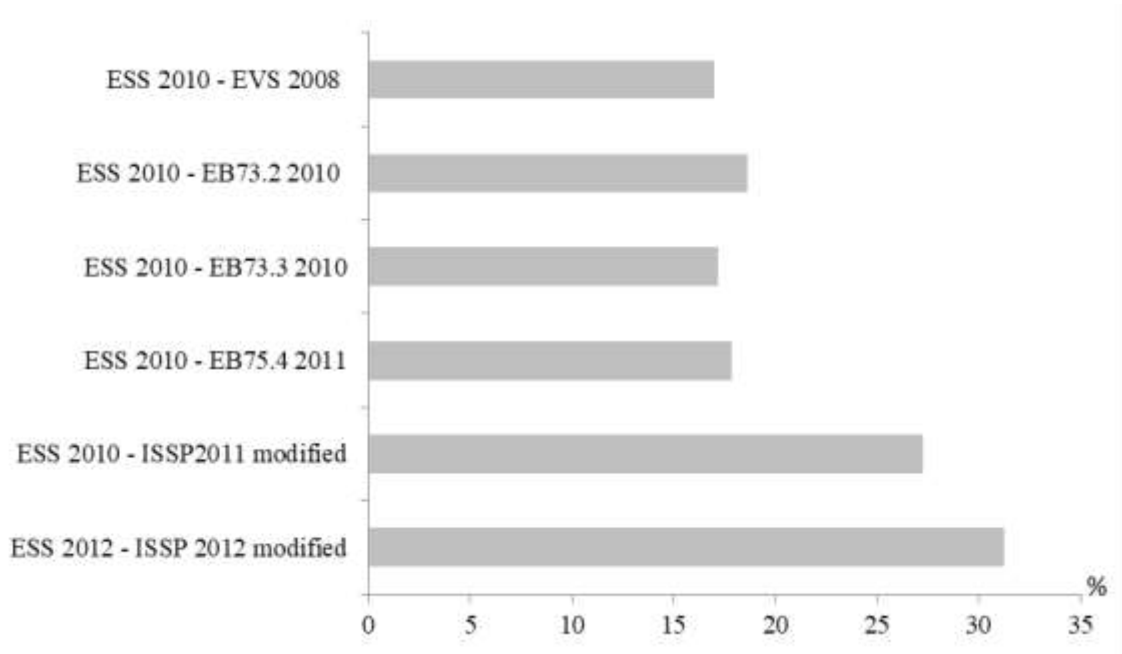


Figure 2.2 Duncan's dissimilarity index for educational attainment distributions comparing EVS, ISSP, and Eurobarometer with ESS averaged across countries

Data sources:
ESS 2010-2012, data file versions: 3.0 (2010), 2.0 (2012) variable edulvlb, weighted using dweight;
EVS 2008, data file version 3.0.0, variable v336; data weighted to correct regional oversampling for Germany, Belgium, and the UK;
Eurobarometer 73.2 (February-March 2010) files from Eurostat, data file versions 2.0.1, variable v362 Eurobarometer 73.3 (March-April 2010) files from Eurostat, data file versions 2.0.1, variable v362 Eurobarometer 75.4 (2011) files from Eurostat, data file; version 3.0.1, variable v105 All Eurobarometer data weighted to correct regional oversampling for Germany and the UK;
ISSP 2011-2012, data file versions: 2.0.0 (2011), 2.0.0 (2012); variable DEGREE, data weighted to correct regional oversampling for Germany
Sample selection: Only respondents aged 25-64.
Countries included: BG, CZ, DE, DK, FI, FR, GB (excluding Northern Ireland in ISSP comparisons), LT, PL, SE, SI, SK

Table 2.5 Duncan's dissimilarity index for educational attainment distributions comparing EVS, ISSP, and Eurobarometer with ESS

	ESS 2010 - EVS 2008	ESS 2010 - EB 2010 (1)	ESS 2010 - EB2010 (2)	ESS 2010 - EB 2011	ESS 2010 - ISSP 2011	ESS 2012 - ISSP 2012	mean per country
BE	11.9	21.6	22.6	11.1	7.9		15.0
BG	4.9	10.9	7.1	9.8	38.5	36.7	18.0
CH	17.0				11.7	9.4	12.7
CY	15.7	12.8	20.1	22.1			17.7
CZ	10.5	16.0	17.2	13.9	45.4	39.2	23.7
DE	17.3	29.6	24.1	28.7	14.8	15.2	21.6
DK	10.8	17.1	15.1	20.8	17.9	13.8	15.9
EE	32.2	12.7	15.3	7.0			16.8
ES	21.4	13.1	12.8	15.2		15.0	15.5
FI	15.7	6.8	5.6	7.6	28.5	25.6	15.0
FR	6.4	12.0	12.6	12.3	23.7	42.4	18.2
GB	32.9	32.0	28.7	27.8	20.5	20.1	27.0
GR	6.8	3.1	4.9	6.8			5.4
HR	2.7				28.1		15.4
HU	1.5	46.2	42.2	39.6		32.6	32.4
IE	12.7	21.2	21.2	22.5		24.3	20.4
IS						6.8	6.8
LT	3.5	29.8	29.8	28.3	27.8	48.8	28.0
NL	13.6	37.5	39.0	41.0	18.0		29.8
NO	14.1	16.3	9.3		44.0	44.4	25.6
PL	41.1	36.6	36.1	34.6	18.5	51.1	36.3
PT	2.9	7.7	5.6	10.4	11.4		7.6
RU	27.3				32.7	50.0	36.6
SE	16.2	11.9	14.2	16.8	32.8	39.5	21.9
SI	30.0	9.2	4.0	2.9	23.4	24.1	15.6
SK	14.2	11.6	11.1	10.3	34.6	34.3	19.4
UA	28.0						28.0
mean value	**15.8**	**18.9**	**18.1**	**18.5**	**25.3**	**30.2**	**20.2**
mean value for countries participating in all comparisons	**16.9**	**18.6**	**17.1**	**17.8**	**27.2**	**32.6**	**21.7**

Data sources:
ESS 2010-2012, data file versions: 3.0 (2010), 2.0 (2012); variable edulvlb, weighted using dweight. For comparison with ISSP Wallonia in BE and Northern Ireland in GB are excluded.
EVS 2008, data file version 3.0.0., variable v336, data weighted to correct regional oversampling for Germany, Belgium, and the UK;
Eurobarometer 73.2 (February-March 2010) files from Eurostat, data file versions 2.0.1, variable v362; Eurobarometer 73.3 (March-April 2010) files from Eurostat, data file versions 2.0.1, variable v362;
Eurobarometer 75.4 (2011) files from Eurostat, data file; version 3.0.1, variable v105 All Eurobarometer data weighted to correct regional oversampling for Germany and the UK;
ISSP 2011-2012, data file versions: 2.0.0 (2011), 2.0.0 (2012); variable DEGREE, data weighted to correct regional oversampling for Germany; Wallonia excluded for BE, Northern Ireland is excluded for GB.
Sample selection: Only respondents aged 25-64.

2.5 Discussion

Comparing the distributions of educational attainment in four large cross-national public opinion surveys over a five-year span, substantial inconsistencies are identified. The first kind are inconsistencies across rounds within individual survey programs. The second are inconsistencies within years across surveys. The size of the discrepancies across surveys sheds substantial doubt on the comparability of educational attainment variables across these data sets: It indicates that more than a fifth of cases on average have to change categories to achieve equal distributions. As a consequence, the comparability of education-related results across studies using these data is in question. It also does not look promising for pooling these data sets.

These inconsistencies cannot reflect 'real' differences in education distributions. Generally, differences in these distributions over short time scales, such as the five-year span we look at, should be minimal because they mostly change through cohort replacement. For identical time points and countries, no real differences in distributions are expected because the samples of the different surveys all follow random sampling techniques, have similar sample sizes, and were harmonized for analysis. The inconsistencies found must therefore have methodological reasons. Inspired by the Total Survey Error Framework (Groves et al., 2009), there are at least four explanations for these inconsistencies: (1) differential unit nonresponse, (2) differential instrument validity, (3) differential measurement error, and (4) differential processing error.

Firstly, on the representation side, differential unit nonresponse is one potential explanation for distributional differences across surveys and even survey waves. The surveys we look at all have substantial amounts of unit nonresponse, which could be differently structured by education across surveys and years. There is, however, no straightforward way to check how much impact this has on our results.

Secondly, on the measurement side, the education question is not standardized across surveys or countries, and there is different wording (sometimes even precise meaning) of this question and the response categories across surveys (and sometimes survey rounds). This may result in differences in validity across surveys, countries, and time points and could explain some of the inconsistencies found. For example, in the ISSP for Slovenia, the question asks about the name of the last school the respondent finished, rather than the highest educational certificate obtained. In the German ISSP, health sector schools are missing on the show card, and therefore, the level of education of nurses is likely to be underestimated compared with other data (see also Schneider, 2009).

Thirdly, also related to the usage of different questionnaire items across surveys and time points, measurement error could differ across data sets because of differential social desirability, item difficulty, or amount of proxy reporting. None of the surveys we analyzed use proxy reporting. We cannot examine the other elements more closely here.

Finally, inconsistencies in harmonization routines, that is, the coding of country-specific education variables over time and surveys into the cross-national education scheme, may also explain the results reported in this study. While all surveys use ISCED as the harmonization framework (apart from ISSP before 2011, which likely explains why the methodological changes in the ISSP in 2011 led to higher inconsistency over time than the change in the ESS 2010), they seem to implement it differently. We can distinguish two kinds of this processing error here: accidental misclassifications because of a lack of information on how to map country-specific education categories to ISCED and deliberate deviations from the official mapping. For Hungary, eight years of basic education is coded as ISCED 1 instead of ISCED 2 and basic vocational education is coded as ISCED 2 instead of 3 in the ISSP 2012, likely continuing to use the coding used in ISSP with the non-ISCED education variable before 2011. Misclassifications such as these are common and have been repeatedly documented (Kerckhoff & Dylan, 1999; Kieffer, 2010; Schneider, 2009). Deliberate deviations from official mappings largely explain the inconsistencies between the ESS and the other surveys for the U.K. and Poland. For the U.K., it was decided to map the General Certificate of Secondary Education grades A–C to ISCED Level 2 rather than 3C, where it is officially coded, which does not make much sense when comparing this with other European countries (Schneider, 2008). Because this is a common qualification, the ESS differs substantially from the ISSP and the Eurobarometer, which use the official mapping. For Poland, basic vocational education completed before 2005 is classified in ISCED Level 2 rather than at ISCED Level 3 in the ESS because the entrance requirements were increased in 2005. Such deviations from official mappings aim at maximizing comparability across countries and time, but they are not coordinated across different public opinion surveys.

2.6 Conclusions

To conclude, we suggest some avenues for improving the measurement, coding, and harmonization of educational attainment in comparative public opinion surveys. We think that cross-national survey management practices and cross-survey cooperation can make a real difference for data consistency by encouraging standardized measurement instruments and harmonization procedures. Firstly, ex-ante output harmonization of complex variables like educational attainment is a challenging task and needs to be trained and quality-controlled (Granda & Blasczyk, 2010), for which it would be worth pooling resources across surveys. Secondly, surveys could learn more from each other. While the Eurobarometer is centrally designed and run by TNS opinion in Brussels, the ESS, ISSP, and EVS consist of rather independent country teams that receive questionnaires and guidelines from the central secretariats and methods groups (e.g., ISSP - Demographic Methods Group, 2010). The ESS however exercises a stronger central overview of country teams than EVS and ISSP. With the development of the new education measures in 2010, a centralized consultation process with several checks in and outside of the national teams, expert workshops, and documentation improvements was performed so that only minor further changes were required in 2012. The education measurement change in the ISSP 2011, in contrast, was less centrally monitored, leading to many readjustments in 2012. These differences in survey management centralization and, in the case of the ESS, quality assurance, may explain the stronger stability over time in the Eurobarometer and ESS compared with the ISSP. Finally, ISCED can legitimately be criticized for being vulnerable to political influence, which is why public opinion survey designers, for the sake of substantive comparability, sometimes deliberately deviate from official ISCED mappings. However, they do so in an uncoordinated way. Agreeing on an ‘academic’ version of ISCED for public opinion surveys like ESS, EVS, and ISSP may be another way forward.

2.7 Funding

This work was supported by the Leibniz Association through the Leibniz Competition (formerly SAW Procedure), grant number SAW-2013-GESIS-5.

2.8 Acknowledgements

We would like to thank the GESIS Data Archive and the Norwegian Social Science Data Services for providing the data sets for this study and for their support. We would also like to thank the anonymous reviewers for their helpful comments and suggestions.

Earlier versions of this article were presented at the European Conference on Quality in Official Statistics 2014 in Vienna and at the 67th Annual Conference of the World Association for Public Opinion Research 2014 in Nice.

2.9 References

Bekhuis, H., Lubbers, M., & Verkuyten, M. (2014). How Education Moderates the Relation Between Globalization and Nationalist Attitudes. *International Journal of Public Opinion Research*, 26(4), 487–500. doi:10.1093/ijpor/edt037

Braun, M., & Müller, W. (1997). Measurement of Education in Comparative Research. *Comparative Social Research*, 16, 163–201.

Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20(2), 210–217. doi:10.2307/2088328

Ehling, M. (2003). Harmonising Data in Official Statistics: Development, Procedures, and Data Quality. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book of Demographic and Socio-Economic Variables* (pp. 17–31). New York: Kluwer Academic/Plenum Publishers.

Granda, P., & Blasczyk, E. (2010). Data Harmonization. *Cross-Cultural Survey Guidelines*. Retrieved from: http://ccsg.isr.umich.edu/harmonization.cfm (accessed July 2015); current version of 2016 can be retrieved from: https://www.ccsg.isr.umich.edu/index.php/chapters/data-harmonization-chapter (accessed April 2020)

Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed). New York: John Wiley & Sons.

Heath, A., Martin, J., & Spreckelsen, T. (2009). Cross-National Comparability of Surveys Attitude Measures. *International Journal of Public Opinion Research*, 21(3), 293–315. doi:10.1093/ijpor/edp034

Hoffmeyer-Zlotnik, J. H. P. (2008). Harmonisation of Demographic and Socio-Economic Variables in Cross-National Survey Research. *Bulletin de Méthodologie Sociologique*, 98(1), 5–24. doi:10.1177/075910630809800103

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research.* Dordrecht: Springer Science+Business Media.

ISSP - Demographic Methods Group. (2010). *ISSP Background Variables Guidelines*. Version of 2010-12-10. Retrieved from: http://www.gesis.org/issp/issp-members-area/ coding-standards/ (accessed July 2015); current version of 2017 can be retrieved from: https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/members/codinginfo/BV_guidelines_for_issp2018.pdf (accessed April 2020)

Kalmijn, M. (2003). Country Differences in Sex-Role Attitudes: Cultural and Economic Explanations. In W. A. Arts, L. C. J. M. Halman & J. A. P. Hagenaars (Eds.), *The Cultural Diversity of European Unity. Findings, Explanations and Reflections from the European Values Study* (pp. 311–337). Leiden: Koninklijke Brill NV.

Kerckhoff, A. C., & Dylan, M. (1999). Problems with International Measures of Education. *Journal of Socio-Economics*, 28(6), 759–775. doi:10.1016/S1053-5357(99)00056-6

Kerckhoff, A. C., Ezell, E. D., & Brown, J. S. (2002). Toward an Improved Measure of Educational Attainment in Social Stratification Research. *Social Science Research*, 31(1), 99–123. doi:10.1006/ssre.2001.0720

Kieffer, A. (2010). Measuring and Comparing Levels of Education: Methodological Problems in the Classification of Educational Levels in the European Social Surveys and the French Labor Force Surveys. *Bulletin de Méthodologie Sociologique*, 107(1), 49–73. doi:10.1177/0759106310369974

Przeworski, A., & Teune, H. (1970). *The Logic of Comparative Social Inquiry. Comparative Studies in Behavioral Science.* New York: John Wiley & Sons.

Schneider, S. L. (2008). The Application of the ISCED-97 to the UK's Educational Qualifications. In S. L. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries* (pp. 281–300). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES).

Schneider, S. L. (2009). *Confusing Credentials: The Cross-Nationally Comparable Measurement of Educational Attainment* (Doctoral thesis). Oxford: University of Oxford, Nuffield College. Retrieved from: https://ora.ox.ac.uk/objects/uuid:15c39d54-f896-425b-aaa8-93ba5bf03529 (accessed April 2020)

Schneider, S. L. (2010). Nominal Comparability is not Enough: (In-)equivalence of Construct Validity of Cross-National Measures of Educational Attainment in the European Social Survey. *Research in Social Stratification and Mobility*, 28(3), 343–357. doi:10.1016/j.rssm.2010.03.001

Schröder, H., & Ganzeboom, H. B. G. (2014). Measuring and Modelling Level of Education in European Societies. *European Sociological Review*, 30(1), 119–136. doi:10.1093/esr/jct026

Smith, T. W. (1995). Some Aspects of Measuring Education. *Social Science Research*, 24(3), 215–242. doi:10.1006/ssre.1995.1008

UNESCO-UIS. (1997 [2006]). *ISCED 1997 International Standard Classification of Education*. Montreal: UNESCO-UIS. Retrieved from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf (accessed April 2020)

Weakliem, D. L. (2002). The Effects of Education on Political Opinions: An International Study. *International Journal of Public Opinion Research*, 14(2), 141–157. doi: 10.1093/ijpor/14.2.141

Data References

ESS. (2012a). *European Social Survey, Round 1, Data 2002.* Data File Version 6.3. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2012b). *European Social Survey, Round 2, Data 2004.* Data File Version 3.3. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2012c). *European Social Survey, Round 3, Data 2006.* Data File Version 3.4. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2012d). *European Social Survey, Round 4, Data 2008.* Data File Version 4.1. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2012e). *European Social Survey, Round 5, Data 2010.* Data File Version 3.0. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2014). *European Social Survey, Round 6, Data 2012.* Data File Version 2.0. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

European Commission. (2012). *Eurobarometer 73.2+73.3 (2-3 2010)*. ZA5236. Data File Version 2.0.1. Cologne: GESIS Data Archive. doi:10.4232/1.11473

European Commission. (2014). *Eurobarometer 75.4 (2011)*. ZA5564, Data File Version 3.0.1. Cologne: GESIS Data Archive. doi:10.4232/1.11851

EVS. (2011). *European Values Study 2008, 4th wave. Integrated Dataset*. ZA4800. Data File Version 3.0.0. Cologne: GESIS Data Archive. doi: 10.4232/1.11004

ISSP - Research Group (2012). I*nternational Social Survey Programme: Religion III - ISSP 2008.* ZA4950. Data File Version 2.2.0. Cologne: GESIS Data Archive. doi:10.4232/1.11334

ISSP - Research Group (2012). *International Social Survey Programme: Social Inequality IV - ISSP 2009*. ZA5400. Data File Version 3.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.11506

ISSP - Research Group (2012). *International Social Survey Programme: Environment III - ISSP 2010*. ZA5500. Data File Version 2.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.11418

ISSP - Research Group (2013). *International Social Survey Programme: Health and Health Care - ISSP 2011*. ZA5800. Data File Version 2.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.11759

ISSP - Research Group (2014). *International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012.* ZA5900. Data File Version 2.0.0, Dataset Hungary: ZA5949. Data File Version 1.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.12022; Dataset Hungary doi:10.4232/1.12117

Paper II

Can We Assess Representativeness of Cross-National Surveys Using the Education Variable?

3 Can We Assess Representativeness of Cross-National Surveys Using the Education Variable?[3]

Achieving a representative sample is an important goal for all surveys. This study asks whether education, a socio-demographic variable covered by virtually every survey of individuals, is a good variable for assessing the realised representativeness of a survey sample, using benchmark data. The main condition for this is that education is measured and coded comparably across data sources. We examine this issue in two steps: Firstly, the distributions of the harmonised education variable in six official and academic cross-national surveys by country-year combination are compared with the respective education distributions of high-quality benchmark data. Doing so, we identify many substantial inconsistencies. Secondly, we try to identify the sources of these inconsistencies, looking at both measurement errors in the education variables and errors of representation. Since in most instances, inconsistent measurement procedures can largely explain the observed inconsistencies, we conclude that the education variable as currently measured in cross-national surveys is, without further processing, unsuitable for assessing sample representativeness, and for constructing weights to adjust for nonresponse bias. The paper closes with recommendations for achieving a more comparable measurement of the education variable.

Keywords: education, comparability, cross-cultural surveys, representativeness, sample quality

3.1 Introduction

How to achieve good survey data quality is an important issue for the whole survey landscape, including official and academic surveys. In addition to reliable and valid measurements, a key criterion for evaluating survey quality is sample representativeness. Commonly response rates are referred to as an important quality indicator for the representativeness of a sample (see Abraham, Maitland, & Bianchi, 2006; Biemer & Lyberg, 2003; Groves, 2006). However, research showed that low response rates do not necessarily lead to nonresponse bias (Bethlehem, Cobben, & Schouten, 2011; Groves & Peytcheva, 2008), so that this indicator alone is insufficient to assess sample representativeness.

Another simple and commonly used approach to evaluate sample representativeness is to compare the data in question to benchmark data by checking

[3] This paper is published as: Ortmanns, V. & Schneider, S. L. (2016). Can We Assess Survey Representativeness of Cross-National Surveys Using the Education Variable? *Survey Research Methods*, 10(3): 189–210. doi: 10.18148/srm/2016.v10i3.6608

descriptive statistics and distributions for core variables (Groves, 2006; Kamtsiuris et al., 2013; Koch, Halbherr, Stoop, & Kappelhof, 2014; Struminskaya, Kaczmirek, Schaurer, & Bandilla, 2014). These benchmark data are often from official sources such as register or census data, and it is assumed that they are the 'gold standard' regarding representativeness (Bethlehem & Schouten, 2016; Billiet, Matsuo, Beullens, & Vehovar, 2009; Groves, 2006). Following this approach, we speak of a representative sample if the relative distributions of a core set of stable (e. g. socio-demographic) variables in the survey are equal to the relative distributions in the target population (Bethlehem et al., 2011). This focus is justified when looking at large-scale general population surveys using probability based (rather than e. g. quota) sampling methods and best available sampling frames and designs. In European comparative research, in the absence of suitable register or census data of the target population, the European Union Labour Force Survey (EU-LFS) is commonly used as the benchmark for this purpose.

Mostly socio-demographic variables are used for the comparisons between benchmark data and the surveys in question (e. g. Koch et al., 2014; Peytcheva & Groves, 2009; Struminskaya et al., 2014). The age and gender variables are especially suitable due to their high measurement quality and straightforward comparability. However, age and gender are insufficient criteria on their own for judging a samples' representativeness. Another commonly-used socio-demographic variable is education, which is also covered in almost all surveys (Homeyer-Zlotnik & Warner, 2014; Smith, 1995). In statistical analyses, education is often used as an independent variable to explain, for example, attitudes, beliefs, and behaviours (Kalmijn, 2003; Kingston, Hubbard, Lapp, Schroeder, & Wilson, 2003). It could be a sensitive marker of representativeness: Several studies show that samples in academic surveys contain an education bias; less educated people are often underrepresented in surveys likely due to selective nonresponse (Abraham et al., 2006; Billiet, Philippens, Fitzgerald, & Stoop, 2007; Couper, 2000; Groves & Couper, 1998). Nonresponse bias, which occurs if the characteristic that influences response propensity is also related to the variables we wish to analyse (Biemer & Lyberg, 2003; Kreuter et al., 2010), is thus particularly likely to occur with respect to education. Being able to use the education variable for constructing weights to adjust for nonresponse would thus be highly desirable.[4] However, the comparison with benchmark data becomes much more challenging with the education variable because it is more complex than the gender and age variables, and therefore contains more possibilities for errors on the measurement side (Billiet et al., 2009; Schneider, 2008b).

From previous research we know that the distributions of the education variable for the same country, year, and age-groups between EU-LFS and other survey data are often not equal, even though supposedly coded in the same way. Kieffer (2010) in her analysis focuses on French data from the EU-LFS and the European Social Survey (ESS) from 2002 to 2008, and identified large discrepancies in the distributions for 2002, 2004 and 2006 but smaller discrepancies for the 2008 data. Schneider (2009) shows inconsistencies between data from the ESS, the EU-LFS, and the European Survey of Income and Living Conditions (EU-SILC) for the years 2002 to 2007 for most European countries. Her analysis uses Duncan's Dissimilarity Index for comparing the distributions of the education variable. Ortmanns and Schneider (2016), using the same method, find inconsistent educational distributions across four mostly European public opinion surveys: the ESS, the European Values Study (EVS), the International Social Survey Programme (ISSP), and the Eurobarometer. All authors attribute those inconsistencies to inconsistent measurement procedures rather than non-representativeness.

We extend the study by Schneider (2009) by using data from 2008 to 2012, and the study by Ortmanns and Schneider (2016) by adding official surveys - the EU-LFS, EU-SILC, and OECD's Programme for the International Assessment of Adult Competencies (PIAAC). The research question of this paper is: Can we use the education variable for assessing the realised representativeness of the samples of cross-national academic and official surveys? If yes, benchmark data could be used for constructing weights to correct for nonresponse bias (Bethlehem & Schouten, 2016; Kreuter et al., 2010).

In order to answer this question, we firstly present the methodological background on sample representativeness and the measurement of education in cross-national

[4] For a discussion of the merits and effectiveness of weighting and weighting techniques see e. g. Bethlehem (2002), Bethlehem et al. (2011), Gelman and Carlin (2002).

surveys. Then we introduce the data sources, our analysis strategy and Duncan's Dissimilarity Index as our measure of consistency across surveys. Then the results are presented and interpreted with regards to possible sources of inconsistencies using the Total Survey Error (TSE) framework (Groves et al., 2009; Groves & Lyberg, 2010). The paper ends with conclusions and some practical recommendations for achieving more comparable education variables in cross-national surveys.

3.2 Methodological Background

This section is structured by the two dimensions of TSE which distinguish survey errors resulting from problems of representation and measurement. We first clarify how we use the term sample representativeness, and review different methods for evaluating it. Then we describe the challenges of measuring education in such a way that it can be compared across countries and surveys.

3.2.1 Sample Representativeness

A representative sample is important for surveys in order to achieve data that allow statistical inferences about the whole target population (Biemer & Lyberg, 2003). The terms 'representative samples' or 'sample representativeness' however have many different interpretations (Kruskal & Mosteller, 1979a, 1979b, 1979c). In this paper, we concentrate on the aspect of achieving equal distributions between the surveyed and the target population in large-scale probability based surveys. If a certain group of the population with specific characteristics is less well covered by the survey sample, it is no longer representative of the population and overrepresents some and underrepresents other groups. Those non-observation errors in principle occur through a combination of coverage, sampling, or nonresponse bias (Bethlehem et al., 2011). There are three main methods for assessing sample representativeness: response rates, R-indicators and benchmark comparisons.

The most commonly used indicator for representativeness is the response rate (Abraham et al., 2006; Biemer & Lyberg, 2003; Groves, 2006). Surveys with very high response rates are commonly regarded as representative, if probability sampling methods are employed and respondent substitution is barred, because they imply a low nonresponse rate. The nonresponse rate indicates the upper limit of the possible nonresponse bias. It "refers to the percentage or proportion of sample cases not included in the eventually realised sample, for whatever reasons (refusals, non-contacts, other reasons)" (Heerwegh, Abts, & Loosveldt, 2007, p. 3). However, from research we know that low response rates do not necessarily lead to a non-representative or biased sample, if nonresponse is random and no bias occurs (Bethlehem et al., 2011; Groves & Peytcheva, 2008). In addition, response rates also ignore errors due to different sampling frames or sampling methods. Therefore response rates alone are an insufficient indicator for evaluating sample representativeness.

A more recently developed set of indicators assessing representativeness of surveys are model-based representativeness measures, such as the R-indicator, partial R-indicator (Bethlehem et al., 2011; Schouten, Cobben, & Bethlehem, 2009), and other balance and distance indices (Lundquist & Särendal, 2013). These indicators compare the set of respondents to a survey to its gross sample, which includes the respondents as well as the nonrespondents (Bethlehem et al., 2011; Schouten et al., 2009). They therefore predominantly assess nonresponse bias while neglecting potential coverage and sampling biases (Nishimura, Wagner, & Elliott, 2016). These sample-based representativeness indicators require auxiliary data for respondents and non-respondents. These auxiliary data are usually taken from the sampling frame, e. g. a population register (Schouten et al., 2009). However, information on the education of survey nonrespondents is not available in most sampling frames, except for some countries' population registers, such as in the Netherlands and the Scandinavian countries. Since we wish to look at a much wider range of countries, for which such auxiliary data is not available, we cannot use this approach for assessing the realised sample representativeness.

The third approach uses benchmark data for evaluating the realised sample representativeness. It compares the distributions of selected variables covered by both the data to be evaluated and the benchmark data. The advantages of this approach are firstly its simplicity from a statistical perspective, and secondly the availability of the required benchmark data. Thirdly, coverage and sampling errors are also reflected in benchmark comparisons. However, using this approach requires that the measurement of the variable(s) to be used is comparable. Another disadvantage of using benchmark data is that these data might not be free from (measurement and representation) errors themselves (Groves, 2006; Koch et al., 2014). Typical variables used for this approach are socio-demographic variables (e. g. Koch et al., 2014; Peytcheva & Groves, 2009; Struminskaya et al., 2014) because these are covered in almost every survey and it is assumed that those are usually measured in an equivalent way. Mostly age and gender are used quite often to evaluate the representativeness of a sample, but also education. However, it is well-known that the measurement of education in cross-national surveys is highly complex, which we turn to next.

3.2.2 The Measurement of Education in Cross-National Surveys

In this paper we thus want to figure out whether the education variable is suitable for evaluating the representativeness of a survey sample. To answer the survey question on educational attainment, respondents typically need to identify their highest formal educational qualification in a list of categories. This list is country-specific, because the national elements of the educational system and the names of the qualifications cannot be input harmonised (Schneider, Joye, & Wolf, 2016). The country-specific answer categories have to be mapped into a standard coding scheme before data collection. This approach is called ex-ante output harmonisation (Ehling, 2003). Therefore the survey team has to agree on such a standard coding scheme, and clear guidelines and rules have to be defined for developing the country-specific answer categories and the coding procedure (Ehling, 2003; Eurostat, 2006; Eurostat & OECD, 2014). Most comparative surveys agree on some variant of the International Standard Classification of Education (ISCED).

ISCED was designed by UNESCO in the 1970s and revised in 1997 and 2011 (for details on the most recent update, see Schneider, 2013). It was developed in order to facilitate comparisons of country-specific educational programmes for comparative education statistics. Therefore ISCED defines international levels and types of education (UNESCO-UIS, 2006), and education ministries and national statistical institutes map national educational programmes to these levels and types. Since ISCED 97 is used in the surveys analysed in this article, we limit our presentation to ISCED 97. The main levels of ISCED 97 are:

- ISCED 0: Pre-primary education (or not completed primary education)
- ISCED 1: Primary education or first stage of basic education
- ISCED 2: Lower secondary or second stage of basic education
- ISCED 3: Upper secondary education
- ISCED 4: Post-secondary non-tertiary education

- ISCED 5: First stage of tertiary education
- ISCED 6: Second stage of tertiary education

We focus on these seven main levels and ignore the additional complementary dimensions of ISCED 97, because most of the surveys we look at do not use them (see section 3.3.1).

3.3 Data and Method

In this section, we introduce the data sources and their education variables in the first part. In the second part, the analysis strategy and the indicator of data consistency are described.

3.3.1 Data and Education Coding

For our analysis we select those cross-national survey data that permit the construction of a common education coding scheme based on ISCED, i.e. that claim to use ISCED for education coding. Further criteria are the application of random probability sampling, no substitution of respondents, and coverage of a wide range of European countries. This resulted in the selection of seven diverse cross-national survey datasets on a wide range of topics and with very different cross-national set-up and organisation: the EU-LFS (Eurostat, 2013a, 2013b, 2013c, 2013d, 2013e) and the EU-SILC (Eurostat, 2008b, 2009b, 2010, 2011, 2012) as official data, the OECD's PIAAC (OECD, 2013) and the Eurobarometer (European Commission, 2012, 2014) as policy related studies, and three academic surveys: ESS (2012a, 2012b, 2014b), EVS (2011), and ISSP (ISSP - Research Group, 2013, 2014). We focus on the years 2008 to 2012.

The EU-LFS provides official quarterly household data for monitoring employment and unemployment in the EU. The data are available from the 1970s onwards and cover all European Union countries. As mentioned above, the EU-LFS is used as benchmark data in this study. We only use the spring (second) quarters of the data in our analyses. On average across years and countries, the response rate for 2008 to 2012 is 78% (also due to compulsory participation in 13 of 31 countries, see Eurostat, 2013f). Because of the relatively high response rates, we expect less error of non-observations of lower educated respondents in this data, especially when participation is mandatory, than in the academic surveys. What has to be kept in mind is that the EU-LFS allows proxy-reports. Those, in general, raise the response rates, but may also result in measurement errors. With regard to the education variable, the EU-LFS provides a harmonised variable for all countries consisting of 13 categories, thus distinguishing ISCED main levels as well as some elements of sub-dimensions. We expect the coding of the country-specific qualifications into the official ISCED classification to follow the official ISCED mappings, using the basic principles for implementing ISCED formulated by Eurostat (2006, 2008a). The harmonisation process of the country-specific education variables takes place in the statistical institutes of the EU member states rather than centrally at Eurostat. In this study we use the EU-LFS as the benchmark data, because of its wide country coverage, probability sampling methods, relatively high response rates, and large sample sizes, supposedly leading to representative data and precise estimates for any given country.

The EU-SILC was launched in 2003 with the aim of providing cross-sectional and longitudinal official micro-data on income, poverty, social exclusion, as well as living and housing conditions in the EU. The average response rate from 2008 to 2012 is around 80%. The EU-SILC also allows proxy-reports. In the EU-SILC, ISCED main levels 5 and 6 are not distinguished. Coding of country-specific education variables into the ISCED categories for the EU-SILC is also done by the national statistical offices (Eurostat, 2008a, 2009a). Therefore we expect a close match with the EU-LFS data, which was demonstrated for earlier years by Schneider (2009).

OECD's PIAAC data were first collected in 2011/ 12. This study measures adults' key cognitive skills across OECD countries. The response rate on average is 60%, and there is neither proxy reporting nor compulsory participation in any country (OECD, 2016). PIAAC's education variable adopts the EU-LFS coding scheme and additionally anticipates ISCED 2011 by providing more differentiation at the tertiary level.

The politically-driven Eurobarometer programme was set up by the European Commission in the 1970s to monitor public opinion towards the EU and related topics in all member states. The European Commission unfortunately does not provide information on the response rates of the Eurobarometer studies. Since they do not measure educational attainment on a regular basis, only three Eurobarometer studies from 2010 and 2011 (European Commission, 2012, 2014), which contain main ISCED 97 levels, are included in this study.

The ESS was set up in 2002 and measures individuals' attitudes, beliefs, and behaviour patterns in around 30 European countries. The response rate on average for the years 2008 to 2012 is around 60% (see e. g. ESS, 2014c). Up to 2008, the harmonised education variable consisted of ISCED 97 main levels, but categories 0 and 6 were integrated in categories 1 and 5 respectively. The education variable was changed in 2010, with the aim of achieving more informative and more comparable education variables, introducing a detailed cross-national variable closely anticipating ISCED 2011.

The EVS, which also covers a large number of European countries, was launched in 1981 and also focuses on respondents' values, attitudes, and beliefs. The average response rate for the latest wave (2008) is 56% (EVS & GESIS, 2010). This is the first EVS wave implementing a harmonised education variable representing main ISCED 97 levels.

The ISSP was set up in 1985 and also measures peoples' attitudes and values and extends beyond Europe. For the European countries covered in the ISSP, the response rate on average for 2011 and 2012 is around 50% (see e. g. ISSP, 2015). Before 2011, the ISSP used an education scheme that was specific to the ISSP, but since 2011 one closely related to ISCED has been implemented for measuring educational attainment. Therefore, we will include only ISSP data from 2011 and 2012. However, all upper secondary (ISCED 3) or post-secondary non-tertiary (ISCED 4) qualifications which give access to tertiary education are coded in category 3, and category 4 contains all other upper secondary and post- secondary non-tertiary qualifications, that are more technically oriented or designed for directly entering the labour market (ISSP – Demographic Methods Group, 2010). Therefore ISSP categories 3 and 4, as well as ISCED levels 3 and 4 of the EU-LFS have to be aggregated to render the coding schemes of both sources comparable.

To summarize, while all these surveys use ISCED 97 as their education coding scheme, each survey defines the specific codes to be used slightly differently. Therefore we further harmonise the different education variables ex-post by focusing on the main ISCED levels, with some adjustments: As the EU-SILC, ESS 2008, and the ISSP do not distinguish between ISCED levels 5 and 6, we combine those two levels. The same is true for ISCED levels 0 and 1, which cannot be differentiated in the ISSP and the ESS 2008 (and many countries in the EU-LFS also fail to make this distinction). The correspondence of the survey-specific ISCED variables and our adapted 5 level version (4 level version for the ISSP) used for the analyses in this study is shown in Table A3.1 and Table A3.2 in the Supplemental Material.

From each survey, respondents aged 25 to 64 are selected to render samples as comparable as possible. Data are weighted using design weights when available. Cases with missing values on the education variable are excluded from the analysis. This is unproblematic because item-nonresponse on the education variable is generally very low.

3.3.2 Analysis Strategy and Method

Our analysis consists of two steps. Firstly, we compare the distributions of the education variable across surveys to see whether the data are consistent across data sources. Secondly, we examine those cases revealing the largest inconsistencies to find out whether these can be explained by differences in measurement procedures, or by lack of representativeness of the sample.

In the first step, for measuring the inconsistencies of the harmonised education variable, we compare the education distributions for the same country and year between the EU-LFS and each other survey presented in section 3.3.1, by calculating Duncan's Dissimilarity Index (Duncan & Duncan, 1955).[5] Originally, Duncan's Dissimilarity Index was developed for measuring residential segregation, but it can also be used to measure differences in the distributions of categorical variables more generally. Formally, if x_i denotes the size of category i out of k ISCED categories for country A in year B in survey S, and y_i denotes the same for country A in year B in survey T, the index is defined as: $D = ½ \sum_{i=1}^{k} |x_i - y_i|$.We rescale the index to range from 0 to100 in order to interpret the resulting number as the percentage of cases that would have to change categories in order to achieve an equal education distribution across the two data sources. This can be regarded as the TSE with respect to the education variable.

[5] In the case that some countries run their fieldwork a year later than foreseen (for example Italy and Finland in 2009 instead of 2008 in the EVS), we stick to the main survey year (in this case 2008). We do not expect a substantial change in the distribution of the education variable across two consecutive years because the actual educational distribution in the population only changes very slowly through cohort replacement.

In the second step, for cases revealing specifically large deviations, we examine whether those are likely to be caused either by measurement errors in the education variable, or by errors of non-observation, or both. For this analysis we try to unpack the overall discrepancies. To do this, we have a closer look at the frequencies of the ISCED variable across the two surveys in question and check whether the inconsistencies are concentrated in specific ISCED levels or whether they are spread across the education spectrum. If we identify an inconsistency in specific ISCED levels, we have a closer look at the country-specific questions and showcards of the survey (if available) and analyse the exact wording of the categories on the showcard in comparison with the respective information for the EU-LFS. We then check to which ISCED levels the qualifications are coded, and whether this coding appears to differ from the official (EU-LFS) coding. For interpreting the coding in the EU-LFS we used the ISCED mappings of 2013, which contain ISCED 1997 codes used in the EU-LFS, as earlier versions are not publicly available. For the ESS, EVS, ISSP and Eurobarometer, the country-specific education variables and the ISCED variable are included in the datasets, so a simple cross tabulation can be made to identify the mapping used. If we do not find any explanation on the measurement side for the inconsistent education distributions, i.e. if the instrument and coding appear equivalent, we conclude that the representativeness of the sample is probably in question.

One challenge is that it is very difficult to disentangle, let alone quantify, measurement and representation errors empirically. Another challenge when comparing the survey data in question with data from official surveys is that the latter are also not free from errors: The variables of interest could be measured differently, e. g. by allowing proxy respondents, or the samples' characteristics regarding coverage and nonresponse may be different, which both leads to discrepancies in the distributions (Billiet et al., 2009; Groves, 2006). We are aware of the fact that these errors also occur in our benchmark data, the EU-LFS, and that register or census data would be better, if they existed in a comparable fashion across Europe. However, for the reasons mentioned above, this is the most adequate benchmark for this task. Rather than naively assuming the EU-LFS as a 'golden standard', when presenting and discussing the results we will try to take potential quality issues with this benchmark data itself into account.

3.4 Results

In line with the two steps of our analysis strategy, we first present the results regarding Duncan's Dissimilarity Index, with which we identify inconsistencies in the education distributions within countries and years across surveys. We then move on to examine more closely several examples of countries and survey years with large inconsistencies.

3.4.1 Comparing Distributions of the Education Variable Across Surveys

For a first overview, Figure 3.1 shows the boxplots of Duncan's Dissimilarity Index across countries in percent for comparisons between the EU-LFS data and the other six surveys for all possible time points in the years 2008 to 2012. Detailed results for individual countries can be found in Table A3.3 in the Supplemental Material. Comparing EU-LFS and EU-SILC, the median across countries of Duncan's Index is between 4 and 5% in years 2008 to 2012. On average, around 4% of the respondents would have to change into another category to reach equal education distributions across the two datasets. The highest inconsistencies, on average over the five years, are observed for Iceland (16%), Switzerland (15%), and Luxembourg (13%). The smallest deviations between EU-LFS and EU-SILC can be found for the Czech Republic (less than 1%), Slovenia and Austria (both around 2%). The education distributions in these latter countries thus lie very close together which means they are almost perfectly consistent across the two surveys.

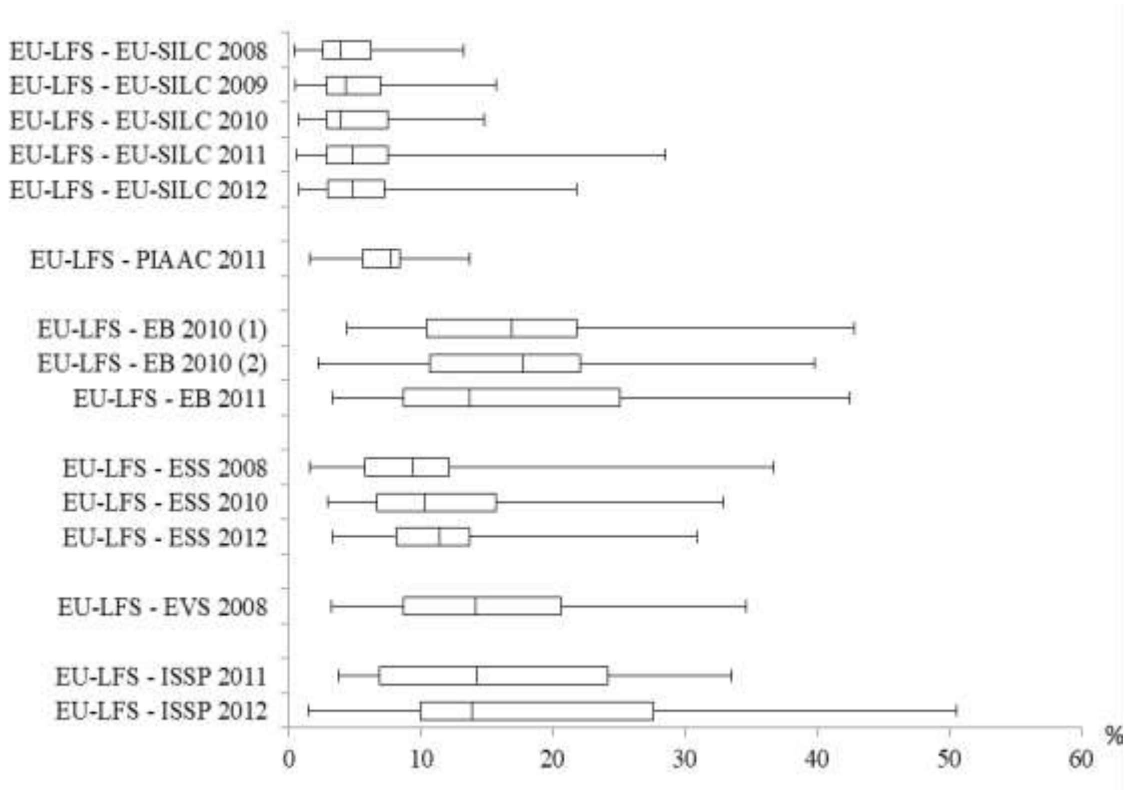


Figure 3.1 Boxplot of Duncan's Dissimilarity Index across countries for all survey comparisons

Data sources:
EU-LFS 2008-2012 (second quarter used only), files from Eurostat, data file versions 2013, variable HATLEVEL, weighted using variable coeff;
EU-SILC 2008-2012, files from Eurostat, data file versions: 2008-3, CROSS-2009-4, CROSS-2010-3, CROSS-2011-1, CROSS 2012, variable PE040, weighted using variable PB040;
PIAAC 2011, file from OECD, data file version 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB);
Eurobarometer 73.2 & 73.3 (2010) files from Eurostat, data file versions 2.0.1, variable v362; Eurobarometer 75.4 (2011) files from Eurostat, data file version 3.0.1, variable v105, data weighted to correct regional oversampling for Germany and the UK;
ESS 2008, data file version 4.1, variable edulvla;
ESS 2010-2012, data file versions: 3.0. (2010), 2.0 (2012) variable edulvlb, weighted using variable dweight;
EVS 2008, data file version 3.0.0, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK;
ISSP 2011-2012, data file versions: 3.0.0 (2011), 2.0.0 (2012) and 1.0.0 (2012) for Hungary, variable DEGREE, data weighted to correct regional oversampling for Germany.
Only respondents aged 25-64 for all surveys.

When comparing data from PIAAC and EU-LFS, the median of Duncan's Dissimilarity Index is 8%. For Norway (14%), England and Northern Ireland[6] (12%), and the Slovak Republic (11%) the highest discrepancies are found. For Austria (2%), France and Sweden (both around 3%) the inconsistencies are smallest.

The median of Duncan's Index in the three Eurobarometer studies in 2010 and 2011 and the EU-LFS of the equivalent years is much higher, between 14 and 18%. We found very high discrepancies between the education distributions of the two data sources of around 40%, averaged over the three comparisons, for the Netherlands, Malta, and Hungary. Small inconsistencies are identified for Slovenia, the Slovak Republic (both around 4%), and Poland (5%).

Comparing ESS 2008, 2010 and 2012 education distributions with those from the corresponding years of the EU-LFS, the median value of Duncan's Index lies between 9 and 11%. High inconsistencies can be found for the UK (25%), Poland (23%), and Denmark (19%) across the three years. The smallest deviations are observed for Bulgaria (3%), Switzerland and Portugal (both around 4%).

With respect to the comparison of EVS and EU-LFS 2008, the median value of Duncan's Index across countries is 14%. We found the largest discrepancies between the two education distributions for Estonia (35%), Finland (30%), and Slovenia (27%) and the smallest for the Czech Republic (3%), the Slovak Republic (4%), and Bulgaria (5%).

Finally, comparing the ISSP and the EU-LFS 2011 and 2012, the median value for the inconsistency of the further aggregated ISCED classification (see Table A3.2 and section 3.3.1) also amounts to 14%. On average, the highest discrepancies are observed for Austria (50%), followed by Denmark (33%), and the Slovak Republic (32%). The lowest inconsistencies are found for Latvia (1%), Bulgaria and Portugal (both around 5%).

The overall median inconsistency of education distributions between the six surveys and the EU-LFS for the time period 2008 to 2012 and across countries lies around 10%. The lowest – and substantively non-problematic – median inconsistencies

[6] For this comparison, Scotland and Wales were excluded from the EU-LFS because they are not covered in PIAAC.

and also the smallest range as well as interquartile range across countries can be observed between the EU-LFS and EU-SILC and the EU-LFS and PIAAC. The large range between EU-LFS and EU-SILC 2011 is the effect of one outlier, namely Iceland. Intermediate median inconsistencies and respectively intermediate ranges and interquartile ranges are identified for the ESS as compared to the EU-LFS. For the comparison of the EU-LFS and the EVS we find a slightly higher median and a higher interquartile range than for the comparison of EU-LFS and ESS data, while the range of inconsistencies is similar. For comparing the EU-LFS with the ISSP (both years) and the Eurobarometer 2011 respectively, we identified the same median inconsistencies. The interquartile range however shows a larger variation of inconsistencies for these benchmark comparisons. For the ISSP 2012 we observe the largest range, caused by the outlier Austria. We find the highest discrepancies when comparing the data of the EU-LFS and the two Eurobarometer studies for 2010, which however show a somewhat lower interquartile range than the comparison with the Eurobarometer 2011 (at constant range).

Overall, the inconsistencies and the interquartile ranges shown in the boxplots vary quite strongly across survey programmes for the same countries and time points. Since the actual education distribution in the population only changes very slowly through cohort replacement, the observed inconsistencies must be ascribed to methodological factors. This may mean either a problem of poor representativeness, or systematic differences in the measurement of education between the surveys. In the next step of the analysis, we will try to disentangle these two factors.

3.4.2 Explaining Inconsistencies Between Education Distributions Across Surveys

As main factors for explaining the inconsistencies, we distinguish between the two dimensions of the TSE - the measurement and the representation sides, where measurement includes instruments and data processing (Groves et al., 2009). We attempt a deeper interpretation of the results for those 35 country-survey-year-comparisons (affecting 18 countries) showing inconsistencies in the education distributions of more than 25% in at least one comparison between one of the six surveys and the EU-LFS. For each of these comparisons, we first check whether we can find signs of systematic errors on the measurement side, i. e. problems regarding measurement instruments, the response process and data processing, which in the case of comparative education measurement refers to output harmonisation.[7] Then, especially if we do not find any hints at measurement and harmonisation problems, we look out for signs of errors of non-observation, especially selective nonresponse. In the following section we present one case per survey error source in more detail. Illustrative results for these selected cases can be seen in Figure 3.2.

Errors Related to Measurement Instruments

The first set of problems that may explain inconsistencies in measured education distributions results from inconsistent or sub-optimal wording of education questions and response categories as well as missing response categories. A rather rare example for *different question wording* (or even choice of different empirical indicators across surveys), which just misses the 25% criterion, is observed for Slovenia in the ISSP, where the education question asks for the last school that was completed rather than the highest educational qualification obtained. While, these two indicators probably correlate highly and this issue thus only explains part of the discrepancy between ISSP and EU-LFS, it is a remarkable lack of input harmonisation.

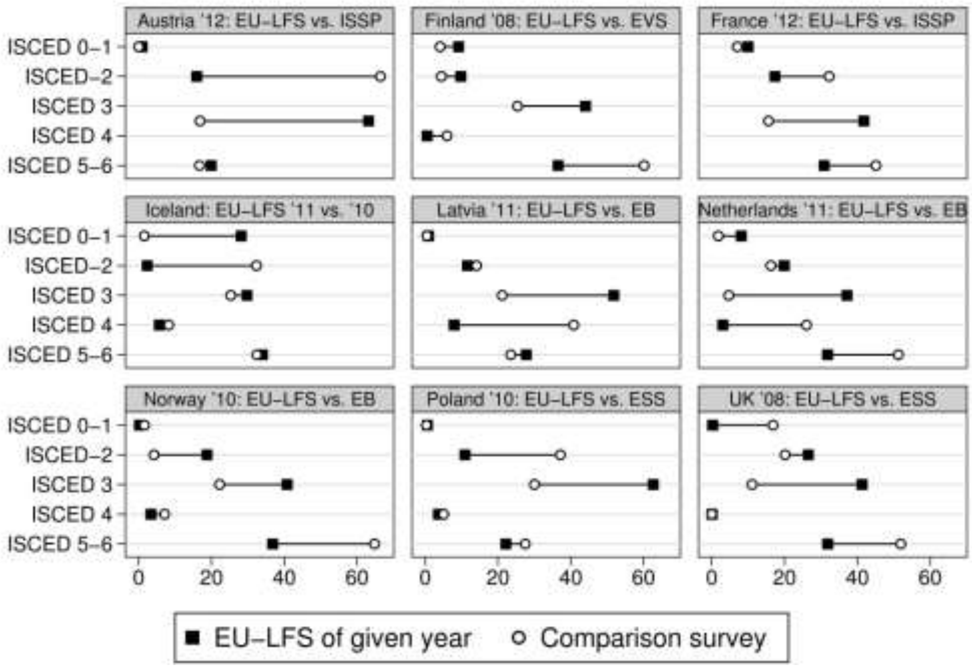
A further example related to the measurement instrument is the *number of questions* asked on the highest educational attainment. Regarding Germany most surveys ask two questions, one on the school leaving certificate, and one on post-school education. Therefore, German respondents are used to individual showcards for the school certificates and vocational and higher education qualifications. In the Eurobarometer only one question is asked and consequently only one (but therefore very long) showcard is presented. This could lead to stronger primacy effects in the Eurobarometer, if respondents select the first matching entry, likely a school-leaving certificate, rather than the highest one (as they should). This could likely explain the large discrepancy which is 24% on average between the three EU-LFS and Eurobarometer studies. The largest deviations are observed for ISCED categories 2 and 3.

[7] Some (especially Nordic) countries in some surveys (mostly the EU-LFS) obtain socio-demographic data from population registers rather than by actually asking respondents. This could explain part of the inconsistency found for these countries between the education distribution in the EU-LFS and the other data sources, which are however rather small (apart from Iceland, see below). While register-based data do not contain survey measurement error, we cannot be sure about the quality either; for example they may be incomplete with regards to the education of migrants, and of nationals who have completed their education abroad.

An example for the use *of vague or ambiguous terms* in the questionnaires and on the showcards is France in the ISSP 2012. The inconsistency in the education distributions compared to the EU-LFS 2012 amounts to 29%. Around one third of the respondents are coded into ISCED level 2 in the ISSP whereas in the EU-LFS only 17% fall into this category. Regarding the combined ISCED levels 3 and 4, 16% of the ISSP respondents and 42% of the EU-LFS respondents are coded here (see Figure 3.2). The ISSP showcard contains ambiguous terms and descriptions rather than specific names of educational qualifications, especially regarding vocational upper secondary and tertiary education. For example, it generically mentions "vocational training after lower secondary school" ("Enseignement professionnel après le college") in two response categories. Such terms do not easily correspond to the specific names of French vocational training certificates, programmes, or institutions that respondents may have in mind, such as CAP ("Certificat d'aptitude professionnelle") or BEP ("Brevet d'études professionnelles"). This could be confusing for respondents who may not find their specific qualification on the showcard and thus may be unsure which category to pick. The EU-LFS showcard is more precise through offering these specific qualifications as response options. However, the discrepancy at ISCED levels 5 and 6, into which 45% of the ISSP respondents and 31% of the EU-LFS are classified, cannot easily be explained by measurement error because the way the categories are worded should lead to underreporting rather than over reporting of tertiary education in the ISSP. Here we rather think of an education bias in the sample of the ISSP. This probably is in line with the large deviation of nearly 50 percentage points in the response rates: 37% in the ISSP in contrast to 85% in the EU-LFS. Further examples of this kind where we find a mix of showcard issues and selective nonresponse are Denmark, the Netherlands and Sweden in the ISSP.

A related kind of measurement error occurs when an a*nswer category is entirely missing* on the showcard. In this situation, some respondents do also not know which category to choose, but here there basically is none that would really fit. This probably happened in Latvia in the three Eurobarometer studies, where the inconsistency between the Eurobarometer and the EU-LFS data is above 30%. The largest discrepancies are observed for ISCED levels 3 and 4. In the Eurobarometer, more than one third of respondents chose one specific response category that is coded to ISCED level 4, while in the EU-LFS only around 8% fall into ISCED level 4. This category on the Eurobarometer showcard, translated into English, reads "Post-secondary education including professional continuing education, but not higher education programmes (1–3 years after general upper secondary school)".[8] Due to the absence of a category referring to vocational training after *lower* secondary school ("pamatskola") on the Eurobarometer showcard, all respondents having vocational training probably pick this category, whether or not they have actually completed general upper secondary education. In contrast to the Eurobarometer, the showcard of the EU-LFS in Latvia contains five categories covering professional programmes that respondents can more easily choose from and will thus be correctly coded in ISCED. A similar problem regarding missing vocational education categories in the Eurobarometer is observed for upper secondary education in Sweden. Such sub-optimal provision for vocational education is quite common in education questions. This may have several reasons: firstly, vocational education may not be considered as formal education; secondly, it may be regarded as irrelevant when the measure of education is only meant as a proxy for academic skills; and thirdly, the number of respondents who have vocational qualifications is estimated to be rather small. All these reasons are problematic in the context of cross-national surveys when different surveys and countries may opt for different solutions in the absence of clear guidance.

[8] Name of the Latvian category on the Eurobarometer showcard: "Pēcvidējā izglītība ieskaitot profesionālās tālākizglītības programmas, bet ne augstākās izglītības programmas (1-3 gadi pēc vidusskolas)". Normally, education answer categories contain country-specific names of educational qualifications, which cannot be translated. Since the Latvian showcard in the Eurobarometer here only provides a generic description, translation is possible in this case.



For Austria and France: Markers between ISCED 3 and ISCED 4 refer to ISCED 3−4

Figure 3.2 Education distributions (in percent) in different surveys and years for selected countries

Data sources:
Second quarter data for all data from the EU-LFS data. Figure shows proportions of variable HATLEVEL weighted by variable coeff for data of the EU-LFS;
Austria, France: Eurostat (2012a), ISSP Research Group (2014, variable DEGREE);
Finland: Eurostat (2008b), EVS (2011, variable v336);
Iceland: Eurostat (2011a), Eurostat (2010a);
Latvia, Netherlands: Eurostat (2011a), European Commission (2014, variable v105);
Norway: Eurostat (2010a), European Commission (2012, variable v362);
Poland: Eurostat (2010a), ESS (2010b, variable edulvlb, weighted using dweight);
United Kingdom: Eurostat (2008b), ESS (2008, variable edulvla, weighted using dweight).
Only respondents aged 25-64 for all surveys.

Errors Related to Data Processing

Inconsistent application of ISCED, 'accidental' or intended, is another important source of inconsistent education distributions on the measurement side. If we find documentation on a decision to deviate from the official ISCED mapping or we find straightforward reasons such as educational reforms, we call this an intended deviation – which should likely not be called an error in the survey in question, but an error or gap in the official ISCED mappings: such deviations are typically made in order to improve comparability across countries and time. This latter situation can only occur in academic surveys because official surveys are bound to use the official ISCED mappings. We thus define misclassifications as 'accidental' when we could not find 'obvious' errors or documentation showing why a certain qualification is coded into a different ISCED category than suggested by the official mappings – we then have no reason to think that the deviation was intended.

Firstly, an example where we identify a processing error in the benchmark data of the EU-LFS: Iceland. This inconsistency is identified because in 2011, Iceland in the EU-LFS produces large discrepancies compared to *all* other surveys. Therefore we have a look at the EU-LFS data over time and spot a high value of Duncan's Dissimilarity Index of 33% comparing EU-LFS data of 2010 and 2011. It seems that a large number of respondents previously coded in ISCED level 2 was downgraded to the combined category of ISCED level 0 and 1 in 2011. The coding in 2010 seems to be correct and is also implemented in the other surveys. We could not identify the reason for the shift of coding in the EU-LFS in 2011. Maybe it is 'just' a coding error made by a human that slipped through any quality check. This example shows that our benchmark data is not free from errors, either.

Another factor which may lead to 'accidental' misclassification is complications in the communication process between the different teams working on the survey. This may be the explanation for the deviation of around 50% for Austria in the ISSP 2012 from the EU-LFS – this is the highest discrepancy identified in the whole analysis. The largest deviation is on ISCED level 2, in which 16% of the respondents in the EU-LFS and 66% of the respondents in the ISSP are found. For ISCED level 3, the distributions are the other way round. What probably happened is that Austria still used the coding scheme of previous ISSP rounds, in which vocational upper secondary school ("berufsbildende mittlere Schule"), was coded to category 2 instead of category 4 (which is where vocational ISCED 3 qualifications are found in the ISSP since 2011, see section 3.3.1) as it now should be. Austria did not participate in the ISSP in 2011 and thus may have missed instructions on the changes of the education variable.

A third example of an 'accidental' misclassification where we could also identify the specific coding error relates to the Netherlands in the Eurobarometer. The overall discrepancy between the Eurobarometer and the EU-LFS 2011 for the Netherlands is over 40%. In the Eurobarometer around one fourth of the respondents are found in ISCED level 4, whereas only 3% of the EU-LFS respondents belong to this category. Instead, around 37% of the EU-LFS respondents and only 5% of the respondents of the Eurobarometer are coded to ISCED level 3. This can be explained by the misclassification of the school-leaving certificates of upper secondary institutions such as the VWO (“Voorbereidend wetenschapplijk onderwijs”), HBS (“Hogereburgerschool”), and the vocational qualification MBO (“Middelbaar beroesponderwijs”). These qualifications are classified into ISCED level 4 in the Eurobarometer instead of ISCED level 3, as they should be according to the official ISCED mappings. The discrepancy at ISCED levels 5 and 6, into which half of the Eurobarometer respondents and 32% of the EU-LFS are classified, cannot be explained by differences between instruments or processing error. Here we assume an education bias in the Eurobarometer sample. Further examples of ‘accidental’ misclassifications, which are not discussed here in detail, can be found for Hungary in the Eurobarometer and the ISSP (see Ortmanns & Schneider, 2016) Slovenia in the ISSP and EVS, Sweden and the Slovak Republic in the ISSP, and Spain in the Eurobarometer.

Intended deviations from the official ISCED coding are a further possible explanation for some discrepancies, which are however well documented only for the ESS data since round 5 (ESS, 2010). For Poland we found inconsistent data with Duncan’s Dissimilarity Index of more than 30% between EU-LFS and ESS 2010 and 2012. The largest deviation found at ISCED levels 2 and 3: In the ESS 2010, 37% are coded to ISCED level 2 and 30% to level 3, whereas in the EU-LFS it is 11% and over 60% respectively. This difference is explained by the decision in the ESS to differentiate between the certificate of basic vocational school before and after an educational reform in 1999. Basic vocational school (“Ukonczona´ szkoła zasadnicza zawodowa”) used to start after 7 years of elementary education before the reform, while in the current system, it starts after 9 years of general education. Before the reform, individuals thus did not complete ISCED level 2 (which typically lasts 9 to 10 years) before entering basic vocational school, but after the reform, they do. This results in ISCED level 2 for the pre-reform vocational qualification, and ISCED level 3 for the post-reform qualification. In the ESS, respondents who achieved the qualification before 2005, when the reform was fully implemented, are therefore coded to ISCED level 2, and all others to ISCED level 3 (ESS, 2010, p. 59). In the EU-LFS, all respondents with this qualification are regarded as reaching ISCED level 3, although the majority still went through the old system. Such reforms, increasing the duration of compulsory schooling, are invisible in official education statistics, which may, from a political point of view, be quite desirable. A similar case is observed for Estonia in the EVS 2008 where the EVS decided to downgrade the basic vocational training to the lower secondary level (“kutseõpe põhihariduse baasil”).

The UK in the ESS is another example of an intended deviation in data processing and of an overrepresentation of the highly educated. Overall, the inconsistency for the UK between EU-LFS and ESS data is 37% in 2008 and around 20% in 2010 and 2012. Focusing on the comparison of the 2008 data there is a discrepancy on ISCED levels 0 and 1; in the ESS around 17% are coded to this category, whereas in the EU-LFS it is less than 1%. This is explained by the ESS decision to classify respondents who finished compulsory schooling without school-leaving certificate into ISCED level 1 instead of ISCED level 2 as is done in the EU-LFS. The main discrepancy of the UK is however on ISCED level 3, in which 11% of the ESS respondents but 41% of the EU-LFS are classified. This inconsistency is caused by the decision of the ESS to put the General Certificate of Secondary Education (GCSE) into ISCED 2, although it is officially mapped to ISCED 3C (ESS, 2010). This latter category describes programmes which do not give access to ISCED level 5, but directly lead to the labour market (or to other programmes at ISCED level 3 or 4). These programmes are thus usually vocational. However, the GCSE is a general school leaving certificate awarded at age 16, which does not specifically prepare for direct labour market entry. In order to improve comparability with other western European countries that offer first school-leaving certificates around age 16 at the end of ISCED level 2, GCSE is classified as ISCED level 2 in the ESS (ESS, 2010; Schneider, 2008a). A further large difference between the two surveys is found at ISCED levels 5 and 6, where around 30% of the respondents in the EU-LFS but over 50% of those in the ESS are found. We cannot identify a systematic measurement or processing error here, and therefore strongly suspect selective nonresponse by education (or, less likely, sampling frame issues).

From the examples of showcards using ambiguous terms, incomplete sets of response categories, harmonisation problems, poor communication, as well as ‘accidental’ and intended misclassification, we can conclude that the education variable is not consistently measured across surveys. However, most of the measurement errors are processing errors, which could even be corrected ex-post. Then, assessing sample representativeness using the corrected education variables would be possible. If the measurement instruments however are the main ‘culprit’, this cannot be repaired ex-post.

Errors of Non-Observation

In some cases, we could not explain the observed inconsistencies of the education distributions even after close examination of the survey instruments and harmonisation routines. Therefore, we now look for further factors influencing sample representativeness. These are, for example, differences *in coverage or sampling frames*, different *sampling designs*, different *survey modes*, as well as *selective nonresponse*.

An example where we think sample representativeness is at risk through the sample design and selective nonresponse is Norway, where we find an inconsistency between the EU-LFS (with mandatory participation in Norway) and the Eurobarometer 73.2 of 2010 of more than 30%. The largest deviation occurs at ISCED levels 5 and 6 to which 37% of the respondents of the EU-LFS and 65% of the Eurobarometer respondents are coded. The EU-LFS uses a random sample from the Norwegian central population register. The Eurobarometer, as in most countries, uses a standard random route procedure by which, in principle, a representative sample can be drawn. However, the success of this approach strongly depends on interviewers implementing it correctly. Here interviewers may have systematically avoided poor neighbourhoods, favoured wealthy ones, or substituted unavailable/ refusing lower educated respondents by willing and available higher educated respondents. Another explanation could be that lower educated may have refused to participate in the Eurobarometer more often. We unfortunately cannot separate the errors due to sampling design from those due to selective nonresponse, also because for the Eurobarometer, response rates are not published. The high inconsistencies for a substantial number of countries between the Eurobarometer and the EU-LFS data are particularly alarming when considering representativeness, however we also found many education measurement problems (as described above) in the Eurobarometer.

Another factor which can influence the representativeness of a sample by introducing differential nonresponse is the *survey mode*. This might explain the high deviation of the education distribution in Finland in the EVS 2008 compared to the EU-LFS of 29%. We found an overrepresentation of higher educated Finnish people in the EVS: over 60% of the respondents stated that they have tertiary education, whereas in the EU-LFS the proportion is 37%. In the EVS 2008, Finland decided to question respondents using a web panel, while all other countries used face-to-face interviews. This Finnish web panel is based on a random selection from earlier telephone or face-to-face samples of which the recruitment criteria are based on figures from Statistics Finland (EVS & GESIS, 2010). However, it seems that this panel is not a representative sample of the Finnish population. In general, web surveys tend to overrepresent highly educated people (Couper, 2000; Dever, Rafferty, & Valliant, 2008).

These examples show that some of the observed inconsistencies are probably caused by errors of non-observation rather than measurement and processing errors. In these cases, we conclude that random route sampling design (via interviewer effects) and selective nonresponse (e. g. if survey modes differ across surveys) might cause the discrepancies, and indeed representativeness is at risk. For those cases it would be possible to design a weighting factor using the education variable based on the distributions of the EU-LFS to correct for the observed inconsistencies, provided we have in fact excluded all measurement sources of the discrepancies of education distributions.

3.5 Conclusions and Recommendations

The aim of this paper was to examine whether the education variable is appropriate for evaluating the realised representativeness of a survey sample. In the first step of the analysis, we detected small median inconsistencies and low variation in the data of EU-SILC and PIAAC as compared with the EU-LFS. We suspect that these surveys use quite similar measurement instruments and coding procedures, as well as state-of-the-art sampling frames and methods. Intermediate median inconsistencies and medium-sized variation are identified when comparing the ESS with the EU-LFS data. Larger median inconsistencies and variation in the distributions are observed for the comparison of EVS, ISSP and Eurobarometer data with the EU-LFS. These could be due to either measurement or representation errors.

Therefore, in the second step, we diagnosed various kinds of measurement errors by having a closer look at the education distributions, measurement instruments and coding (harmonisation) decisions in individual countries, years and surveys with very high inconsistencies between two education distributions. On the measurement side, we find more processing than instrument-related measurement errors, which hints at a potential to correct errors in the data ex-post. Doing so, assessing sample representativeness would become possible. Obviously, these issues imply a lack of substantive comparability of the education variable (Billiet et al., 2009; Ortmanns &

Schneider, 2016). Only for a few cases with large inconsistencies we conclude that these are mostly caused by errors of non-observation alone, so that here the education variable can be used for assessing sample representativeness.

Therefore, there is strong evidence that educational attainment in many cases is not a good variable for evaluating the representativeness of a survey sample. Consequently, the education variable should not be used when designing weights to adjust for nonresponse bias without the necessary measurement comparability checks. The ESS, for instance, since round 5 adjusts education in only three broad education categories to the EU-LFS (Billiet et al., 2009; ESS, 2014a), and they also reversed intended deviations from official ISCED mappings before doing so. From the results of this study we consider this to be quite a suitable solution (which will however result in somewhat less effective nonresponse-adjustment).

One important limitation of this study is that our benchmark data, the EU-LFS, are not free from errors as demonstrated by the example of Iceland. Especially the use of proxy reporting could lead to measurement error in the reference data because proxies may not know the target person’s educational attainment well enough. Another limitation of this study is that errors appearing in every survey are not observed, because the value of Duncan’s Dissimilarity Index will be unremarkable in this case. Also, we could not systematically examine all survey error sources because some are not observable with our data, for example social desirability bias (Biemer & Lyberg, 2003; Tourangeau, Rips, & Rasinski, 2000). Social desirability could e. g. express itself by respondents reporting the level of education required for their current job rather than their actual level of education. This could upwardly bias respondents’ self-reported attainment (Huddy et al., 1997). However, we do not expect that the prevalence of socially desirable responding would differ across the surveys we examine: they are almost all interviewer-administered and thus prone to similar bias. As another example, older respondents may have difficulties remembering their education level. They may also have more difficulty reporting it, especially if formal qualifications have changed over time and the measurement instrument does not mention the names of outdated qualifications explicitly.[9] We used the same age range across surveys to minimize the impact of such issues on our results. These response effects cannot be studied in detail

[9] Educational reforms may actually be one reason for using rather vague terms in education questions, the problematic implications of which we discussed above.

using quantitative data, but call for more systematic cognitive pretesting of education questions in all countries.

To conclude, we would like to make some recommendations for improving the consistency of education data across surveys, to improve its substantive comparability and to facilitate the use of this variable for checking sample quality and constructing weights correcting for nonresponse bias. While some of these recommendations address the survey community as a whole and also international official statistics, others can be implemented by each survey directly.

First of all, surveys need good instruments and showcards which avoid the use of ambiguous terms and unspecific, vague wording, or incomplete sets of response categories. The showcards should instead contain the names of educational qualifications, including formal vocational qualifications, or summary terms that are generally understood by respondents and easily codable to ISCED. Therefore, country teams need the ISCED mappings and guidelines for the development of measurement instruments before developing their questionnaire or should adopt existing instruments from other surveys. Also, more research should be conducted comparatively studying educational systems, qualifications, and careers, including vocational ones, with education measurement in mind.

Secondly, we recommend standardising country-specific education response categories and showcards across surveys in order to elicit more similar kinds of measurement errors in different surveys. No instrument will be without measurement error, but it would be good to produce minimal and consistent errors. Such standardised showcards of course need rigorous testing and regular updates to ensure quality.

Thirdly, we recommend more effective quality assurance and control procedures for background variables and their harmonisation in all surveys. Consistency checks such as those described in this article should be standard for a range of socio-demographic variables, so that especially ‘accidental’ misclassifications can be fixed before data release. Regarding the education variable we strongly recommend ex-post corrections of existing data, and improvements of measurement instruments for future data collections, especially for the Eurobarometer and in the ISSP.

Finally, we would like to question the capability of ISCED to ensure substantive comparability of education data in cross-national surveys. ISCED is, during its development and implementation, vulnerable to political influence, chiefly because education ministries or national statistical institutes determine which national qualifications to map to which ISCED level, and in the latter case, statistical institutes don't always seem to act independently in doing so. At the same time, political education targets that are directly related to ISCED, such as the Europe 2020 goal of reducing the numbers of 'early school leavers' (i. e. students leaving education with less than ISCED level 3) to below 10% (Eurostat, 2016), provide an incentive to classify educational programmes at ISCED level 3 even though ISCED level 2 may be substantively more accurate in terms of ISCED classification criteria.

If the international official statistics community does not achieve stricter quality control of national ISCED mappings, the international survey community may need to find solutions that more reliably produce comparable education data for their own purposes. International academic surveys such as ESS, EVS, ISSP, and the Survey of Health, Ageing and Retirement in Europe (SHARE) could agree on one 'alternative' ISCED scheme and adjust the official ISCED mappings to optimise comparability over time and space. Thereby, these surveys would be more comparable with each other. If this alternative variable is coded in detail, it would still be possible to also derive the official ISCED variable in order to check sample representativeness by comparing with official data. For such an academic survey version of ISCED, good documentation is required and the recodes to the official version would have to be published. The ESS since 2010 has tried to go down this route with a number of surveys following suit - SHARE, and probably also the EVS 2017.

Following these recommendations, the statistical consistency and substantive comparability of cross-national education data could be greatly improved. The education variable in academic surveys could then reliably be used for evaluating the realised representativeness of survey samples.

3.6 Acknowledgements

We are thankful to Carina Cornesse for her helpful advice and suggestions. We also would like to thank the editor and especially the two anonymous reviewers for their questions and remarks which helped us improving this paper. This research was supported by the Leibniz Association through the Leibniz Competition (formerly SAW Procedure); grant number SAW-2013-GESIS-5.

3.7 References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American Time Use Survey: Who is Missing From the Data and How Much Does it Matter? *Public Opinion Quarterly*, 70(5), 676–703. doi:10.1093/poq/nfl037

Bethlehem, J. D. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 275–287). New York: John Wiley & Sons.

Bethlehem, J. D., Cobben, F., & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. Hoboken: John Wiley & Sons.

Bethlehem, J. D., & Schouten, B. (2016). Nonresponse Error: Detection and Correction. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (pp. 558–578). London: SAGE Publications.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Hoboken: John Wiley & Sons.

Billiet, J., Matsuo, H., Beullens, K., & Vehovar, V. (2009). Non-Response Bias in Cross-National Surveys: Designs for Detection and Adjustments in the ESS. *ASK Research & Methods*, 18(1), 3–42.

Billiet, J., Philippens, M., Fitzgerald, R., & Stoop, I. (2007). Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents. *Journal of Official Statistics*, 23(2), 135–162.

Couper, M. P. (2000). Web surveys – A Review of Issues and Approaches. *Public Opinion Quarterly*, 64(4), 464–494. doi: 10.1086/318641

Dever, J. A., Rafferty, A., & Valliant, R. (2008). Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? *Survey Research Methods*, 2(2), 47–62. doi:10.18148/srm/2008.v2i2.128

Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20(2), 210–217. doi:10.2307/2088328

Ehling, M. (2003). Harmonising Data in Official Statistics: Development, Procedures, and Data Quality. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book of Demographic and Socio-Economic Variables* (pp. 17–31). New York: Kluwer Academic/Plenum Publishers.

ESS. (2010). *European Social Survey Round 5 Technical Report (2010/2011) Appendix A1 Education* (4.0 ed). Retrieved from: http://www.europeansocialsurvey.org/docs/round5/survey/ESS5_appendix_a1_e04_0.pdf (accessed September 2016); current version 4.3 can be retrieved from: https://www.europeansocialsurvey.org/docs/round5/survey/ESS5_appendix_a1_e04_2.pdf (accessed April 2020)

ESS. (2014a). *Documentation of ESS Post-Stratification Weights*. Retrieved from: https://www.europeansocialsurvey.org/docs/methodology/ESS1-5_post_stratification_weights_documentation.pdf (accessed April 2020)

ESS. (2014c. *European Social Survey Round 6 Technical Report (2012/2013)* (2.2 ed.). Retrieved from:

http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_data_documentation_report_e02_2.pdf (accessed September 2016); current version 2.4 can be retrieved from: https://www.europeansocialsurvey.org/docs/round6/survey/ESS6_data_documentation_report_e02_4.pdf (accessed April 2020)

Eurostat. (2006). *LFS Users Guide. Labour Force Survey Anonymised Data Sets*. Luxembourg: European Commission.

Eurostat. (2008a). *Description of SILC User Database Variables: Cross-Sectional and Longitudinal*. Luxembourg: European Commission.

Eurostat. (2009a). *Description of SILC User Database Variables: Cross-Sectional and Longitudinal*. Luxembourg: European Commission.

Eurostat. (2013f). *Labour Force Survey in the EU, Candidate and EFTA Countries – Main Characteristics of National Surveys, 2012*. Luxembourg: European Commission. Retrieved from: https://ec.europa.eu/eurostat/documents/3888793/5856085/KS-TC-13-003-EN.PDF/15feeb32-717d-4326-8e7b-da6e2530ed4c (accessed April 2020)

Eurostat. (2016). *Europe 2020 Indicators – Education*. Retrieved from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Europe_2020_indicators_-_education (accessed April 2020)

Eurostat, & OECD. (2014). *Joint Eurostat-OECD Guidelines on the Measurement of Educational Attainment in Household Surveys*. Version of September 2014. Retrieved from: http://ec.europa.eu/eurostat/documents/1978984/6037342/Guidelines-on-EA-final.pdf (accessed April 2020)

EVS, & GESIS. (2010). EVS 2008 *Method Report: Documentation of the Full Data Release 30/11/10*. GESIS-Technical Reports 2010/17. Cologne: GESIS Data Archive for the Social Sciences.

Gelman, A., & Carlin, J. B. (2002). Poststratification and Weighting Adjustments. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 289–302). New York: Wiley & Sons.

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646–675. doi:10.1093/poq/nfl033

Groves, R. M. & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken: John Wiley & Sons.

Groves, R. M., & Lyberg, L. E. (2010). Total Survey Error – Past, Present, and Future. *Public Opinion Quarterly*, 74(5), 849–879. doi:10.1093/poq/nfq065

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. *Public Opinion Quarterly*, 72(2), 167–189. doi: doi.org/10.1093/poq/nfn011

Heerwegh, D., Abts, K., & Loosveldt, G. (2007). Minimizing Survey Refusal and Noncontact Rates: Do Our Efforts Pay off? *Survey Research Methods*, 1(1), 3–10. doi: 10.18148/srm/2007.v1i1.46

Hoffmeyer-Zlotnik, J. H. P., & Warner, U. (2014). *Harmonising Demographic and Socio-Economic Variables for Cross-National Comparative Survey Research.* Dordrecht: Springer Science+Business Media.

Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P. J., & Pugliani, P. (1997). The Effect of Interviewer Gender on the Survey Response. *Political Behavior,* 19(3), 197–220. doi:10.1023/A:1024882714254

ISSP. (2015). *Study Monitoring Report 2012, Family & Changing Gender Roles IV*.

ISSP - Demographic Methods Group. (2010). *ISSP Background Variables Guidelines* (press release). Retrieved from: http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/members/codinginfo/BV_guidelines_for_issp2011.pdf (accessed September 2016); current version of 2017 can be retrieved from: https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/members/codinginfo/BV_guidelines_for_issp2018.pdf (accessed April 2020)

Kalmijn, M. (2003). Country Differences in Sex-Role Attitudes: Cultural and Economic Explanations. In W. A. Arts, L. C. J. M. Halman & J. A. P. Hagenaars (Eds.), *The Cultural Diversity of European Unity. Findings, Explanations and Reflections from the European Values Study* (pp. 311–337). Leiden: Koninklijke Brill NV.

Kamtsiuris, P., Lange, M., Hoffmann, R., Schaffrath Rosario, A. S., Dahm, S., Kuhnert, R., & Kurth, B.-M. (2013). Die erste Welle der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). Stichprobendesign, Response, Gewichtung und Repräsentativität. *Bundesgesundheitsblatt* 2013. 56: 620–630. doi:10.1007/s00103-012-1650-9

Kieffer, A. (2010). Measuring and Comparing Levels of Education: Methodological Problems in the Classification of Educational Levels in the European Social Surveys and the French Labor Force Surveys. *Bulletin de Méthodologie Sociologique*, 107(1), 49–73. doi:10.1177/0759106310369974

Kingston, P. W., Hubbard, R., Lapp, B., Schroeder, P., & Wilson, J. (2003). Why Education Matters. *Sociology of Education*, 76(1), 53–70. doi: 10.2307/3090261

Koch, A., Halbherr, V., Stoop, I. A. L., & Kappelhof, J. W. S. (2014). *Assessing ESS Sample Quality by Using External and Internal Criteria.* Mannheim: European Social Survey, GESIS. Retrieved from: https://www.europeansocialsurvey.org/docs/round5/methods/ESS5_sample_composition_assessment.pdf (accessed April 2020)

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T. M., Casas-Cordero, C., Lemay, M., Peytchev, A., Groves, R. M. & Raghunathan, T. E. (2010). Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-Response: Examples From Multiple Surveys. *Journal of the Royal Statistical Society*, 173(2), 389–407.

Kruskal, W., & Mosteller, F. (1979a). Representative Sampling, I: Non-Scientific Literature. *International Statistical Review,* 47(1), 13–24. doi:10.2307/1403202

Kruskal, W., & Mosteller, F. (1979b). Representative Sampling, II: Scientific Literature, Excluding Statistics. *International Statistical Review*, 47(2), 111–127. doi:10.2307/1402564

Kruskal, W., & Mosteller, F. (1979c). Representative Sampling, III: The Current Statistical Literature. *International Statistical Review*, 47(3), 245–265. doi: 10.2307/1402647

Lundquist, P., & Särendal, C.-E. (2013). Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey. *Journal of Official Statistics*, 29(4), 557–582. doi:10.2478/jos-2013-0040.

Nishimura, R., Wagner, J., & Elliott, M. R. (2016). Alternative Indicators for the Risk of Non-Response Bias: A Simulation Study. *International Statistical Review*, 84(1), 43– 62. doi:10.1111/insr.12100

OECD. (2016). Technical Report of the Survey of Adult Skills (PIAAC) (2nd ed.). Retrieved from: https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf (accessed April 2020)

Ortmanns, V., & Schneider, S. L. (2016). Harmonization still failing? Inconsistency of education variables in cross-national public opinion surveys. *International Journal of Public Opinion Research*, 28(4), 562–82. doi:10.1093/ijpor/edv025

Peytcheva, E., & Groves, R. M. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. Journal of Official Statistics, 25(2), 193–201.

Schneider, S. L. (2008a). The Application of the ISCED-97 to the UK's Educational Qualifications. In S. L. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries* (pp. 281–300). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES).

Schneider, S. L. (2008b). *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). Oxford: University of Oxford, Nuffield College. Retrieved from: http://ora.ox.ac.uk/objects/uuid:15c39d54-f896-425b-aaa8-93ba5bf03529 (accessed April 2020)

Schneider, S. L. (2013). The International Standard Classification of Education 2011. In G. E. Birkelund (Ed.), *Class and Stratification Analysis* (pp. 365–379). Bingley: Emerald.

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. Fu (Eds.), The SAGE Handbook of Survey Methodology (pp. 288–307). London: SAGE Publications.

Schouten, B., Cobben, F., & Bethlehem, J. D. (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology*, 35(1), 101–113.

Smith, T. W. (1995). Some Aspects of Measuring Education. *Social Science Research*, 24(3), 215–242. doi:10.1006/ssre.1995.1008

Struminskaya, B., Kaczmirek, L., Schaurer, I., & Bandilla, W. (2014). Assessing Representativeness of a Probability-Based Online Panel in Germany. In M.

Callegaro, P. J. Lavrakas, J. A. Krosnik, R. Baker, J. D. Bethlehem, & A. S. Göritz (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 61–85). West Sussex: John Wiley & Sons.

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

UNESCO-UIS. (2006). *ISCED 1997 International Standard Classification of Education*. Montreal: UNESCO-UIS. Retrieved from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf (accessed April 2020)

Data References

ESS. (2012a). *European Social Survey, Round 4, Data 2008.* Data File Version 4.1. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2012b*). European Social Survey, Round 5, Data 2010.* Data File Version 3.0. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2014b). *European Social Survey, Round 6, Data 2012.* Data File Version 2.0. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

European Commission. (2012). *Eurobarometer 73.2+73.3 (2-3 2010)*. ZA5236. Data File Version 2.0.1. Cologne: GESIS Data Archive. doi:10.4232/1.11473

European Commission. (2014). *Eurobarometer 75.4 (2011)*. ZA5564, Data File Version 3.0.1. Cologne: GESIS Data Archive. doi:10.4232/1.11851

Eurostat. (2008b). *European Union Statistics on Income and Living Conditions (EU-SILC) 2008*. User Database. Data File Version 2008-3.

Eurostat. (2009b). *European Union Statistics on Income and Living Conditions (EU-SILC) 2009*. User Database. Data File Version CROSS-2009-4.

Eurostat. (2010). *European Union Statistics on Income and Living Conditions (EU-SILC) 2010*. User Database. Data File Version CROSS-2010-3.

Eurostat. (2011). *European Union Statistics on Income and Living Conditions (EU-SILC) 2011*. User Database. Data File Version CROSS-2011-1.

Eurostat. (2012). *European Union Statistics on Income and Living Conditions (EU-SILC) 2012*. User Database. Data File Version 2012.

Eurostat. (2013a). E*uropean Union Labour Force Survey (EU-LFS) 2008*. User Database. Data File Version 2013.

Eurostat. (2013b). *European Union Labour Force Survey (EU-LFS) 2009*. User Database. Data File Version 2013.

Eurostat. (2013c). *European Union Labour Force Survey (EU-LFS) 2010.* User Database. Data File Version 2013.

Eurostat. (2013d). *European Union Labour Force Survey (EU-LFS) 2011*. User Database. Data File Version 2013.

Eurostat. (2013e). *European Union Labour Force Survey (EU-LFS) 2012*. User database. Data File Version 2013.

EVS. (2011). *European Values Study 2008, 4th wave. Integrated Dataset*. ZA4800. Data File Version 3.0.0. Cologne: GESIS Data Archive. doi: 10.4232/1.11004

ISSP - Research Group (2013). *International Social Survey Programme: Health and Health Care - ISSP 2011*. ZA5800. Data File Version 2.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.11759

ISSP - Research Group (2014). *International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012*. ZA5900. Data File Version 2.0.0, Dataset Hungary: ZA5949. Data File Version 1.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.12022; Dataset Hungary doi:10.4232/1.12117

OECD. (2013). *Programme for the International Assessment of Adult Competencies (PIAAC)* 2011 - *Public Use File*. Data File Version 1.0.

3.8 Supplemental Material

Table A3.1 Categories and recodes of the education variables across surveys into 5-level version of ISCED 97

	5-level version of ISCED97		EU-LFS		EU-SILC		PIAAC		Eurobarometer		ESS until 2008		ESS since 2010		EVS
1	Pre-primary and primary or first stage of basic education	0	No formal education or below ISCED 1	0	Pre-primary education	1	Primary or less (ISCED 1 or less)	0	Pre-primary education	1	Less than lower secondary education (ISCED 0-1)	0	Not completed ISCED level 1	0	Pre-primary education or none education
		11	ISCED 1	1	Primary education			1	Primary education or first stage of basic education			113	ISCED 1, completed primary education	1	Primary education or first stage of basic education
												129	Vocational ISCED 2C < 2 years, no access ISCED 3		
2	Lower secondary or second stage of basic education	21	ISCED 2	2	Lower secondary education	2	Lower secondary (ISCED 2, ISCED 3C short)	2	Lower secondary or second stage of basic education	2	Lower secondary education completed (ISCED 2)	212	General/pre-vocational ISCED 2A/2B, access ISCED3 vocational	2	Lower secondary or second stage of basic education
												213	General ISCED 2A, access ISCED 3A general/all 3		
												221	Vocational ISCED 2C >= 2 years, no access ISCED 3		
												222	Vocational ISCED 2A/2B, access ISCED 3 vocational		
												223	Vocational ISCED 2, access ISCED 3 general/all		
		22	ISCED 3c (shorter than 2 years)									229	Vocational ISCED 3C < 2 years, no access ISCED 5		
3	(Upper) Secondary education	30	ISCED 3 (without distinction a, b or c possible, 2 years and more)	3	(Upper) Secondary education	3	Upper secondary (ISCED 3A-B, C long)	3	(Upper) secondary education	3	Upper secondary education completed (ISCED 3)			3	(Upper) secondary education
		31	ISCED 3c (2 years and more)									311	General ISCED 3 >=2 years, no access ISCED 5		

5-level version of ISCED97	EU-LFS	EU-SILC	PIAAC	Eurobarometer	ESS until 2008	ESS since 2010	EVS
						321 Vocational ISCED 3C >= 2 years, no access ISCED 5	
	32 ISCED 3 a, b					312 General ISCED 3A/3B, access ISCED 5B/lower tier 5A	
						322 Vocational ISCED 3A/3B, access 5B/lower tier 5A	
						313 General ISCED 3A, access upper tier ISCED 5A/all 5	
						323 Vocational ISCED 3A, access upper tier ISCED 5A/all 5	
4 Post-secondary non-tertiary education	43 ISCED 4 (without distinction a, b or c possible)	4 Post-secondary non-tertiary education	4 Post-secondary, non-tertiary (ISCED 4A-B-C)	4 Post-secondary, non-tertiary education	4 Post-secondary non-tertiary education completed (ISCED 4)		4 Post-secondary non-tertiary education
	42 ISCED 4c					421 ISCED 4 programmes without access ISCED 5	
	41 ISCED 4a, b					412 General ISCED 4A/4B, access ISCED 5B/lower tertiary 5A	
						413 General ISCED 4A, access upper tier ISCED 5A/all 5	
						422 Vocational ISCED 4A/4B, access ISCED 5B/lower tertiary 5A	
						423 Vocational ISCED 4A, access upper tier ISCED 5A /all 5	
5 First and second stage of tertiary education	51 ISCED 5b	5 1st & 2nd stage of tertiary education	8 Tertiary - bachelor/ master/research degree (ISCED 5A/6)	5 First stage of tertiary education	5 Tertiary education completed (ISCED 5-6)	510 ISCED 5A short, intermediate/academic/general tertiary below	5 First stage of tertiary education

5-level version of ISCED97	EU-LFS	EU-SILC	PIAAC	Eurobarometer	ESS until 2008	ESS since 2010	EVS
			5 Tertiary – professional degree (ISCED 5B)			520 ISCED 5B short, advanced vocational qualifications	
	52 ISCED 5a		6 Tertiary – bachelor degree (ISCED 5A)			620 ISCED 5A medium, bachelor/equivalent from upper/single tertiary	
						710 ISCED 5A long, master/equivalent from lower tertiary	
			7 Tertiary – master/research degree (ISCED 5A/6)			720 ISCED 5A long, master/equivalent from upper/single tertiary	
	60 ISCED 6			6 Second stage of tertiary education		800 ISCED 6, doctoral degree	6 Second stage of tertiary education

Data sources:
EU-LFS 2008-2012 (second quarter used only), files from Eurostat, data file versions 2013, variable HATLEVEL;
EU-SILC 2008-2012, files from Eurostat, data file versions: CROSS 2008-3, CROSS-2009-4, CROSS-2010-3, CROSS-2011-1, CROSS 2012, variable PE040;
PIAAC 2011, file from OECD, data file version 2013, variable edcat7;
Eurobarometer 73.2 & 73.3 (2010) files from Eurostat, data file versions 2.0.1, variable v362; Eurobarometer 75.4 (2011), files from Eurostat, data file version 3.0.1, variable v105; ESS 2002-2006, data file versions: 6.2 (2002), 3.3 (2004), 3.4 (2006), 4.1 (2008), variable edulvla;
ESS 2008-2012, data file versions: 3.0. (2010), 2.0 (2012), variable edulvlb; EVS 2008, data file version 3.0.0, variable v336

Table A3.2 Categories and recodes of the education variables in ISSP and EU-LFS into 4-level version of ISCED 97

	4-level version of ISCED 97	EU-LFS		ISSP since 2011	
1	Pre-primary and primary or first stage of basic education	0	No formal education or below ISCED 1	0	No formal education
		11	ISCED 1	1	Primary school
2	Lower secondary or second stage of basic education	21	ISCED 2	2	Lower secondary (secondary education completed that does not allow entry to university: end of obligatory school but also short programs (less than 2 years))
		22	ISCED 3c (shorter than 2 years)		
3	(Upper) Secondary education and post-secondary non-tertiary education	32	ISCED 3 a, b	3[a]	Upper secondary (programs that allow entry to university)
		41	ISCED 4 a, b		
		30	ISCED 3 (without distinction a, b or c possible, 2 years and more)	4	Post-secondary, non-tertiary (other upper secondary programs toward the labour market or technical formation)
		43	ISCED 4 (without distinction a, b or c possible)		
		31	ISCED 3c (2 years and more)		
		42	ISCED 4c		
4	First and second stage of tertiary education	51	ISCED 5b	5	Lower level tertiary, first stage (also technical schools at a tertiary level)
		52	ISCED 5a	6	Upper level tertiary (Master, Dr.)
		60	ISCED 6		

Notes:[a] ISCED 3B and 4B are included in ISSP DEGREE variable category 4, not 3, which cannot be differentiated in the ESS. Therefore ISCED 3 and 4 are summarized.

Table A3.3 Duncan's Dissimilarity Index for educational attainment distributions across surveys and years per country

Survey	SILC-LFS					PIAAC[c]-LFS	EB-LFS			ESS-LFS			EVS-LFS	ISSP-LFS[d]		Weighted mean
Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2011**	**2010(1)**	**2010(2)**	**2011**	**2008**	**2010**	**2012**	**2008**	**2011**	**2012**	
AT	3.3	1.5	1.9	1.4	1.8	1.6	17.2	14.4	12.3				12.9		**50.5**	16.3
BE[a]	6.1	6.5	3.9	4.9	6.3	8.4	16.3	18.2	8.0	6.2	12.3	11.1	8.2	6.9		8.8
BG	4.6	2.7	1.2	1.5	2.0		9.6	8.9	8.0	1.6	3.6	3.5	4.9	4.7	5.1	4.8
CH				16.1	16.2					4.9	3.0	3.5	10.8	6.9	4.0	9.1
CY	2.6	3.7	4.5	3.2	4.2	7.7	15.4	20.7	24.9	8.1	5.8	4.6	18.9			11.3
CZ	0.4	0.8	0.9	0.6	0.8	5.9	10.4	11.1	7.1	5.8	12.4	13.6	3.2	24.4	24.6	9.1
DE	1.4	2.3	3.6	3.1	2.4	6.9	25.6	19.3	26.4	9.7	7.9	10.5	10.5	5.0	6.7	9.8
DK	3.2	2.2	2.6	2.3	6.1	10.2	12.0	16.1	11.9	17.2	19.3	20.3	10.7	33.5	32.2	14.9
EE	5.3	5.1	4.7	5.8	4.0	5.5	21.8	24.6	14.4	11.0	10.2	11.4	34.6			15.2
ES	4.0	4.0	3.2	2.9	2.8	5.8	22.8	22.1	25.2	10.2	10.1	12.6	20.4	9.3	10.0	12.3
FI	9.8	8.6	7.7	6.5	5.6	7.6	17.0	18.5	16.6	5.5	15.2	14.0	**29.1**	9.7	11.6	14.0
FR	4.0	2.9	8.2	8.1	7.6	2.7	9.9	8.3	10.0	9.4	11.5	13.3	14.1	27.7	**29.2**	12.0
GB[b]	11.0	9.1	8.0	9.4	9.6	12.3	20.8	18.1	19.2	**36.7**	21.5	17.8	23.5	20.8	13.9	17.9
GR	6.5	8.1	8.6	9.0	7.5		12.6	10.7	10.6	19.3	13.4		15.2			12.7
HR				5.9	4.4					10.6	7.0		6.5	9.2	11.5	7.7
HU	3.2	4.4	3.3	4.4	3.1		42.8	38.8	36.7	5.5	5.4	3.3	5.2		30.0	16.6
IE	5.9	5.0	6.0	7.4	6.2	8.3	17.6	19.9	17.2	18.8	7.0	8.9	12.4		12.7	11.6
IS	9.8	10.7	8.4	28.5	21.8		19.5	17.7				24.1	17.6		19.7	19.8
IT	3.1	3.6	3.3	2.9	3.2	7.8	16.9	15.7	17.2			12.6	16.1	14.1		11.7
LT	3.2	5.4	3.5	5.1	5.7		11.4	11.5	13.7		18.5	11.4	21.2	14.4	10.6	13.1
LU	7.7	15.7	14.8	13.5	15.0		8.9	7.4	4.8				7.1			9.2
LV	2.5	4.3	4.7	3.0	3.9		31.6	32.5	**35.4**	7.6			20.7		1.5	13.3
MT		8.6	7.5	3.9	3.0		41.6	37.1	39.6							22.6

Survey	SILC-LFS					PIAAC[c]-LFS	EB-LFS			ESS-LFS			EVS-LFS	ISSP-LFS[d]		Weighted mean
Year	2008	2009	2010	2011	2012	2011	2010(1)	2010(2)	2011	2008	2010	2012	2008	2011	2012	
NL	1.9	3.0	4.0	5.5	4.9	4.7	38.3	39.8	**42.4**	12.1	17.9	14.1	23.0	25.7		18.7
NO	2.3	0.5	0.8	1.4	5.6	13.7	**33.1**	26.1		13.7	17.0	13.2	19.5	20.1	20.4	16.6
PL	13.2	13.1	12.3	11.3	11.3	7.7	4.8	3.3	6.2	4.9	**32.9**	30.9	14.1	3.8	9.4	11.4
PT	2.2	3.1	2.9	1.6	2.8		8.3	7.3	13.1	3.4	3.6	5.7	5.1	6.3		5.5
RO	4.6	3.4	2.3	3.0	3.0		14.1	12.4	9.3	10.6			10.1			9.0
SE	7.2	6.9	7.2	7.7	7.8	2.9	21.1	23.5	25.3	13.6	9.8	10.4	22.0	24.1	27.6	15.4
SI	1.4	2.1	1.7	2.7	1.6		5.8	2.2	3.3	6.0	3.8	3.7	26.6	19.8	23.7	11.7
SK	5.9	7.0	5.5	6.4	7.0	10.8	4.4	4.0	4.5	8.7	10.4	11.1	3.9	31.3	32.0	11.2
Median	4.0	4.4	4.0	4.9	4.9	7.7	16.9	17.7	13.7	9.4	10.3	11.4	14.1	14.3	13.9	12.0
Mean	4.9	5.3	5.1	6.1	6.0	7.3	18.3	17.6	17.2	10.4	11.6	11.9	14.9	15.9	18.4	12.7
Min	0.4	0.5	0.8	0.6	0.8	1.6	4.4	2.2	3.3	1.6	3.0	3.3	3.2	3.8	1.5	4.8
Max	13.2	15.7	14.8	28.5	21.8	13.7	42.8	39.8	42.4	36.7	32.9	30.9	34.6	33.5	50.5	22.6

Notes :[a] For PIAAC and EU-LFS only Flanders, excluding Wallonia and Brussels; for ISSP and EU-LFS Flanders and Brussels, excluding Wallonia; [b] For PIAAC and EU-LFS only England and Northern Ireland, excluding Scotland and Wales; for ISSP and EU-LFS excluding Northern Ireland; [c] For PIAAC, DE and AT use age group 25 to 65 instead of 25 to 64;[d] For ISSP, adapted ISCED97_4 level is used (see Table A3.2);
Cells shaded in grey show discrepancies above 25. Cells with bold print are included in Figure 3.2 and discussed in more detail in section 3.4.2.
Data sources:
EU-LFS 2008-2012 (second quarter used only), files from Eurostat, data file versions 2013, variable HATLEVEL, weighted using variable coeff;
EU-SILC 2008-2012, files from Eurostat, data file versions: 2008-3, CROSS-2009-4, CROSS-2010-3, CROSS-2011-1, CROSS 2012, variable PE040, weighted using variable PB040;
PIAAC 2011, file from OECD, data file version 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB); Eurobarometer 73.2 & 73.3 (2010) files from Eurostat, data file versions 2.0.1, variable v362;
Eurobarometer 75.4 (2011) files from Eurostat, data file version 3.0.1, variable v105, data weighted to correct regional oversampling for Germany and the UK; ESS 2008, data file version 4.1, variable edulvla;
ESS 2010-2012, data file versions: 3.0. (2010), 2.0 (2012) variable edulvlb, weighted using variable dweight;
EVS 2008, data file version 3.0.0, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK;
ISSP 2011-2012, data file versions: 3.0.0 (2011), 2.0.0 (2012) and 1.0.0 (2012) for Hungary, variable DEGREE, data weighted to corret regional oversampling for Germany;
Only respondents aged 25-64 for all surveys.

Paper III

Explaining Inconsistencies in the Education Distributions of Ten Cross-National Surveys – The Role of Methodological Survey Characteristics

4 Explaining Inconsistencies in the Education Distributions of Ten Cross-National Surveys – The Role of Methodological Survey Characteristics[10]

Surveys measuring the same concept using the same measure on the same population at the same point in time should result in highly similar results. If this is not the case, this is a strong sign of lacking reliability, resulting in non-comparable data across surveys. Looking at the education variable, previous research has identified inconsistencies in the distributions of harmonised education variables, using the International Standard Classification of Education (ISCED), across surveys within the same countries and years. These inconsistencies are commonly explained by differences in the measurement, especially in the response categories of the education question, and in the harmonisation when classifying country-specific education categories into ISCED. However, other methodological characteristics of surveys, which we regard as 'containers' for several characteristics, may also contribute to this finding. We compare the education distributions of nine cross-national surveys with the European Union Labour Force Survey (EU-LFS), which is used as benchmark. This study analyses 15 survey characteristics to better explain the inconsistencies. The results confirm a predominant effect of the measurement instrument and harmonisation. Different sampling designs also explain inconsistencies, but to a lesser degree. Finally, we discuss the results and limitations of the study and provide ideas for improving data comparability.

Keywords: comparative research, cross-national surveys, survey characteristics, education

4.1 Introduction

Education is a key socio-demographic variable that is measured in nearly every survey (Smith, 1995). Education is central in social stratification research, for instance, when analysing educational inequalities and how social class of origin affects education (Breen & Jonsson, 2000, 2005; Müller & Karle, 1993), or when analysing returns to education for example how education determines individuals' income and socio-economic status (Becker 1993; Blau & Duncan, 1967; Bol & van de Werfhorst, 2013). Outside of stratification research, the education variable is an important proxy variable for another concept, such as cognitive competencies, and it is also widely used as a background or control variable. Quite often studies find a substantial effect of the education variable, for example when analysing values and behaviours, e.g. political

[10] This paper is published as: Ortmanns, V. (2020). Explaining Inconsistencies in the Education Distributions of Ten Cross-National Surveys – the Role of Methodological Survey Characteristics. *Journal of Official Statistics*, 36(2): 379-409. doi:10.2478/JOS-2020-0020

attitudes or voting behaviours (Bekhuis, Lubbers, & Verkuyten, 2014; Weakliem, 2002), gender role attitudes (Bolzendahl & Myers, 2004; Kalmijn, 2003) or attitudes towards minorities and immigrants (Coenders & Scheepers, 2003; Semyonov, Raijman, & Gorodzeisky, 2008; Hyman & Wright, 1979). In survey methodological research, the education variable is important because together with sex and age, it is often used to assess the comparability of the survey data, for instance with official data sources (Peytcheva & Groves, 2009). Furthermore, education is often included when calculating post-stratification weights, which aim to correct for non-sampling errors such as nonresponse and may decrease the variance of a survey's estimate (e.g. ESS, 2014b). Clearly the education variable is important for different purposes, and ideally should be of high measurement quality.

Previous research compared the education distribution across surveys within countries and years to assess how reliable the distribution of education is measured across surveys and thus how comparable the data are. For identical populations and time points, one would expect only minimal variation in the data. However, studies repeatedly revealed inconsistencies in education distributions across surveys even when they use the same harmonised education variables (Kieffer, 2010; Ortmanns & Schneider, 2016a, 2016b; Schneider, 2009). These discrepancies indicate that the data cannot be comparable in some way. However, especially for cross-national comparative research, data need to be comparable. In more detail, the study of Kieffer (2010) observed discrepancies when comparing the distribution for the European Social Survey (ESS) with the EU-LFS for France. Large deviations were identified for the first three waves of the ESS in 2002, 2004 and 2006; while for 2008, the deviation was smaller. Schneider (2009), who compared data from 2002 to 2007, also identified inconsistencies when comparing the distributions for most countries in the European Union Statistics on Income and Living Conditions (EU-SILC), and in the ESS with the EU-LFS. Ortmanns and Schneider (2016b) replicated and extended this work by comparing education distributions for European countries included in four public opinion surveys between 2008 and 2012. They analysed the Eurobarometer, the European Values Study (EVS), the International Social Survey Programme (ISSP) and the ESS, which was used as the reference survey. In the most comprehensive study to date, Ortmanns and Schneider (2016a) analysed seven cross-national survey programmes, again looking at the period 2008 to 2012. They included OECD's Programme for the International Assessment of Adult Competencies (PIAAC), EU-

SILC, Eurobarometer, ESS, EVS and ISSP, and compared the education distributions for the same countries and years to the respective distribution in the EU-LFS. Since this study is the basis for this article, we will briefly summarise the main result to illustrate the problem. Ortmanns and Schneider (2016a) found that on average, 13% of respondents would have to change education categories to achieve an equal distribution with the EU-LFS. They also found substantial variation across surveys, ranging from 1% to almost 50%. These inconsistencies cannot reflect actual differences in the education distribution because it should be rather stable for the same country and year. Instead, these inconsistencies indicate a severe problem with data comparability across surveys, and thus methodological differences between the surveys must explain the observed deviations.

To date, researchers explain those inconsistencies commonly by differences in the measurement of education or the way country-specific response categories are classified into the International Standard Classification of Education (ISCED) (Kieffer, 2010; Ortmanns & Schneider, 2016a, 2016b). However, we cannot be sure that these are the only or most important factors just because they can be observed easily and are reported more often. Ortmanns and Schneider (2016a) identify single cases where they hypothesise that differences in the survey characteristics such as data collection modes, sampling designs, as well as selective unit nonresponse might also explain the inconsistencies because they do not find any problem in the measurement or the assignment of ISCED codes. Those survey characteristics refer to methodological aspects of a survey, and they differ across surveys because they are designed and organised differently, and apply different methodological standards. Thus, the survey characteristics influence the quality of the survey and its data. To systematically analyse and test the impact of surveys' methodological characteristics, we need an in-depth, quantitative and comprehensive analysis.

Such an analysis is conducted in this study, which analyses the impact of 15 survey characteristics and how they contribute to inconsistent education distributions across surveys within countries and years. As a starting point, we use the results from Ortmanns and Schneider (2016a), comparing the education distributions of six surveys with the EU-LFS for the years 2008 to 2012. We further extend the range of cross-national surveys by adding the Adult Education Survey (AES), the European Quality of Life Survey (EQLS), and the European Working Condition Survey (EWCS). Hence, this study compares the education distributions of ten cross-national surveys for 31 European countries. The research question is: Can survey characteristics explain the inconsistencies identified in the education distributions across surveys? Thirteen hypotheses are formulated and tested by estimating regression models.

The next section describes these cross-national surveys and how they measure education. It also introduces the challenges of comparing the education distributions and the survey characteristics across surveys. In the third section, we present several different survey characteristics and derive our hypotheses regarding their contribution to the inconsistencies in education distributions. We use the Total Survey Error (TSE) framework (Groves et al., 2009; Groves & Lyberg, 2010) to structure this presentation. In the fourth section, the variables and methods are described, before presenting the results in section five. In section six, we discuss the results and limitations of the study and provide ideas for improving data comparability.

4.2 The Cross-National Surveys and Their Education Measures

4.2.1 The Cross-National Surveys Covered in This Study

This study compares the education distributions of nine large-scale, cross-national surveys to the EU-LFS (Eurostat 2016a, 2016b, 2016c, 2016d), which we use as a benchmark, and estimates the impact of survey characteristics on the observed inconsistencies in the education distributions. To better understand the challenges of estimating the impact of survey characteristics when using the EU-LFS as a benchmark, and the consequences for the design of this study, we start with a brief description of the survey programmes.

Since the beginning of the EU-LFS in the 1970s, it has provided official household data for monitoring employment and unemployment in all EU countries and some European non-EU countries. The large number of countries included in the survey, the large sample sizes, the relatively high response rates and the probability-based sampling should produce representative high-quality data and thus an accurate education distribution for each country. Furthermore, the EU-LFS provides annual data, is fairly well documented, and it applies the official ISCED mappings. Thus, it is the most authoritative source regarding education data in Europe. Statistics based on the EU-LFS are, for instance, used in the annual OECD reports “Education at a Glance” (e.g., OECD, 2015, 2016, 2017). EU-LFS data are also used when defining goals of the

Europe 2020 strategy to enhance participation in education in all European countries (Eurostat, 2019). The distribution of the EU-LFS education variable is also used as reference for other surveys, such as the ESS, when comparing or weighting data (ESS, 2014a, 2014b). We are not aware of another official cross-national survey that fulfils all these criteria. Census data, for instance, typically do not provide harmonised data, which can be used for international comparisons; those have to be generated by the researcher herself. More important, to our knowledge, researchers cannot simply access an integrated dataset of the latest official census data for all European countries. Hence, we use the EU-LFS as the benchmark survey in this study.

However, the EU-LFS also does not reflect the ‘true’ education distributions of the countries. The EU-LFS is an output harmonised survey, meaning the national surveys, to a large extent, are independent of each other and follow different national regulations. This applies for nearly all survey characteristics. Survey participation, for instance, is mandatory in roughly half of the countries the EU-LFS, but it is voluntary for the other countries. The response rate also varies greatly across countries between 30% and 98%. Furthermore, the countries use different sampling designs (simple or complex designs), as well as different modes of data collection (face-to-face, telephone, self-administered or mixed-mode). Of course, some guidelines and rules are specified to achieve as much comparable statistics as possible across countries, but the national survey designs entail quite different survey characteristics across the countries participating in the EU-LFS. This considerable variation in the survey characteristics of the EU-LFS forces us to analyse the impact of these survey characteristics with a rather broad approach. Therefore, we cannot assess which data collection mode causes more or fewer inconsistencies in the education distribution. Instead, we can only analyse whether mode differences between the survey in question and the EU-LFS affect the education distribution. As indicated, this applies to all survey characteristics; thus, we can only assess whether differences in the survey characteristics can contribute to inconsistencies in the education distributions across surveys within the same countries and years. This has to be considered when developing the hypotheses, and it adds complexity when operationalising the variables and interpreting the results. Nevertheless, it is important to mention that for all surveys, good documentation of the survey characteristics is an essential precondition for this study to identify how the survey characteristics differs across surveys within the same countries and years.

Another official survey included in this analysis is the EU-SILC (Eurostat, 2010). It was launched in 2003 with the aim of providing cross-sectional and longitudinal official micro-data on income, poverty, social exclusion, as well as living and housing conditions in the EU. We also analyse data from PIAAC (OECD, 2013) and the AES (Eurostat, 2011), which focus on education. PIAAC is an OECD survey that measures adults' general basic skills, and first collected data in 2011/ 12 across OECD countries. The AES is a Eurostat survey that covers participation in formal and non-formal education and training of adults in EU countries. It began in 2007 and has been repeated nearly every fifth year. We also analyse data of the Eurobarometer (European Commission, 2012), which was set up by the European Commission in the 1970s to monitor public attitudes towards the EU and related topics in all Member States. So far, the ISCED classification has only been implemented in three Eurobarometer studies, two of them have been conducted in 2010 and one in 2011. Additionally, we also analyse date from the EQLS (Eurofound, 2014) and the EWCS (Eurofound, 2011). Both surveys include all EU countries and they are funded through Eurostat and realised by Eurofound. The EQLS is conducted every four to five years since it was launched in 2003. The survey questions European citizens on general circumstances of their lives, such as employment, income, housing, family, happiness, and well-being. The EWCS was launched in 2005 and also runs quinquennially. It focuses on different aspects of employment, such as working time, learning and training, earnings and financial security, as well as work-life balance and health.

Lastly, three data sources from the academic community are included that cover different topics related to individuals' attitudes, beliefs, values and behaviour: the ESS (ESS 2016a, 2016b, 2016 c), the EVS (EVS, 2016), and the ISSP (ISSP - Research Group, 2015, 2016). The ESS was set up in 2002 and runs every second year in around 30 European countries. The EVS was launched in 1981, and data from five rounds of the survey are now available. The ISSP is an annual survey set up in 1985, and like PIAAC, it extends beyond Europe.

These surveys partly differ in the definition of their target population, for instance with regard to age groups. To render the samples as comparable as possible, we include only respondents aged 25 to 64 in all surveys. The EWCS focuses on people who are employed and thus, we restrict the analytic sample of the EU-LFS to employed respondents when comparing it to the EWCS.

4.2.2 Measuring and Comparing Educational Attainment in Cross-National Surveys

Asking respondents about their educational attainment is standard in almost all surveys in the social sciences. This question often refers to individuals' highest formal qualification or their highest completed educational level for which a diploma or certificate from a school, a formal vocational training or an institution of higher education or university is awarded. Respondents usually answer this question by selecting a category from a list. Those lists are necessarily country-specific, as education systems differ in their institutions and the names of the qualifications, which cannot be accurately translated (Braun & Mohler, 2003; Schneider, Joye, & Wolf, 2016). Therefore, the ex-ante output harmonisation approach (Ehling, 2003) is commonly used in cross-national surveys. Before data collection, the survey teams agree on a standard classification or a coding scheme and ideally set up guidelines specifying what has to be considered when developing the country-specific answer categories. The mapping of these categories to the standard classification, which is used to compare education across countries, is also developed in advance (Ehling, 2003; Eurostat & OECD, 2014). To harmonise the education categories across countries, most surveys choose the ISCED classification. This was designed by UNESCO in the 1970s and revised in 1997 and 2011. It aims to enable comparisons of country-specific education programmes for producing international education statistics. The ISCED classification defines international levels and types of education (UNESCO-UIS, 2006), and education ministries and national statistical institutes map their educational programmes and qualifications to it. The most recent version of the classification was not yet implemented in most surveys for the years analysed, thus limiting this research to ISCED 97.

The main levels of ISCED 97 are:

- ISCED 0: Pre-primary education (or not completed primary education)
- ISCED 1: Primary education or first stage of basic education
- ISCED 2: Lower secondary or second stage of basic education
- ISCED 3: Upper secondary education
- ISCED 4: Post-secondary non-tertiary education
- ISCED 5: First stage of tertiary education
- ISCED 6: Second stage of tertiary education

The focus here is on comparing the main levels of ISCED 97, ignoring the additional complementary dimensions on programme orientation, destination, duration and position in the national qualification structure, as most of the surveys analysed do not use them. All surveys we analysed implement the main levels of the ISCED classification or a variant thereof, from which we can derive the main level of ISCED 1997 for comparing the distributions. We need to aggregate ISCED levels 0 and 1 and levels 5 and 6 because those categories are not separated in all surveys. When comparing the EU-LFS and the ISSP, we also need to aggregate ISCED levels 3 and 4 (see Tables A4.3 and A4.4 in the Supplemental Material).

Following Ortmanns and Schneider (2016a, 2016b) we calculate Duncan's Dissimilarity Index (Duncan & Duncan, 1955) to compare the education distributions between the EU-LFS, used as the benchmark survey, and the other surveys, which also use the ISCED classification. The index is defined as: D = ½ $\sum_{i=1}^{k} |x_i - y_i|$ where x_i denotes the number of observations in category i out of k ISCED categories for country A in year B in survey S, and y_i denotes the same for country A in year B in survey T. To interpret the resulting numbers as percentages, the index is rescaled to range from 0 to 100. This tells us how large the percentage is that needs to change categories to achieve equal education distributions between the EU-LFS and the survey in question.

Figure 4.1 shows the summary statistics of Duncan's Dissimilarity Index when comparing the education distributions between the EU-LFS and the other surveys within the same countries and years. The exact values can be found in Table A4.5 in the Supplemental Material; these are used later as the dependent variable. We observe the smallest value of 1% in Duncan's index when comparing data for the Czech Republic from the 2010 EU-LFS and EU-SILC; this indicates nearly perfectly consistent data. The largest deviation of 59% is found when comparing EU-LFS and EWCS data for Germany from 2010, which is even higher than the highest deviation identified by Ortmanns and Schneider (2016a). Overall, the mean inconsistency is almost 13%, meaning that on average 13% of respondents would need to change categories to achieve a distribution equal to that in the EU-LFS, which is the same result as found by Ortmanns and Schneider (2016a) based on a more limited set of international surveys. Duncan's Dissimilarity Index should, however, be close to zero because the education distributions should not vary across surveys when analysing the same country and year. This is clearly not the case. Looking at the individual surveys, we find the lowest discrepancy of roughly 6% when comparing the education distributions of the EU-LFS and the EU-SILC. When comparing the distributions of PIAAC and the AES to the EU-LFS, the discrepancy is 8%. We interpret these deviations as relatively small because they are clearly below the mean value of 13%. Duncan’s index indicates a discrepancy of 12% between the ESS and the EU-LFS, 14% between the EQLS and the EU-LFS and 15% between the EVS and the EU-LFS. These percentages are around the mean value (between 10 and 15%) and, thus, we regard those as intermediate discrepancies. The comparison between the EWCS and the EU-LFS indicates a discrepancy of 16% and between the ISSP and the EU-LFS the discrepancy is 17%. We find the largest discrepancy of 19% when comparing the education distributions of the EU-LFS and the Eurobarometer. We interpret these deviations, which are above 15%, as larger inconsistencies.

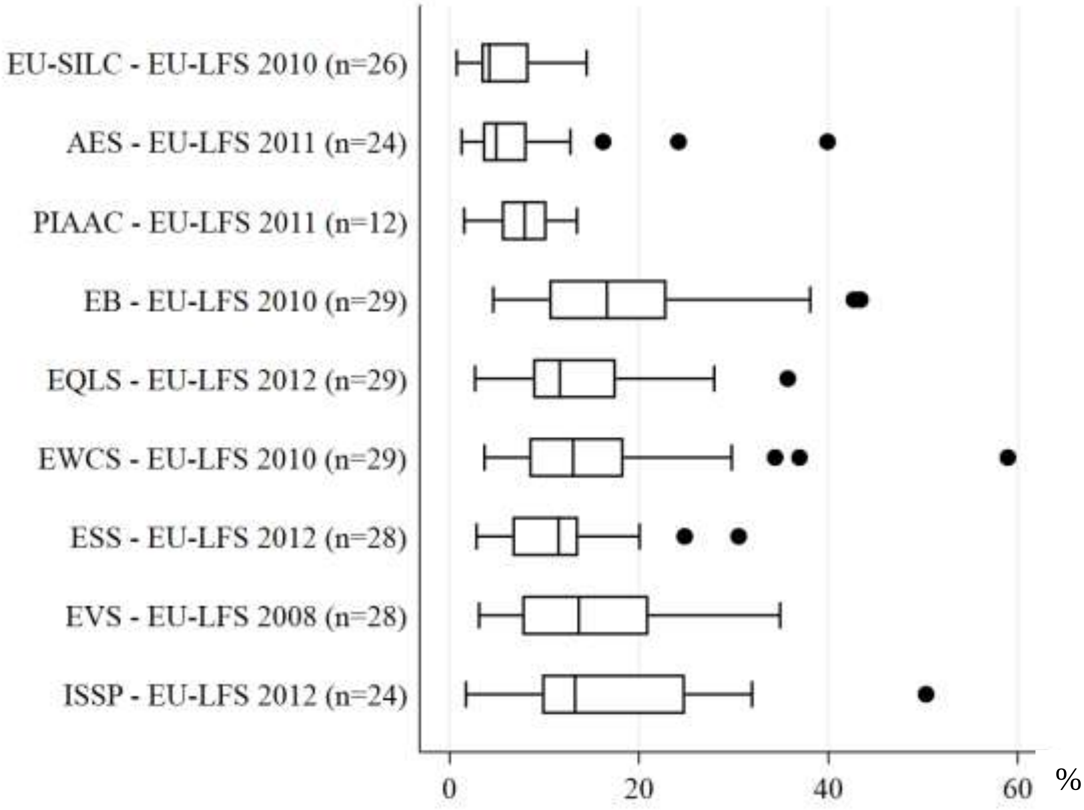


Figure 4.1 Boxplots of Duncan's Dissimilarity Index across countries for all survey comparisons

Notes: Here 'n' indicates the number of countries included in the analysis.
Data sources: EU-LFS 2008-2012, annual data, files from Eurostat, data file versions 2016, variable HATLEVEL, weighted using variable COEFF;
EU-SILC 2010, file from Eurostat, data file version CROSS-2010-6, variable PE040, weighted using variable PB040; AES 2011, file from Eurostat, data file version 1.0, variable HATLEVEL;
PIAAC 2011, file from OECD, data file version of 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB);
Eurobarometer 73.2 & 73.3 (2010), file from Eurostat, data file version 2.0.1of 2012, variable v362, data weighted to correct regional oversampling for Germany and the UK;
EQLS 2012, data file version 3 of 2014, variable Y11_ISCEDsimple, weighted using variable w1;
EWCS 2010, data file version of 2011, variable ef1_isce, weighted using variable w1;
ESS 2012, data file version 2.3 of 2016, variable edulvlb, weighted using variable dweight; for Greece and Croatia data from 2010 were used (data file versions: 3.3., variable edulvlb) and for Latvia and Romania data from 2008 were used (data file version 4.4, variable edulvla);
EVS 2008, data file version 4.0.0 of 2016, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK;
ISSP 2012, data file version 4.0.0 of 2016, variable DEGREE, data weighted to correct regional oversampling for Germany; for Italy data from 2011 were used (data file version 3.0.0, variable DEGREE);
Only respondents aged 25-64 for all surveys, apart from DE and AT in PIAAC including age 65.

4.3 Survey Characteristics

In order to explain differences between surveys, countries and years in terms of how well their education distribution matches that produced by the EU-LFS for the respective country and year, we refer to the Total Survey Error framework (Groves et al., 2009; Groves & Lyberg, 2010) that describes different sources of errors that can appear at different stages of a survey. We use this framework for structuring the survey characteristics according to the different error sources, following the dimensions of representation of the population and measurement. An overview of all survey characteristics analysed in this study can be found in Table 4.1.

Considering that all surveys we analysed in this study are cross-national, we have to be aware that the survey characteristics do not only vary across surveys, but also across participating countries (Kohler, 2008; Słomczyński et al., 2016). Different errors in the countries also reduce quality in terms of comparability across countries and/ or surveys, as described in the application of the TSE approach to cross-national surveys (Smith 2010, 2011).

Some methodological survey characteristics are design features of the survey that can be changed in principle, such as the mode of data collection or fieldwork duration. Other survey characteristics, such as response rate, cannot be changed directly by the survey organisers. In methodological studies, the relationship between different kind for survey characteristics haven been examined as well as the impact of single characteristics on the data quality. For instance, studies have assessed whether the mode of data collection or offering incentives have an impact on response rates (Church, 1993; Daikeler, Bosnjak, & Lozar-Manfreda, 2019 online first). Other studies evaluate the representation of the population of cross-national surveys by systematically comparing single survey characteristics across countries for a single survey (Kaminska & Lynn, 2017) or across several surveys (Kohler, 2007). Based on this research, best practice guidelines for survey organisers are developed (see e.g. Groves & Couper, 1998, chapter 11).

Table 4.1 Overview of the survey characteristics and their operationalisation

Dimension and errors of the TSE		Survey characteristic	Values	Values when comparing with EU-LFS
representation of the population	sampling error	Sampling design	simple, complex	0=equal , 1=unequal
		Final sampling unit	individual, household, dwelling/ address	0=equal , 1=unequal
		Sample size	n	absolute difference in the sample size divided by 1000
	nonresponse error	Response rates	in percent	0=equal response rate, 1=higher, < 30 percentage points, 2= lower, < 30 percentage points, 3= lower, ≥ 30 percentage points, 4= not available
		Survey participation	mandatory, voluntary	0=equal , 1=unequal
		Fieldwork duration	days	0=equal duration, 1=shorter, < 90 days, 2=longer, > 90 days, 3=longer, ≥ 90 days
	sampling and nonresponse error	Index to validate probability sampling	chance of interviewing a man/ woman of a married couple living together in a two-person household	0=equal , 1=unequal
		Index on gender and age	distribution of men and women for following age groups: 25-34, 35-44, 45-54, 55-64	deviations in percent, indicating differences in the gender and age distribution
measurement	measurement error	Response categories of the education question		0=same, 1=similar, 2=different
		Proxy-reporting	yes, no	0=equal , 1=unequal
		Information taken from register	yes, no	0=equal , 1=unequal
	processing/ harmonisation error	Applying official ISCED mapping	official ISCED mapping is applied, intended/ accidental deviation	0=equal , 1=unequal
		Degree of centralisation when applying ISCED	decentralised, partly/ entirely centralised	0=equal , 1=unequal
representation & measurement	sampling, nonresponse, measurement and processing error	Mode of data collection	face-to-face, telephone, self-administered, mixed-mode	0=equal , 1=unequal
		Fieldwork agency	institute of public authority, university/ scientific institute, commercial institute	0=equal , 1=unequal

4.3.1 Survey Characteristics Related to the Representation of the Population

In this section, we present several survey characteristics related to the representation of the population and how they could, theoretically, explain the inconsistencies in the education distributions between the EU-LFS, our benchmark, and the survey in question. When developing our hypotheses on the impact of the survey characteristics, we have to consider that those differ across countries also for the EU-LFS (see section 4.2.1). Thus, we will only formulate undirected hypotheses indicating that differences in the survey characteristics of the EU-LFS and the survey in question might explain discrepancies in the education distributions across surveys within the same country and year.

Looking at the dimension of representation in the TSE approach, four kinds of errors are distinguished: coverage, sampling, unit nonresponse, and adjustment error (Groves et al., 2009). Coverage error emerges at an early stage even before drawing a sample; it arises when there is a discrepancy between the sampling frame and the target population. Sampling error occurs when randomly taking a subset of sampling units from the sampling frame. When assessing sampling error, it is important to notice that most surveys analysed here use probability-based sampling methods, but that in the last stage, random-route approaches are applied in a few surveys. The survey characteristics on the sampling design and the final sampling unit reflect both coverage and sampling error and sample size only sampling error.

The *sampling design* influences the composition of the sample and thus also the education distribution. Almost every sampling design excludes some people from the target population, which might cause under- or over-coverage of certain groups (Groves & Couper, 1998; Lohr, 2009). In this article, we only distinguish between simple and complex sampling designs. In a simple design, the respondent is selected directly from an official register by means of a simple random sample. This is usually the case in the Scandinavian countries, which have central population registers. Ten countries of the EU-LFS have such a sampling design. In contrast, a complex sampling design might also use an official register, but multiple stages are used in the selection process. Other examples of a complex design are random digit dialing, and those where in the final stage a random route technique is applied. If the sampling design differs between the EU-LFS and the other survey, differences in the sample composition are likely, which might contribute to inconsistencies in the education distributions across surveys within the same countries and years (Hypothesis 1). Differences in the sample composition can also occur when both surveys apply complex sampling designs that differ from each other, for example through using different sampling frames. Unfortunately, generating a more detailed differentiation, for example by including additional information on the sampling frame, was not possible due to unstandardised or lacking information. For instance, it was also not possible to consider the information on how the surveys deal with institutionalised population because this often is not a central aspect in the documentation, although it is important to better assess errors in coverage and sampling (Schanze, 2017).

Next, we look at the *final sampling unit*. We differentiate between an individual, a household or a dwelling/ address. In most countries, the EU-LFS and the EU-SILC are household surveys and the dwelling/ address or the household are the final sampling unit. Usually, in those surveys all respondents in a household above a specified age (15 in the EU-LFS, 16 in the EU-SILC), and more than one respondent at the same address or dwelling, are interviewed. This increases the chance of being selected to answer the questionnaire. In contrast, most other surveys use the individual respondent as the final sampling unit, and the individual probability of being selected is lower in these surveys (Groves et al., 2009). The different selection probabilities can influence the sample compositions and thus also the education distribution. To not overestimate the effect of the different sampling units, especially for the household surveys, data are weighted using available design weights. Therefore, we hypothesise that differences in the final sampling units across surveys might not affect the inconsistencies in the education distributions across surveys (Hypothesis 2).

The *sample size* of a survey matters because previous research shows that surveys with a larger sample size are more accurate, as the sampling error decreases (Biemer & Lyberg, 2003). Surveys with smaller samples are more likely to have a sampling error that can lead to a slightly different sample composition and thus to a slightly different education distribution. All analysed surveys have rather large samples; however, the EU-LFS has by far the largest sample size for each country. Thus, we will definitely observe deviations in the sample size across the surveys. However, we estimate that these differences in the sample size might not contribute to the discrepancies in the education distribution (Hypothesis 3).

The nonresponse error, focusing on unit nonresponse, results in lacking representativeness of the sample. This error occurs if respondents systematically differ from non-respondents, that is sample members who refuse to participate in the survey or who cannot be interviewed. Here, we look at the following survey characteristics: mandatory survey participation, fieldwork duration and response rate. The survey characteristic on *mandatory survey participation* indicates that respondents are forced to participate in the survey. Usually, those surveys achieve higher response rates, and the nonresponse error is low because respondents who would refuse in voluntary surveys are often included in mandatory ones. Thus, we hypothesise that differences in mandatory survey participation across the EU-LFS and the other surveys might explain inconsistencies in the education distribution (Hypothesis 4). In the analysed surveys, participation is mandatory for only a small number of countries and surveys, namely 13 countries in the EU-LFS and nine in the AES.

Regarding *fieldwork duration* previous research indicate that longer field periods increase the chance of contacting and interviewing hard-to-reach respondents, whereas shorter fieldwork durations often leave less time for follow-ups. Thus, for surveys having a shorter fieldwork duration, errors of nonresponse become more likely (Biemer & Lyberg, 2003). In the EU-LFS, fieldwork duration is usually three months and we distinguish whether the fieldwork compared to the EU-LFS is longer or shorter. We expect that different fieldwork durations – either considerably shorter or considerably longer than the benchmark – might increase inconsistencies in the education distribution across surveys within the same countries and years (Hypothesis 5).

The *response rate* is an important quality indicator and survey organisers invest a great deal of money in increasing it, for instance, by offering incentives to the respondents (Singer & Ye, 2013; Groves et al., 2006). The response rate of the EU-LFS is relatively high, due to mandatory survey participation in some countries and because proxy-reporting is generally permitted. In contrast, for most other surveys the response rates are much lower and this might indicate that their realised samples can differ from the sample of the EU-LFS, that is, there is a higher risk of nonresponse error. Thus, we hypothesise that large differences in the response rates between the EU-LFS and the other surveys within countries and years could contribute to explaining inconsistencies in the education distributions (Hypothesis 6). However, we know that a high response rate alone is not enough to avoid nonresponse error (Bethlehem, Cobben, & Schouten, 2011; Groves & Peytcheva, 2008). Nevertheless, we decided to include this survey characteristic because we have no better indicator of the nonresponse bias.

The last error related to representation of the population is adjustment error. It emerges after data collection when calculating weights. This error is not taken into account in this study, because data are only weighted using design weights that correct for different inclusion probabilities due to different sampling designs across countries. Applying post-stratification weights that also correct for nonresponse errors is not feasible because those often correct for education, frequently by using the distribution of the EU-LFS as benchmark (e.g., ESS, 2014b). This would lead to an (almost) equal distribution of the two surveys that are being compared.

Some specifications of the described survey characteristics relating to representation of the population are rather broad, for instance regarding the sampling design and sampling unit. This is caused by vague and sometimes also questionable documentation, particularly the design of the sampling process (for more information on the different standards in documentation, see Kohler, 2008; Słomczyński et al., 2016). Therefore, it is advisable to also look directly into the data and check the realised representation. Firstly, we generate Sodeur's Index to validate probability sampling of the survey (Sodeur, 1997, 2007). This index is based on the assumption that in a random sample, the chance of interviewing a man or a woman in a married couple living together in a two-person household is equal, namely 50:50. We adapt this and define the observed distribution of the EU-LFS as a benchmark. For calculation, we firstly restrict all samples to the 25 to 64 age group and married couples living in two-person households. Unfortunately but not unexpectedly, the required variables on marital status and household composition differ greatly across surveys, so adaptations are needed (for details see Annexe 1 in the Supplemental Material). We calculate the gender distribution of this restricted sample and compare it to the distribution identified in the respective sample of EU-LFS, applying the following formula: $B_{UNR}=\frac{\hat{p}-p}{\sqrt{var\ (\hat{p})}}$ where p is the proportion of women in the EU-LFS and $\hat{p}$ is the proportion of women in the survey in question for the same country and year. Finally, the 95% confidence interval is calculated so we can decide whether the gender distribution between the EU-LFS and the other survey is equal or not within the same country and year. Secondly, we calculate an index to compare the gender and age distributions for four age groups (25-

34, 35-44, 45-54, and 55-64) across surveys. Here, we again calculate Duncan's Dissimilarity Index (Duncan & Duncan, 1955) and we use the distribution of the EU-LFS as benchmark.

4.3.2 Survey Characteristics Related to Measurement

On the measurement dimension of the TSE framework, there are three kinds of error that can occur: invalidity, measurement error, and processing error (Groves et al., 2009). Invalidity occurs when there is a disparity between the theoretical construct (what is intended be measured) and what is actually measured by the indicator. In this study, we do not expect to find invalidity because every survey asks respondents for their highest educational attainment in an equivalent way, asking respondents for their highest certificate/ degree or their achieved educational level.

Measurement error occurs when a mismatch exists between the ideal measurement and the actual response obtained from the respondent. A potential source of measurement error across surveys is differences in the *response categories* in the education question. Previous research shows many examples pointing at differences in the measurement instrument as a source of inconsistent education data (Kieffer, 2010; Ortmanns & Schneider, 2016a, 2016b; Schneider, 2009). For instance, when surveys use ambiguous terms or generic descriptions of educational qualifications, instead of the official name of the qualifications, the chance that the response categories differ across surveys is quite high. Thus, this survey characteristic seems to be of some importance when explaining inconsistencies in the education distributions. In the education question, the response categories are the key element influencing respondents' answers. All analysed surveys use country-specific response categories for the education question. To assess the similarity of the response categories of the EU-LFS and the other surveys, we qualitatively compared the education categories for every survey, country and year and generated an index. It distinguishes whether the categories are the same as, similar to, or different from the categories used in the EU-LFS. Detailed information on this index is provided in Annexe 2 in the Supplemental Material. In general, we know that different stimuli can affect respondents' answers (Groves et al., 2009) and this also seems to occur with the education question, even though it is a factual question. Thus, different response categories are a probable explanation for inconsistencies in the education distributions (Hypothesis 7).

Relating to the measurement, we also measure whether *proxy-reporting* is allowed or prohibited. If the survey allows proxy-reporting, a respondent's partner or (adult) child might answer the questions instead of the selected respondent, or the 'head of the household' responds for every household member. Proxy-reporting can only be used in household surveys; thus, it applies to the EU-LFS, EU-SILC and the AES. Proxy-reporting is cognitively demanding, and measurement errors are likely due to lack of knowledge leading to incorrect answers (Blair, Menon, & Bickart, 2011; Kreuter et al., 2010; Moore, 1988). Thus, we expect that differences in the allowability of proxy-reporting can contribute to inconsistencies in the education distribution across surveys (Hypothesis 8).

The last survey characteristic related to measurement error distinguishes *whether respondents' educational attainment is retrieved from a register* or not. Some countries, mostly Scandinavian ones, have population registers from which socio-demographic information, including education, can be directly retrieved. Register information is regarded as high quality and trustworthy (Biemer & Lyberg, 2003). Therefore, differences in this survey characteristic on retrieving information from a register may explain inconsistencies in the education distribution (Hypothesis 9). However, we also have to be aware that register information is not free of errors either, due to delayed updates, especially for younger people who are currently in education (Kleven & Ringdal, 2017). Only four countries of the EU-LFS use register information.

Next, we look at errors in the data processing, including harmonisation, these emerge while transforming responses into the final dataset to be used for analysis. Processing errors seem to be of great importance: previous studies have repeatedly reported errors when classifying the country-specific educational qualifications into ISCED (Kieffer, 2010; Ortmanns & Schneider, 2016a; Schneider, 2009; Hoffmeyer-Zlotnik, 2008). Those errors directly influence the education distributions. We distinguish two survey characteristics here. The first one indicates *whether the official ISCED mapping is applied.* This is important because only if the educational qualifications are classified to ISCED in a consistent way, for example by following the official mappings, the education distributions are comparable across surveys (Schneider, 2009). This characteristic distinguishes whether the assignment of ISCED codes to national education categories follows the official mapping or whether we find deviations from the official mapping. The EU-LFS and EU-SILC are conducted by the national statistical offices, which are also often responsible for developing countries' ISCED mapping, meaning they determine the ISCED code for each country-specific educational qualification. Therefore, we expect that the EU-LFS and the EU-SILC follow the official mapping and that processing errors are rare in these surveys. In the other surveys, classification errors may occur more often because of lack of expertise in implementing the ISCED classification, which might lead to so-called 'accidental' errors. The other reason for this processing error is lack of trust in the official mappings and this might lead to intended deviations from the official ISCED mapping. This deviation is more common in academic surveys such as ESS, EVS and ISSP, which are not obliged to follow the official ISCED mappings. Therefore, we estimate that differences in the application of the official ISCED mappings across surveys can contribute to inconsistencies in the education distribution (Hypothesis 10).

The second survey characteristic indicating processing or harmonisation error describes the *degree of centralisation when applying the ISCED classification* for the survey. It distinguishes between decentralised, partly centralised and centralised processing. In the decentralised approach, the country teams, who are familiar with their education system, are responsible for assigning the ISCED codes to national education categories. The EU-LFS and most other surveys implemented this approach. In contrast, in the centralised approach, one institute is responsible for assigning the ISCED codes for all countries of the survey. The Eurobarometer follows this method. Applying ISCED codes for several countries requires much expertise in ISCED and in the different educational systems. If one of these components is lacking, the chance of processing or harmonisation errors increases. Another approach combines both methods: classifying the national education category in ISCED is carried out by the country teams, but it is also checked centrally. This is beneficial because it involves country experts and an expert in the application of ISCED, and aims to optimise cross-national comparability. The ESS implemented this approach. Hence, differences in the degrees of centralisation across the surveys can increase inconsistencies in the education distributions across surveys within the same countries and years (Hypothesis 11).

4.3.3 Survey Characteristics Related to Both Measurement and Representation

Two survey characteristics are related to both dimensions of the TSE framework: mode of data collection and fieldwork organisation. Regarding the *mode of data collection*, we distinguish between face-to-face interviews, telephone interviews, self-administered modes (including web and postal surveys), and mixed-mode designs. The mode is a relevant factor for representation because different modes tend to systematically over- or under-represent certain groups, for example web surveys tend to over-represent more highly educated respondents (Couper, 2000; Dever, Rafferty, & Valliant, 2008). Regarding the measurement dimension, the mode indicates the presence of an interviewer and the communication channel used. In face-to-face or telephone interviews, the presence of an interviewer makes socially desirable answering and interviewer effects more likely (de Leeuw & van der Zouwen, 2001; Lyberg & Kasprzyk, 2011), however, interviewers may also help the respondent identify a suitable answer. In face-to-face or self-administered modes, respondents usually see a list of education categories, while in telephone interviews, these categories are read out or an open response is coded by the interviewer, which is more error-prone and primacy or recency effects can occur in the former case (Noelle-Neumann & Petersen, 2000). Therefore, we expect that different modes of data collection across the surveys within the same countries and years can increase inconsistencies in the education distributions across surveys (Hypothesis 12).

Fieldwork agencies are responsible for conducting the survey and are thereby involved in several aspects of sample representation and measurement. Therefore, the fieldwork agency can be seen as indicator for the standard of the survey and as proxy for different aspects, including those could not be specified as survey characteristic due to a lack of information. This, for instance, applies to the availability of information on interviewer training. Concerning the EU-LFS, we would expect the overall standard to be quite high, largely because the fieldwork is done by a public authority, mostly the national statistical offices. This also applies to the second official survey, the EU-SILC. For the other surveys, commonly other fieldwork agencies are responsible, e.g. universities, other scientific or commercial institutes. We hypothesise that different kinds of fieldwork agencies can contribute to inconsistencies in the education distributions across surveys within the same countries and years (Hypothesis 13).

4.4 Data, Variables and Methods

In this study, we analyse the impact of surveys' methodological characteristics on discrepancies between the distributions of the harmonised education variable when comparing the EU-LFS with nine other surveys within the same countries and years. A description of the EU-LFS and the other surveys was already given in section 4.2.1. This study focuses on these surveys from the period 2008 to 2012. If a survey was run several times during this time, such as the EU-SILC, the Eurobarometer, the ESS and the ISSP, it is only included once in order not to overestimate its effect. For most surveys the education distribution is stable over the years, as long as the country-specific measurement instruments and the harmonised education variable do not change (Ortmanns & Schneider, 2016a, 2016b). When deciding which year to include, we consider the following factors: (a) number of countries covered, (b) completeness of documentation of survey characteristics, (c) whether its harmonised education variable has systematically changed (as in the ESS 2010 and the ISSP 2011), in which case the most recent year is included, (d) when a single country is not present in the selected year, information from an earlier round is used for this country. Due to a consequential processing error in the ISCED variable for Iceland in the EU-LFS 2011 and 2012 (for details see Ortmanns & Schneider, 2016a), data before 2011 are included as far as possible. Thus, we include the EU-SILC and the Eurobarometer of 2010, and the ESS and ISSP of 2012.

As described in section 4.2.2, the dependent variable is Duncan's Dissimilarity Index that compares the education distributions for each country and year of the EU-LFS with the respective country and year of each other survey. The independent variables reflect the survey characteristics (see section 4.3) that differ across surveys for the same country-year comparison. Annexe 3 in the Supplemental Material provides basic descriptions of each survey characteristic. As mentioned, we focus on whether the survey characteristics differ between the EU-LFS and the respective other survey. Thus, most variables are coded as binary and distinguish whether the survey characteristics are 'equal' (0) or 'unequal' (1). The variables on response categories, fieldwork duration, response rates, sample size and the index of gender and age distribution are operationalised in a slightly more nuanced way. As described in section 4.3.2, we generate an index to assess the comparability of the response categories and distinguish between equal, similar and different. When comparing the fieldwork duration of the

EU-LFS with the other surveys, we distinguish between the following categories: ‘equal fieldwork duration to the EU-LFS’, including up to five percentage points more or fewer days than the EU-LFS, ‘longer duration: up to 90 days’ and ‘longer duration: 90 days or more’, ‘shorter duration: up to 90 days’. These four categories cover all comparisons. Regarding response rates, we use the ones reported in the survey documentation, even when we do not know exactly how these have been calculated, which may hamper their comparability. For the comparison of the response rates, we generate the following categories: ‘equal response rate to the EU-LFS’ if the response rate is up to 5 percentage points lower or higher than in the EU-LFS, ‘lower response rate: up to 30 percentage points, ‘lower response rate: 30 percentage points or more’ and ‘higher response rate: up to 30 percentage points’. A category indicating a higher response rate of more than 30 percentage points was not required. Unfortunately, the Eurobarometer does not provide information on response rates and for some countries of the other surveys the response rates are not documented. In order to be able to include those anyway, we generate an additional category ‘information not available’. The categories of the variables on fieldwork duration and response rate are based on their distributions, and in order to avoid small or empty categories, they are rather broad. We include these categories as dummy variables in the analysis, and the categories indicating equal response rate or fieldwork duration are used as reference categories. When comparing the sample sizes of the EU-LFS with the other surveys, we calculate the absolute differences in the sample size and then divide by 1,000 because of the very high number of respondents in the EU-LFS. We then include this as a continuous variable. Duncan’s index on the gender and age distribution delivers percentages and these are directly included in the regression models.

For many of the survey characteristics analysed, it would be desirable to use a higher level of detail. Unfortunately, this is not possible due to large variation in the accessibility of information, and especially the quality and the richness of the documentation. Still we had to exclude single countries in single surveys from the analysis when the information on a survey characteristic was not available. Thereby the dataset is reduced from 248 to 229 survey comparisons and their respective comparisons of survey characteristics. The highest number of countries covered for one comparison is 29 when comparing EU-LFS with the Eurobarometer, or the EQLS or the EWCS, whereas the comparison between EU-LFS and PIAAC contains only 12 countries. An overview of the countries participating in the surveys and those included in the analysis can be found in Table A4.6 in the Supplemental Material.

Survey characteristics may correlate with each other and also with the survey programmes. Multicollinearity could make it hard to properly disentangle the effects of individual variables. Therefore, we checked the correlations between the different survey characteristics beforehand and Cramer’s V was below 0.65. More details can be found in the Tables showing cross tabulations and correlations for selected survey characteristics in Annexe 4 in the Supplemental Material. Additionally, we calculate the Variance Inflation Factor (VIF) after each regression model.

In the analysis, we estimate four multiple OLS regression models to explore the impact of different survey characteristics on inconsistencies in the education distributions. The first model shows the impact of the survey programmes alone and thereby illustrates the large variation in the education distributions across surveys. The survey comparisons are included as dummy variables, and the comparison of EU-SILC and EU-LFS is used as reference. To explain these inconsistencies through differences in the survey characteristics, the second model adds the survey characteristics related to representation of the population. The third model includes survey characteristics related to measurement and survey programmes. To further reduce multicollinearity we calculate the final model excluding the dummy variables of the survey programmes. This model focuses on the survey characteristics that show statistically significant effects in models 2 and 3.

4.5 Results

4.5.1 Impact of the Survey Programmes

As seen in the boxplot diagram (see Figure 4.1; section 4.2.2) the inconsistencies in the education distributions differ strongly across surveys within the same countries and years. As expected, this pattern recurs when running a linear regression to predict Duncan’s Dissimilarity Index by the survey programmes alone.

Model 1 in Table 4.2 shows low values for the regression coefficients for PIAAC (b=2.30) and the AES (b=2.38) and these survey comparisons are not statistically significant. The regression coefficients of the comparisons to the other survey programmes are higher (b>5.00) indicating larger inconsistencies in the education distribution than in the reference comparisons of EU-LFS and EU-SILC. The comparison of the EU-LFS and the ESS is significant at the five percent level ($p<.05$), and the comparisons of the EU-LFS to the Eurobarometer, the EQLS, the EWCS, the EVS and the ISSP are highly significant ($p<.001$).

The adjusted R^2 of this model is 17%, meaning 17% of the variance can be explained by just the surveys themselves. This is unexpected because we can imagine the survey programmes as 'containers' for different survey characteristics. To identify which survey characteristics contribute to the inconsistencies in the education distributions, we estimate further regression models.

Table 4.2 Results from regression analyses estimating the impact of survey characteristics on the inconsistencies in the education distribution across surveys

predictor	Model 1				Model 2				Model 3				Model 4			
	b		**SE**	**p**	**b**		**SE**	**p**	**b**		**SE**	**p**	**b**		**SE**	**p**
Survey: (ref: SILC)																
AES	2.38		2.47	.337	4.25		3.10	.172	-1.05		3.06	.731				
PIAAC	2.30		3.04	.451	4.42		3.72	.237	-3.08		4.21	.464				
EB	12.97	***	2.36	<.001	14.94	**	4.63	.001	3.71		7.19	.606				
EQLS	8.33	***	2.36	<.001	11.70	**	4.01	.004	-0.70		4-34	.872				
EWCS	10.21	***	2.36	<.001	14.22	**	4.06	.001	1.40		4.30	.745				
ESS	5.94	*	2.38	.013	8.55	*	3.62	.019	-2.74		6.78	.686				
EVS	9.20	***	2.38	<.001	12.34	**	3.69	.001	1.59		4.22	.708				
ISSP	11.57	***	2.47	<.001	13.95	**	3.88	<.001	0.50		4.36	.909				
Different sampling design (ref: equal)					3.67	*	1.52	.016	3.37	**	1.25	.007	3.43	**	1.25	.007
Different sampling unit (ref: equal)					0.51		1.75	.772								
Differences in Sodeur's index (ref: equal)					-1.11		2.57	.665								
Duncan's index age/ gender					-0.10		0.19	.594								
Sample size/ 1000					0.00		0.01	.647								
Fieldwork duration: (ref: equal)																
Shorter, < 90 days					1.09		2.28	.631								
Longer, < 90 days					0.16		2.22	.941								
Longer, ≥ 90 days					-0.73		2.53	.772								
Response rate: (ref: equal)																
Higher, < 30 percentage points					-0.90		3.11	.772								
Lower, < 30 percentage points					-3.55		2.66	.183								
Lower, ≥ 30 percentage points					-5.14		2.93	.081								
Not available					-4.01		3.91	.307								

predictor	Model 1 b		Model 1 SE	Model 1 p	Model 2 b	Model 2 SE	Model 2 p	Model 3 b		Model 3 SE	Model 3 p	Model 4 b		Model 4 SE	Model 4 p
Differences in mandatory participation (ref: equal)					0.82	1.32	.537								
Different mode (ref: equal)					2.15	1.43	.134	1.44		1.19	.225	1.46		1.19	.217
Different agency (ref: equal)					-1.42	2.42	.557	-0.31		2.17	.886	1.34		1.71	.432
Differences in proxy-reporting (ref: equal)								3.97		3.21	.217	2.89		1.98	.148
Difference in using register (ref: equal)								-0.65		1.93	.735	-0.47		1.92	.806
Response categories: (ref: equal)															
Similar								0.87		2.39	.714	0.67		2.37	.776
Different								5.13	*	2.33	.029	5.28	*	2.24	.020
Differences in centralised coding (ref: equal)								0.53		5.67	.926	0.86		1.25	.494
Differences in ISCED coding (ref: equal)								9.20	***	1.31	<.001	9.39	***	1.27	<.001
Constant	5.58	**	1.71	.001	5.63	3.62	.122	2.15		2.27	.346	1.84		2.18	.399
Adjusted R^2 (%)	16.61				16.67			35.59				34.00			
Akaike information criterion (AIC)	1650.76				1664.42			1600.04				1598.15			
Mean of variance inflation factor (VIF)	1.75				3.43			7.17				1.79			
Number of observations	229				229			229				229			

Data sources: EU-LFS 2008-2012, files from Eurostat, data file versions 2016, variable HATLEVEL, weighted using variable COEFF;
EU-SILC 2010, file from Eurostat, data file versions CROSS-2010-6, variable PE040, weighted using variable PB040;
AES 2011, files from Eurostat, data file version 1.0, variable HATLEVEL; PIAAC 2011, file from OECD, data file version of 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB);
Eurobarometer 73.2 & 73.3 (2010), file from Eurostat, data file versions 2.0.1 of 2012, variable v362, data weighted to correct regional oversampling for Germany and the UK; EQLS 2012, data file version 3 of 2014, variable Y11_ISCEDsimple, weighted using variable w1;EWCS 2010, data file version of 2011, variable ef1_isce, weighted using variable w1; ESS 2008, data file version 4.4, variable edulvla; ESS 2010-2012, data file versions: 3.3. (2010), 2.3 (2012) variable edulvlb, weighted using variable dweight; EVS 2008, data file version 4.0.0 of 2016, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK; ISSP 2011-2012, data file versions 3.0.0 (2011), 4.0.0 (2012), variable DEGREE, data weighted to correct regional oversampling for Germany;
Only respondents aged 25-64 for all surveys, apart from DE and AT in PIAAC including age 65.

4.5.2 Impact of Survey Characteristics Related to the Representation of the Population

In addition to the first model, this model (Model 2 in Table 4.2) includes the survey characteristics related to the representation of the population, namely: sampling design, final sampling unit, sample size, mandatory survey participation, fieldwork duration, response rate, Sodeur's Index and Duncan's Dissimilarity Index for the age and gender distributions. Mode of data collection and fieldwork agency are also included.

This model shows that adding variables related to representation does not improve model fit: The adjusted R^2 of this model is also 17%. To estimate the quality of this model relative to the first model, we calculate the Akaike Information Criterion (AIC). For model 1, the AIC is 1650.8 and for this model the AIC slightly increases to 1664.4. The model that shows the lowest value of the AIC, here model 1, performs best. Regarding multicollinearity, the highest value of the VIF in this model is 7.1, which we observe for the dummy variable of the Eurobarometer. This indicates that the Eurobarometer correlates with the analysed survey characteristics. The mean value of the VIF of this model is 3.4, which is higher than in model 1 (mean VIF of 1.8) but still unproblematic.

The only survey characteristic that has a statistically significant impact ($p<0.05$) in this model is different sampling designs across the surveys. The regression coefficient of 3.7 indicates that different sampling designs increase the inconsistencies in the education distributions by roughly four percentage points, compared with equal designs. Thus, we do not reject hypothesis H1. From the results of this model, we find no evidence that the survey characteristics contribute to a higher inconsistency of the education distribution and therefore we do not reject H2 and H3 and we reject hypotheses H4 to H6, H12 and H13. In contrast to most survey characteristics, the survey effects remain significant and their regression coefficients even increase. Overall, this model shows that even when controlling for a substantial number of survey characteristics related to the representation of the population, the survey programmes themselves have by far the largest impact on the observed inconsistencies in the education distributions across surveys.

4.5.3 Impact of Survey Characteristics Related to Measurement

The third regression model shown in Table 4.2 focuses on the survey characteristics related to measurement. The following survey characteristics are included in this model: different response categories of the education question, proxy reporting, use of register information, applying of the official ISCED mappings and the degree of centralisation when applying ISCED. Also included are mode of data collection and fieldwork agency, which refer to both dimensions of the TSE, as well as the sampling design, which was significant in the second model. This model also controls for the survey programmes again.

This model has an adjusted R^2 of 36%, meaning more than one-third of the variance can now be explained. This is an increase of 19 percentage points compared to the previous models. The increase of the adjusted R^2 indicates a strong impact of survey characteristics related to measurement, over and above the effects of the surveys themselves. Compared to model 1 and 2 the AIC decreases to 1600.0, which indicates a higher quality of this model. Concerning multicollinearity, the mean value of the VIF is 7.2, which is higher than in models 1 and 2. In detail, we find high VIF values of around 20 for the dummy variables of the survey programmes for the Eurobarometer and the ESS, as well as the survey characteristic on the degree of centralisation when applying ISCED. This is not surprising because we know that this survey characteristic is strongly associated with the survey programme.

In this model, three survey characteristics have a statistically significant impact: different sampling designs, different response categories in the education item(s) and application of the official ISCED mapping. We find the strongest impact from the survey characteristic that indicates differences in whether the official ISCED mappings were applied between the EU-LFS and the surveys in question. This variable shows a high regression coefficient of 9.2, meaning inconsistency in the mapping of the national educational qualification into ISCED increases inconsistencies in the education distributions by roughly ten percentage points compared to consistent mapping. This effect is highly significant ($p<0.001$). Thus, whether the official ISCED mappings are applied is a crucial factor that explains deviations in the education distributions across surveys within countries and years. Therefore, we do not reject hypothesis H10.

The survey characteristic indicating different response categories in the education items between the EU-LFS and the other surveys is also significant ($p<0.05$). The regression coefficient of 5.1 indicates that using different response categories raises inconsistencies in the education distribution across surveys by roughly five percentage points compared to equal response categories. Thus, we also do not reject hypothesis H7.

The survey characteristic assessing different sampling designs between the EU-LFS and other surveys, which was the only significant factor in model 2, is again significant. The regression coefficient increases to 3.4 and the p-value is smaller in this model ($p<0.01$), thus we again do not reject hypothesis H1 in this model. Nevertheless, the effect of sampling design is smaller compared to the coefficients related to measurement.

All other survey characteristics are not statistically significant. The survey comparisons themselves are also not significant any more. Thus, in this model we identified the survey characteristics causing inconsistencies in the education distributions across surveys, and we successfully opened ‘the black box of the surveys’.

In the final model (Model 4 in Table 4.2) the adjusted R^2 slightly decreases to 34%. The AIC declines to 1598.2, which is lowest value across all models, indicating that this is the best model estimated. Though excluding the survey programmes, we also reduce multicollinearity and the mean value of the VIF decreases to 1.8. The statistical significance of the variables assessing different sampling designs ($p<0.01$), different response categories ($p<0.05$) and differences in the application of the official ISCED mapping ($p<0.001$) between the EU-LFS and the other surveys remain. This highlights the importance of these three survey characteristics independently of the survey programmes. Thus, we do not reject hypotheses H1, H2, H3, H7 and H10, but according to this analysis, we can reject all other hypotheses. This result emphasises a predominant effect of measurement, especially the consistency of applying the official ISCED mappings and consistent response categories in the education question. Those are the key elements when it comes to explaining the inconsistencies in the education distributions across surveys within countries and years.

4.6 Conclusion and Discussion

This article asked which survey characteristics could explain the inconsistencies in the education distributions when comparing nine cross-national surveys to the EU-LFS. To answer that question, the impact of 15 survey characteristics and the survey programmes themselves were estimated. The dataset used for this analysis contains detailed macro-information concerning the survey characteristics for the countries and years of the ten surveys. The main finding of this study is that differences in applying the official ISCED mappings (H10), differences in the response categories of the education question across surveys (H7), as well as – but to a lesser degree – differences in the sampling designs of the surveys (H1), are systematically related to inconsistencies in the education distributions across surveys within the same countries and years. These results are in line with our expectation and also with previous research (Kieffer, 2010; Schneider, 2009; Ortmanns & Schneider, 2016a, 2016b) that focused on the measurement of the education variable to explain inconsistent education distributions. Hence, the focus of previous studies was well justified. The comprehensive analysis of survey characteristics in this study additionally shows that apart from the sampling design, the survey characteristics related to the representation of the population do not cause inconsistencies in the education distribution across surveys.

To achieve higher consistency in the education distributions across surveys, survey organisers should, firstly, reduce the processing error by improving the assignment of the response categories of the education item to the ISCED classification. To make recommendations on how to reduce the processing error, we further need to distinguish whether the deviation from the official ISCED mapping occurs accidentally or whether it is intended. 'Accidental' errors, which are often caused by limited knowledge when assigning the national educational qualification to the ISCED classification, can be avoided through implementing additional quality checks and the application of the official ISCED mappings in principal (Ortmanns & Schneider, 2016a).

In contrast, the intended deviations applied by some academic surveys aim to enhance comparability of cross-national education data across countries (Ortmanns & Schneider, 2016a). This is justified because during the development and the implementation of the ISCED mappings it is vulnerable to political influence of education ministries and national statistical offices. The latter often develop the national

ISCED mappings and they do not equally strictly apply the ISCED criteria. At the same time, some criteria formulated in the ISCED classification are rather vague and thus leave some room for interpretation. This explains why countries with similar qualification nevertheless classify them to different ISCED codes. The intended deviations made by academic surveys attempt to correct for this. However, these deviations also introduce incomparability across survey, notably with official surveys applying the official ISCED mappings, such as the EU-LFS and the EU-SILC. Intended deviations could be avoided when the quality control of the national ISCED mappings, for example through UNESCO, would become stricter. As this is currently not ensured, the international survey community has good reasons to find solutions to produce comparable education data for their own purpose. Academic surveys, for instance, could agree on applying an ‘alternative’ ISCED scheme that adjusts the official mappings to optimise comparability over time and space. This alternative version should be well-documented and contain recodes to the official mappings in order to still compare them with official education data.

The second important recommendation to achieve higher consistency in the education distributions across surveys is to improve the education item itself. We should aim for standardised country-specific education categories, which use a terminology that is equally understandable for everyone and avoid generic terms and descriptions. These categories can then be implemented in all surveys, national as well as international, that measure education as a background variable. Of course, no instrument will be without measurement error; however, if every survey uses the same instrument, the error will be consistent and this enhances data comparability. The development of these country-specific education categories and their assignment to ISCED should be done by a national expert group, which should consist of experts of the country-specific educational system, experts of ISCED and also representatives of the national statistical office, the education ministry as well as a survey expert. Ideally, also an expert in cross-national surveys should be included in the discussion to consider comparability in international surveys. Additionally, for countries having a similar educational system, for instance Germany, Austria and Switzerland or the UK and Ireland, it is also worthwhile to exchange their suggestions and, even better, to discuss shared issues. Then we can also better consider comparability across *countries*, which we did not look at in this article.

This study also faces some limitations. An obvious one is the small number of cases (n=229), which might be problematic for testing such a large number of survey characteristics. However, focusing on whether the survey characteristics are equal or unequal across surveys prevents us from having small or even empty cells. The disadvantage of these variables is that they are quite generic, and it is not possible to, for instance, to identify which kind of fieldwork agency (public authority including statistical office, university or other scientific institute, commercial institute) causes more or less inconsistent education distributions. We can only tell whether differences in the fieldwork agencies between the survey in question and the EU-LFS affect deviations in the education distribution. This structure of the variables and the low case number furthermore do not allow calculation of more complex models or application of multilevel modeling.

Another limitation of this study is that it compares the education distribution using the 1997 version of ISCED, whereas surveys are increasingly implementing the more recent version – ISCED 2011. However, we are convinced that the current results would not be very different and we would still find inconsistencies when comparing the education distributions across surveys within countries and years. One change in ISCED 11 is a better differentiation of levels within tertiary education, so when surveys implement this new version, they will be paying particular attention to the codes of tertiary education. However, we observe the greatest inconsistencies for ISCED level 3 (upper secondary education), and also find deviations in the adjacent categories ISCED level 2 (lower secondary) and ISCED level 4 (post-secondary, non-tertiary). At these levels we find most of the ambiguous terms and generic descriptions used in the response categories of the surveys, especially with the vocational qualifications. These can also cause errors when assigning ISCED codes. The inconsistencies on these levels will not disappear when implementing ISCED 11, unless surveys start primarily to correct for accidental errors when assigning ISCED codes and update the country-specific response categories alongside the implementation of the new ISCED version. The ESS in 2010 undertook such a detailed check and updated its variables, and a similar review took place for the EVS 2017. The ISSP is currently considering how best to implement ISCED 11. The effort invested in the education variables in these surveys is likely to reduce inconsistencies in the education distribution in the future.

An output of this study is the data file of survey characteristics that is available at the SowiDataNet|datorium (Ortmanns, 2020). Until recently, survey characteristics have rarely been considered in substantive data analyses, and only few studies exist that include them (e.g., Heath, Martin, & Spreckelsen, 2009; van Tuyckom & Bracke, 2014). The main reason that survey characteristics are often neglected is probably that collecting and harmonising this information requires considerable effort. Often the documentation of survey characteristics is neglected, meaning we have to look at several documents of varying quality, to be found on different webpages of the surveys or data archives. Sometimes we still cannot find complete information, and it is little standardised. More systematic and easily accessible documentation would be very helpful. This would enhance transparency and increase the possibility of developing standards on how to report survey characteristics. Some initiatives have begun by collecting, documenting and publishing information on methodological survey characteristics relevant for their specific projects. Such an initiative exists for official statistics within the online platform MISSY, which provides metadata of the EU-LFS and EU-SILC. A further initiative that recently has been completed is part of the EU project ‘Synergies for Europe’s Research Infrastructures in the Social Sciences’. In work package two, the sampling practices of European surveys haven been documented to compare and finally improve them (Scherpenzeel et al., 2017). The ongoing research project on survey data harmonisation of the Polish Academy of Sciences in cooperation with Ohio State University also devotes substantial effort to documenting and harmonising data related to democratic values and protest behaviours (Słomczyński et al., 2018). Unfortunately, this study was already underway, so the outcomes of these initiatives could only be used for cross-checking. Finally, the IPUMS-International project, a collaboration of the University of Minnesota, National Statistical Offices, international data archives as well as other international organisations, harmonises publicly available census data and provides a systematic inventory (Minnesota Population Center, 2019). Unfortunately, it does not (yet) offer a harmonised ISCED variable that can be used for cross-national comparisons. However, all these projects will facilitate future studies like this, as well as substantive (rather than methodological) studies that would like to control for the impact of a single survey characteristic.

4.7 Acknowledgements

I especially thank Silke Schneider, Michael Braun, Matthias Sand, Roberto Briceno-Rosas and Patricia Hadler for their helpful inputs and suggestions on this study as well as their feedback on earlier versions of this paper. I also would like to thank the editor and especially the four anonymous reviewers for their remarks which helped to improve this paper. I also thank my student assistants, Tim Dennenmoser, Svenja Friedel, Zeynep Umuc Demirci, Paweł Komendziński and Kimberly Herbst for their help with researching, collecting and coding the survey characteristics. I would also like to thank the UK Data Archive, Eurofound, the Norwegian Centre for Research Data, the GESIS Data Archive and the German Microdata Lab at GESIS for providing the data sets, the documentation of the survey characteristics and for their support. Finally, I thank Jane Roberts for proofreading.

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 programme under grant agreement No 654221.

4.8 References

Becker, G. S. (1993). *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education*. (3rd ed.). Chicago: The University of Chicago Press.

Bekhuis, H., Lubbers, M., & Verkuyten, M. (2014). How Education Moderates the Relation Between Globalization and Nationalist Attitudes. *International Journal of Public Opinion Research*, 26(4), 487–500. doi:10.1093/ijpor/edt037

Bethlehem, J. D., Cobben, F., & Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*. Hoboken: John Wiley & Sons.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Hoboken: John Wiley & Sons.

Blair, J., G. Menon, & Bickart, B. (2011). Measurement Effects in Self vs. Proxy Response to Survey Questions: An Information-Processing Perspective. In P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Matthiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp.145–166). Hoboken: John Wiley & Sons.

Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: John Wiley & Sons.

Bol, T., & van de Werfhorst, H. G. (2013). Educational Systems and the Trade-Off Between Labor Market Allocation and Equality of Educational Opportunity. *Comparative Education Review*, *57*(2), 285–308.

Bolzendahl, C. I. & Myers D. J. (2004). Feminist Attitudes and Support for Gender Equality: Opinion Change in Women and Men, 1974-1998. *Social Forces*, *83*(2), 759–789. doi: 10.1353/sof.2005.0005

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.

Breen, R., & Jonsson, J. O. (2000). Analyzing Educational Careers: A Multinomial Transition Model. *American Sociological Review*, *65*(5), 754–772. doi: 10.2307/2657545

Breen, R., & Jonsson, J. O. (2005). Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility. *Annual Review of Sociology* 31: 223–243. doi:10.1146/annurev.soc.31.041304.122232

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly* 57(1): 62–79. doi: 10.1086/269355

Coenders, M., & Scheepers, P. (2003). The Effect of Education on Nationalism and Ethnic Exclusionism : An International Comparison. *Political Psychology*, *24*(2), 313–343.

Couper, M. P. (2000). Web surveys – A Review of Issues and Approaches. *Public Opinion Quarterly*, 64(4), 464–494. doi: 10.1086/318641

Daikeler, J., Bosnjak, M., & Lozar Manfreda, K. (2019 online first) Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology*. doi:10.1093/jssam/smz008.

de Leeuw, E. D., & van der Zouwen, J. (2001). Data Quality in Telephone and Face-to-Face Surveys: A Comparative Meta-Analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.) *Telephone Survey Methodology* (pp. 283–300). New York: John Wiley & Sons.

Dever, J. A., Rafferty, A., & Valliant, R. (2008). Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? *Survey Research Methods*, 2(2), 47–62. doi:10.18148/srm/2008.v2i2.128

Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20(2), 210–217. doi:10.2307/2088328

Ehling, M. (2003). Harmonising Data in Official Statistics: Development, Procedures, and Data Quality. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Eds.), *Advances in Cross-National Comparison. A European Working Book of Demographic and Socio-Economic Variables* (pp. 17–31). New York: Kluwer Academic/Plenum Publishers.

ESS. (2014a). *Documentation of ESS Post-Stratification Weights*. Retrieved from: https://www.europeansocialsurvey.org/docs/methodology/ESS1-5_post_stratification _weights_documentation.pdf (accessed April 2020)

ESS. (2014b). *Weighting European Social Survey Data*. Retrieved from: https://www.europeansocialsurvey.org/docs/methodology/ESS_weighting_data_1.pdf (accessed April 2020)

Eurostat. (2019). *Smarter, Greener, More Inclusive? Indicators to Support The Europe 2020 Strategy*. Luxembourg: Publication Office of the European Union. Retreived from: https://ec.europa.eu/eurostat/documents/3217494/10155585/KS-04-19-559-EN-N.pdf/b8528d01-4f4f-9c1e-4cd4-86c2328559de (accessed April 2020)

Eurostat & OECD. (2014). *Joint Eurostat-OECD Guidelines on the Measurement of Educational Attainment in Household Surveys*. Version of September 2014. Retrieved from: http://ec.europa.eu/eurostat/documents/1978984/6037342/Guidelines-on-EA-final.pdf (accessed April 2020)

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta G. P., & Nelson, L. (2006). Experiments in Producing Nonresponse Bias. *Public Opinion Quarterly*, 70(5), 720–736. doi: 10.1093/poq/nfl036

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken: John Wiley & Sons.

Groves, R. M., & Lyberg, L. E. (2010). Total Survey Error – Past, Present, and Future. *Public Opinion Quarterly*, 74(5), 849–879. doi:10.1093/poq/nfq065

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. *Public Opinion Quarterly*, 72(2), 167–189. doi: doi.org/10.1093/poq/nfn011

Heath, A., Martin, J., & Spreckelsen, T. (2009). Cross-National Comparability of Surveys Attitude Measures. *International Journal of Public Opinion Research*, 21(3), 293–315. doi:10.1093/ijpor/edp034

Hoffmeyer-Zlotnik, J. H. P. (2008). Harmonisation of Demographic and Socio-Economic Variables in Cross-National Survey Research. *Bulletin de Méthodologie Sociologique*, 98(1), 5–24. doi:10.1177/075910630809800103

Hyman, H. H., & Wright, C. R. (1979). *Education's Lasting Influence on Values*. Chicago: University of Chicago Press.

Kalmijn, M. (2003). Country Differences in Sex-Role Attitudes: Cultural and Economic Explanations. In W. A. Arts, L. C. J. M. Halman, & J. A. P. Hagenaars (Eds.), *The Cultural Diversity of European Unity. Findings, Explanations and Reflections from the European Values Study* (pp. 311–337). Leiden: Koninklijke Brill NV.

Kaminska, O., & Lynn, P. (2017). Survey-Based Cross-Country Comparisons Where Countries Vary in Sample Design: Issues and Solutions. *Journal of Official Statistics,* 33(1), 123–136. doi:10.1515/JOS-2017-0007

Kieffer, A. (2010). Measuring and Comparing Levels of Education: Methodological Problems in the Classification of Educational Levels in the European Social Surveys and the French Labor Force Surveys. *Bulletin de Méthodologie Sociologique*, 107(1), 49–73. doi:10.1177/0759106310369974

Kleven, Ø., & Ringdal, K. (2017). *Level of Education – Measuring the Quality of Questions in Survey Interviews by Administrative Records on Education. Experiences from the Norwegian European Social Survey 2004 - 2014*. In Proceedings of the 7th Conference of the European Survey Research Association (ESRA), July 20, 2017. Lisbon.

Kohler, U. (2007). Surveys from Inside: An Assessment of Unit Nonresponse Bias With Internal Criteria. *Survey Research Methods,* 1(2), 55–67. doi:10.18148/srm/2007.v1i2.75

Kohler, U. (2008). Assessing the Quality of European Surveys - Towards an Open Method of Coordination for Survey Data. In J. Albers, T. Fahey, and C. Saraceno (Eds.), *Handbook of Quality of Life in the Enlarged European Union* (pp.405–423). London, New York: Routledge.

Kreuter, F., Eckman, S., Maaz, K., & Watermann, R. (2010). Children's Reports of Parents' Education Level: Does It Matter Whom You Ask and What You Ask About?. *Survey Research Methods,* 4(3), 127–138. doi:10.18148/srm/2010.v4i3.4283.

Lohr, S. L. (2009). *Sampling: Design and Analysis* (2nd ed.) Boston: Brooks/ Cole Cengage Learning.

Lyberg, L. E., & Kasprzyk, D. (2011). Data Collection Methods and Mesurement Error: An Overview." In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowwetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 235–257). Hoboken: John Wiley & Sons.

Minnesota Population Center. (2019). *Integrated Public Use Microdata Series, International*. Data File Version 7.2. Data Minneapolis, MN: IPUMS. doi:10.18128/D020.V7.2

Moore, J. C. (1988) Self/Proxy Response Status and Survey Response Quality: A Review of the Literature. *Journal of Official Statisics,* 4(2), 155–172.

Müller, W., & Karle, W. (1993). Social Selection in Educational Systems in Europe. *European Sociological Review*, *9*(1), 1–23. doi:10.1093/oxfordjournals.esr.a036652

Noelle-Neumann, E., & Petersen, T. (2000). Das halbe Instrument, Die halbe Reaktion. Zum Vergleich von Telefon- und Face-to-Face Umfragen. In V. Hüfken (Ed.), *Methoden in Telefonumfragen* (pp.183–200). Wiesbaden: Westdeutscher Verlag.

OECD. (2015). *Education at a Glance 2015: OECD Indicators*. OECD Publishing. Retrieved from: https://www.oecd-ilibrary.org/education/education-at-a-glance-2015_eag-2015-en (accessed April 2020)

OECD. (2016). *Education at a Glance 2016: OECD Indicators*. OECD Publishing. Retrieved from: https://www.oecd-ilibrary.org/education/education-at-a-glance-2016_eag-2016-en (accessed April 2020)

OECD. (2017). *Education at a Glance 2017: OECD Indicators*. OECD Publishing. Retrieved from: https://www.oecd-ilibrary.org/education/education-at-a-glance-2017_eag-2017-en (accessed April 2020)

Ortmanns, V., & Schneider, S. L. (2016a). Can We Assess Survey Representativeness of Cross-National Surveys Using the Education Variable? *Survey Research Methods*, 10(3): 189–210. doi: 10.18148/srm/2016.v10i3.6608

Ortmanns, V., & Schneider, S. L. (2016b). Harmonization Still Failing ? Inconsistency of Education Variables in Cross-National Public Opinion Surveys. *International Journal of Public Opinion Research*, 28(4), 562–82. doi:10.1093/ijpor/edv025

Peytcheva, E., & Groves, R. M. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. Journal of Official Statistics, 25(2), 193–201.

Schanze, J.-L. (2017). *Report on Sampling Practices for the Institutionalized Population in Social Surveys*. Deliverable 2.16 of the SERISS project funded under the European Union’s Horizon 2020 research and innovation programme GA No: 654221. Retrieved from: www.seriss.eu/resources/deliverables (accessed April 2020)

Scherpenzeel, A., Maineri, A. M., Bristle, J., Pflüger, S., Mindarova, I., Butt, S., Zins, S., Emery, T. & Luijkx, R. (2017). *Report on the Use of Sampling Frames in European Studies*. Deliverable 2.1 of the SERISS project funded under the European Union’s Horizon 2020 research and innovation programme GA No: 654221. Retrieved from: www.seriss.eu/resources/deliverables (accessed April 2020)

Schneider, S. L. (2009). *Confusing Credentials: The Cross-Nationally Comparable Measurement of Educational Attainment* (Doctoral thesis). Oxford: University of Oxford, Nuffield College. Retrieved from: http://ora.ox.ac.uk/objects/uuid:15c39d54-f896-425b-aaa8-93ba5bf03529 (accessed April 2020)

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. Fu (Eds.), The SAGE Handbook of Survey Methodology (pp. 288–307). London: SAGE Publications.

Semyonov, M. Raijman, R. & Gorodzeisky, A. (2008) Foreigners' Impact on European Societies Public Views and Perceptions in a Cross-National Comparative Perspective. *International Journal of Comparative Sociology*, 49(1), 5–29. doi: 10.1177/0020715207088585

Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112–141. doi:10.1177/0002716212458082

Słomczyński, K. M., Jenkins, J. C., Tomescu-Dubrow, I., Kołczyńska, M., Wysmułek, I., Oleksiyenko, O., Powałko, P., & Zielinski, M. W. (2018). *Survey Data Recycling Dataverse - Master Box*. Data File Version 1.1. Harvard Dataverse. doi: 10.7910/DVN/VWGF5Q

Słomczyński, K. M., Tomescu-Dubrow, I., Jenkins, J. C., Kołczyńska, M., Powałko, P., Wysmułek, I., Oleksiyenko, O., Zielinski, M. W. & Dubrow, J. K. (2016). *Democratic Values and Potest Behavior - Harmonization of Data from International Survey Projects*. Warsaw: IFiS Publishers.

Smith, T. W. (1995). Some Aspects of Measuring Education. *Social Science Research*, 24(3), 215–242. doi:10.1006/ssre.1995.1008

Smith, T. W. (2010) The Globalization of Survey Research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 477–484). Hoboken: John Wiley & Sons.

Smith, T. W. (2011). Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research*, 23(4), 464–484. doi: 10.1093/ijpor/edq052.

Sodeur, W. (1997). Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. In Zentralarchiv Für Empirische Sozialforschung (Ed.), *ZA-Information* 41, 58–82. Cologne. Retrieved from: https://www.ssoar.info/ssoar/handle/document/19999 (accessed April 2020)

Sodeur, W. (2007). Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl - Ein Vergleich von ALLBUS-Erhebungen. *MDA - Methoden Daten Analysen*, 1(2), 107–130.

UNESCO-UIS. (2006). *ISCED 1997 International Standard Classification of Education*. Montreal: UNESCO-UIS. Retrieved from: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-1997-en_0.pdf (accessed April 2020)

van Tuyckom, C., & Bracke, P. F. (2014). Survey Quality and Cross-National Sports Research: A Case Study of the 2007 ISSP Survey. *European Journal of Sport Science*, 14(1), 228–234. doi:10.1080/17461391.2012.683814

Weakliem, D. L. (2002). The Effects of Education on Political Opinions: An International Study. *International Journal of Public Opinion Research*, 14(2), 141–157. doi: 10.1093/ijpor/14.2.141

Data References

ESS. (2016a). *European Social Survey, Round 4, Data 2008.* Data File Version 4.4. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2016b*). European Social Survey, Round 5, Data 2010.* Data File Version 3.3. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

ESS. (2016c). *European Social Survey, Round 6, Data 2012.* Data File Version 2.3. Norway: Norwegian Centre for Research Data, Data Archive and Distributor of ESS Data for ESS ERIC.

Eurofound. (2011). *European Working Conditions Survey (EWCS) 2010*. Study number 6971. Data File Version 1. Colchester: UK Data Archive.

Eurofound. (2014). *European Quality of Life Survey (EQLS) 2011-2012*. Study number 7316. Data File Version 3. Colchester: UK Data Archive.

European Commission. (2012). *Eurobarometer 73.2+73.3 (2-3 2010)*. ZA5236. Data File Version 2.0.1. Cologne: GESIS Data Archive. doi:10.4232/1.11473

Eurostat. (2010). *European Union Statistics on Income and Living Conditions (EU-SILC) 2010*. User Database. Data File Version CROSS-2010-6.

Eurostat. (2011). *Adult Education Survey (AES) 2011. User Database. D*ata File Version 1.

Eurostat. (2016a). *European Union Labour Force Survey (EU-LFS) 2008*. User database. Data File Version 2016.

Eurostat. (2016b). *European Union Labour Force Survey (EU-LFS) 2010*. User database. Data File Version 2016.

Eurostat. (2016c). *European Union Labour Force Survey (EU-LFS) 2011*. User database. Data File Version 2016.

Eurostat. (2016d). *European Union Labour Force Survey (EU-LFS) 2012*. User database. Data File Version 2016.

EVS. (2016). *European Values Study 2008, 4th wave. Integrated Dataset*. ZA4800. Data File Version 4.0.0. Cologne: GESIS Data Archive. doi:10.4232/1.12458

ISSP - Research Group (2015). *International Social Survey Programme: Health and Health Care - ISSP 2011*. ZA5800. Data File Version 3.0.0. Cologne: GESIS Data Archive. doi: 10.4232/1.12252

ISSP - Research Group (2016). *International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012*. ZA5900. Data File Version 4.0.0, Cologne: GESIS Data Archive. doi: 10.4232/1.12661

OECD. (2013). *Programme for the International Assessment of Adult Competencies (PIAAC)* 2011 - *Public Use File*. Data File Version 1.0.

Ortmanns, V. (2020). Education Distributions and Survey Characteristics of Ten Cross-National Surveys. GESIS, SowiDataNet|datorium. Data File Version: 1.0.0, Cologne: GESIS Data Archive. doi: 10.7802/1.2002

4.9 Supplemental Material

Table A4.3 Categories and recodes of the education variables across surveys into 5-level version of ISCED 97

5-level version of ISCED 97		**EU-LFS**		**EU-SILC**		**AES**		**PIAAC**		**Eurobarometer**		**EQLS**		**EWCS**		**ESS**		**EVS**	
1	Pre-primary and primary or first stage of basic education	0	No formal education or below ISCED 1	0	Pre-primary education	1	No formal education or below ISCED 1	1	Primary or less (ISCED 1 or less)	0	Pre-primary education	0	No education completed (ISCED 0)	0	Pre-primary education	0	Not completed ISCED level 1	0	Pre-primary education or none education
		11	ISCED 1	1	Primary education	11	ISCED 1			1	Primary education or first stage of basic education	1	Primary education (ISCED 1)	1	Primary education or first stage of basic education	113	ISCED 1, completed primary education	1	Primary education or first stage of basic education
																129	Vocational ISCED 2C < 2 years, no access ISCED 3		
2	Lower secondary or second stage of basic education	21	ISCED 2	2	Lower secondary education	21	ISCED 2	2	Lower secondary (ISCED 2, ISCED 3C short)	2	Lower secondary or second stage of basic education	2	Lower secondary education (ISCED 2)	2	Lower secondary or second stage of basic education	212	General/pre-vocational ISCED 2A/2B, access ISCED3 vocational	2	Lower secondary or second stage of basic education
																213	General ISCED 2A, access ISCED 3A general/all 3		
																221	Vocational ISCED 2C >= 2 years, no access ISCED 3		
																222	Vocational ISCED 2A/2B, access ISCED 3 vocational		
																223	Vocational ISCED 2, access ISCED 3 general/all		
		22	ISCED 3c (shorter than 2 years)			22	ISCED 3c (shorter than 2 years)									229	Vocational ISCED 3C < 2 years, no access ISCED 5		

5-level version of ISCED 97	EU-LFS	EU-SILC	AES	PIAAC	Eurobarometer	EQLS	EWCS	ESS	EVS
3 (Upper) Secondary education	30 ISCED 3 (without distinction a, b or c possible, 2 years and more)	3 (Upper) Secondary education	30 ISCED 3 (without distinction a, b or c possible, 2 years and more)	3 Upper secondary (ISCED 3A-B, C long)	3 (Upper) secondary education	3 Upper secondary education (ISCED 3)	3 (Upper) secondary education		3 (Upper) secondary education
	31 ISCED 3c (2 years and more)		31 ISCED 3c (2 years and more)					311 General ISCED 3 >=2 years, no access ISCED 5	
								321 Vocational ISCED 3C >= 2 years, no access ISCED 5	
	32 ISCED 3 a, b		32 ISCED 3 a, b					312 General ISCED 3A/3B, access ISCED 5B/lower tier 5A	
								322 Vocational ISCED 3A/3B, access 5B/lower tier 5A	
								313 General ISCED 3A, access upper tier ISCED 5A/all 5	
								323 Vocational ISCED 3A, access upper tier ISCED 5A/all 5	

5-level version of ISCED 97	EU-LFS	EU-SILC	AES	PIAAC	Eurobarometer	EQLS	EWCS	ESS	EVS
4 Post-secondary non-tertiary education	43 ISCED 4 (without distinction a, b or c possible)	4 Post-secondary non-tertiary education	40 ISCED 4	4 Post-secondary, non-tertiary (ISCED 4A-B-C)	4 Post-secondary, non-tertiary education	4 Post-secondary including pre-vocational or vocational education but not tertiary (ISCED 4)	4 Post-secondary, non-tertiary education		4 Post-secondary non-tertiary education
	42 ISCED 4c							421 ISCED 4 programmes without access ISCED 5	
	41 ISCED 4a, b							412 General ISCED 4A/4B, access ISCED 5B/lower tertiary 5A	
								413 General ISCED 4A, access upper tier ISCED 5A/all 5	
								422 Vocational ISCED 4A/4B, access ISCED 5B/lower tertiary 5A	
								423 Vocational ISCED 4A, access upper tier ISCED 5A /all 5	
5 First and second stage of tertiary education	51 ISCED 5b	5 1st & 2nd stage of tertiary education	50 ISCED 5, 6	8 Tertiary - bachelor/ master/research degree (ISCED 5A/6)	5 First stage of tertiary education	5 Tertiary education – first level (ISCED 5)	5 First stage of tertiary education	510 ISCED 5A short, intermediate/academic/general tertiary below	5 First stage of tertiary education
				5 Tertiary – professional degree (ISCED 5B)				520 ISCED 5B short, advanced vocational qualifications	

5-level version of ISCED 97	EU-LFS		EU-SILC	AES	PIAAC		Eurobarometer		EQLS		EWCS		ESS		EVS	
	52	ISCED 5a			6	Tertiary – bachelor degree (ISCED 5A)							620	ISCED 5A medium, bachelor/equivalent from upper/single tertiary		
					7	Tertiary - master/research degree (ISCED 5A/6)							710	ISCED 5A long, master/equivalent from lower tertiary		
													720	ISCED 5A long, master/equivalent from upper/single tertiary		
	60	ISCED 6					6	Second stage of tertiary education	6	Tertiary education - advanced level (ISCED 6)	6	Second stage of tertiary education	800	ISCED 6, doctoral degree	6	Second stage of tertiary education

Data sources: EU-LFS 2008-2012, files from Eurostat, data file versions 2016, variable HATLEVEL; EU-SILC 2010, file from Eurostat, data file version CROSS-2010-6, variable PE040; AES 2011, files from Eurostat, data file version 1.0, variable HATLEVEL; PIAAC 2011, file from OECD, data file version of 2013, variable edcat7; Eurobarometer 73.2 & 73.3 (2010), file from Eurostat, data file version 2.0.1 of 2012, variable v362; EQLS 2012, data file version 3 of 2014, variable Y11_ISCEDsimple; EWCS 2010, data file version of 2011, variable ef1_isced; ESS 2012, data file version 2.3 of 2016, variable edulvlb; EVS 2008, data file version 4.0.0 of 2016, variable v336

Table A4.4 Categories and recodes of the education variables in ISSP and EU-LFS into 4-level version of ISCED 97

	4-level version of ISCED 97	**EU-LFS**		**ISSP since 2011**	
1	Pre-primary and primary or first stage of basic education	0	No formal education or below ISCED 1	0	No formal education
		11	ISCED 1	1	Primary school
2	Lower secondary or second stage of basic education	21	ISCED 2	2	Lower secondary (secondary education completed that does not allow entry to university: end of obligatory school but also short programs (less than 2 years))
		22	ISCED 3c (shorter than 2 years)		
3	(Upper) Secondary education and post-secondary non-tertiary education	32	ISCED 3 a, b	3[a]	Upper secondary (programs that allow entry to university)
		41	ISCED 4 a, b		
		30	ISCED 3 (without distinction a, b or c possible, 2 years and more)	4	Post-secondary, non-tertiary (other upper secondary programs toward the labour market or technical formation)
		43	ISCED 4 (without distinction a, b or c possible)		
		31	ISCED 3c (2 years and more)		
		42	ISCED 4c		
4	First and second stage of tertiary education	51	ISCED 5b	5	Lower level tertiary, first stage (also technical schools at a tertiary level)
		52	ISCED 5a		
				6	Upper level tertiary (Master, Dr.)
		60	ISCED 6		

Notes:[a] ISCED 3B and 4B are included in ISSP DEGREE variable category 4, not 3, which cannot be differentiated in the ESS. Therefore ISCED 3 and 4 are summarised.
Data sources: EU-LFS 2011-2012, files from Eurostat, data file versions 2016, variable HATLEVEL; ISSP 2012, data file version 4.0.0 of 2016, variable DEGREE

Table A4.5 Duncan's Dissimilarity Index for educational attainment distributions across surveys and years per country

Survey **Year**	**SILC-LFS** **2010**	**AES-LFS** **2011**	**PIAAC[c]-LFS** **2011**	**EB-LFS** **2010**	**EQLS-LFS** **2012**	**EWCS-LFS** **2010**	**ESS-LFS** **2012**	**EVS-LFS** **2008**	**ISSP-LFS[d]** **2012**	**Mean**
AT	1.60	2.60	1.53	16.63	11.06	14.95		12.78	50.32	13.93
BE[a]	4.26	6.32	9.25	16.46	6.67	9.61	10.77	8.33	8.20	8.87
BG	0.80	2.01		9.98	10.40	8.13	2.84	5.23	4.48	5.49
CH	8.03[f]	4.82					3.34	10.81	4.34	6.27
CY	4.92	3.37	7.69[f]	15.83	11.64	6.03	4.21	19.27	24.43	10.82
CZ	0.75	3.34	5.59[f]	10.50	15.77	12.44	13.53	3.10		8.13
DE	4.12	5.78	6.54	26.13	20.65	58.96	10.04	10.66	6.34	16.58
DK	3.37[f]	7.36	9.86	12.06	24.32	18.36	20.06	10.79	31.93	15.35
EE	3.73	5.52	6.71[f]	22.59	35.71	29.80	11.10	34.90		18.76
ES	3.05	16.20	6.36[f]	24.09	22.34	21.15	13.90	21.63	10.45	15.46
FI	7.56	2.89[f]	7.82[f]	17.76	15.81	8.72	13.64	28.38	11.91	12.72
FR	8.34	5.00	2.69[f]	9.87	10.79	7.64	13.49	14.50	29.32[f]	11.29
GB[b]	11.13	1.99 [f]	12.95	22.91	27.95	34.38	17.77	20.76	16.38	18.47
GR	9.01	4.38		12.44	15.01	13.38	13.26[e]	14.98		11.53
HR	5.93[f]				17.56	36.95	5.99[e]	6.98	11.90	15.86
HU	3.34	30.09[f]		42.70	6.84	8.41	3.19	5.18	29.90	16.21
IE	5.61[f]	1.59[f]	8.16	17.61	4.58	14.97	8.90	12.52	12.41	9.59
IS	8.79			19.06	20.79		24.83	18.02	25.28	19.46
IT	3.42	8.92	7.69	16.62	11.13	11.71	12.66	15.71	14.20[e]	10.98
LT	5.17	2.85		10.89	19.74	14.28	11.90	21.24	10.63	12.09
LU	14.47	7.35[f]		8.49	7.56	6.44		7.00		8.55
LV	3.89	3.63		30.94	11.49	22.13	7.28[e]	20.30	1.72	13.44
MT	10.59	4.60		43.38	9.58	13.04				16.24
NL	3.77	4.61	4.45	38.09	13.16	16.59	13.92	22.92	14.09	14.62
NO	1.78[f]	39.91	13.45	33.53		18.42	13.32	20.22[f]	20.49	20.14
PL	11.88	12.76	7.25	5.02	17.24	11.78	30.53	14.22[f]	9.07	13.31
PT	3.52	5.58		8.43	3.66	5.75	5.66	5.17	12.12	6.24
RO	1.65	1.27		14.39	8.81	10.01	10.56[e]	10.13		7.71
SE	7.50	11.64	2.79	21.03	14.64	15.37	10.20	22.13	25.48	14.53
SI	1.90	4.27		6.01	2.69	3.67	3.37	26.29	23.59	8.97
SK	5.90	24.17	10.62	4.61	5.75	4.77	12.27	4.05	31.94	11.56
Median	4.26	4.82	7.47	16.62	11.64	13.04	12.08	14.36	13.25	12.72
Mean	5.48	8.10	7.30	18.55	13.91	15.79	11.89	14.94	17.78	12.68
Min	0.75	1.27	1.53	4.61	2.69	3.67	2.84	3.10	1.72	5.49
Max	14.47	39.91	13.45	43.38	35.71	58.96	30.53	34.90	50.32	20.14

Notes

[a] For PIAAC and EU-LFS only Flanders, excluding Wallonia and Brussels;

[b] For PIAAC and EU-LFS only England and Northern Ireland, excluding Scotland and Wales; for ISSP and EU-LFS excluding Northern Ireland; for AES and EU-LFS only England, excluding Scotland, Wales and Northern Ireland;

[c] For PIAAC, DE and AT use age group 25 to 65 instead of 25 to 64;

[d] For ISSP, adapted ISCED97_4 level is used (see Table A4.4);
[e] For GR and HR data of ESS data retrieved from 2010, for LV and RO of ESS data retrieved from 2008, for IT in ISSP data retrieved from 2011;
[f] Not included in the analysis due to missing information on survey characteristics

Data sources:
EU-LFS 2008-2012, files from Eurostat, data file versions 2016, variable HATLEVEL, weighted using variable COEFF;
EU-SILC 2010, file from Eurostat, data file versions CROSS-2011-6, variable PE040, weighted using variable PB040;
AES 2011, file from Eurostat, data file version 1.0, variable HATLEVEL;
PIAAC 2011, file from OECD, data file version 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB);
Eurobarometer 73.2 & 73.3 (2010), file from Eurostat, data file versions 2.0.1, variable v362, data weighted to correct regional oversampling for Germany and the UK;
EQLS 2012, data file version 3 of 2014, variable Y11_ISCEDsimple, weighted using variable w1
EWCS 2010, data file version of 2011, variable ef1_isce, weighted using variable w1;
ESS 2012, data file version 2.3 of 2016, variable edulvlb, weighted using variable dweight; for Greece and Croatia data from 2010 were used (data file versions: 3.3., variable edulvlb) and for Latvia and Romania data from 2008 were used (data file version 4.4, variable edulvla);
EVS 2008, data file version 4.0.0, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK;
ISSP 2012, data file version 4.0.0 of 2016, variable DEGREE, data weighted to correct regional oversampling for Germany; for Italy data from 2011 were used (data file version 3.0.0, variable DEGREE);
Only respondents aged 25-64 for all surveys, apart from DE and AT in PIAAC including age 65.

Table A4.6 Overview of the surveys and the participating countries

survey	year	number of European countries	participating European countries
European Labour Force Survey (EU-LFS)	2008-2012	30-31	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MT (not in 2008), NL, NO, PL, PT, RO, SE, SI, SK, UK
European Union Statistics on Income and Living Conditions (EU-SILC)	2010	31	AT, BE, BG, CH*, CY, CZ, DE, DK*, EE, ES, FI, FR, GB, GR, HR*, HU, IE*, IS, IT, LT, LU, LV, MT, NL, NO*, PL, PT, RO, SE, SI, SK
Adult Education Survey (AES)	2011	29	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI*, FR, GB (only England)*, HU*, IE*, IS, IT, LT, LU*, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK
Programme for the International Assessment of Adult Competencies (PIAAC)	2011	18	AT, BE (only Flanders), CY*, CZ*, DE, DK, EE*, ES*, FI*, FR*, GB (only England and Northern Ireland) IE, IT, NL, NO, PL, SE, SK
Standard Eurobarometer 73.2&73.3	2010	29	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK
European Quality of Life Survey (EQLS)	2012	29	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MT, NL, PL, PT, RO, SE, SI, SK
European Working Condition Survey (EWCS)	2010	29	AT, BE, BG, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK
European Social Survey (ESS)	2012	28	BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR (from 2010), HR (from 2010), HU, IE, IS, IT, LT, LV (from 2008), NL, NO, PL, PT, RO (from 2008), SE, SI, SK
European Values Study (EVS)	2008	30	AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, NL, NO*, PL*, PT, RO, SE, SI, SK

Notes on countries abbreviations: AT=Austria, BE=Belgium, BG=Bulgaria, CH=Switzerland, CY = Cyprus (Republic only), CZ= Czech Republic, DE=Germany, DK=Denmark, EE=Estonia, ES=Spain, FI=Finland, FR=France, GB=United Kingdom, GR=Greece, HR=Croatia, IE=Ireland, IS=Iceland, IT=Italy, LT=Lithuania, LU=Luxembourg, LV=Latvia, NL=Netherlands, No=Norway, PL=Poland, PT=Portugal, RO=Romania, SE=Sweden, SI=Slovenia, SK=Slovakia
*not included in the analysis due to missing information on survey characteristics

***Annexe 1:* Information on the adaptations made to the variables on marital status and household composition when calculating Sodeur's Index**

For Denmark (until 2009), Iceland, Sweden and Switzerland in the EU-LFS no information on the household composition is provided, so we include all married couples. To render fair comparisons, we drop the condition of two-person households for those countries in all other surveys as well. Unfortunately, the EVS did not ask about household composition either. Again we drop that condition for all countries when calculating the index for EVS and EU-LFS. In general, we do not consider same-sex couples because the index does not work for them.

***Annexe 2*: Information on the index generated to assess consistency of the response categories across surveys and countries**

The index assesses the similarity between the response categories of the EU-LFS and those in the other surveys, and it distinguishes whether the categories are the same, similar or different. The coding rules assigned response categories as equal when the terms and the number of categories are identical or only minor changes in the order are identified, e.g. two categories are in reversel order.

According to the coding rules, response categories are similar if they use slightly different terms for the same qualification. For instance in Poland the category of bachelor or equivalent degree is named in the EU-LFS ‘university - bachelor degree or engineer’ (szkoły wyższej - studia licencjackie lub inżynierskie). In the EVS the category emphasizes the status of bachelor degree by mentioning that this is below a master’s degree. Besides, a more general term for including vocational programmes (zawodowe) is used instead of explicitly mentioning engineer (wyższe licencjackie lub zawodowe - bez magisterium). We also allow the number of categories to differ because they can be split or aggregated. For instance, Portugal in the EVS distinguishes between general and technological upper secondary education (‘Ensino Secundário Cursos Gerais’ and ‘Ensino Secundário Cursos Tecnológicos’, whereas in the EU-LFS these are included in one category (‘Secundário’). Response categories are also similar when they are ordered slightly differently and when a single category is missing or added, but overall no more than three categories are changed. For Belgium we decide that the response categories between the EU-LFS and the AES are similar; although the EU-LFS asks four questions and the AES only one, the names of the qualifications and also the order is nearly identical.

Lastly, we code response categories as different, if the wording of the qualification across surveys is different. For example, Lithuania in the Eurobarometer offers one category of ‘vocational school’ (‘Profesinė mokykla’), whereas the EU-LFS offers five categories specifying the different programmes and levels of that school because those are also coded differently in ISCED. Another example for the code different can be found in the EWCS for Germany where the education question excludes qualifications of the vocational training. Perhaps the education question follows different scope and focuses only on general education. Response categories that are regarded as different also often have different numbers of categories as well as a different order. In Spain, the EU-LFS, the EU-SILC and the ISSP ask an open-ended question rather than offering answer categories whereas all other surveys provide a list of categories. This is another example of when the code for the ‘different’ category is assigned. It is important to note that the category ‘different’ does not tell us whether the response categories are more or less detailed than those of the EU-LFS. We also cannot assess what amount of detail is helpful for the respondents and when additional information is confusing. This also depends on the education system of the countries.

Two persons independently assigned the codes to the index; the overlap of the two coders was 80%. All differing coding decisions were discussed and a final code was agreed.

Annexe 3: Tables on basic statistical descriptives of for each survey characteristic

Number of countries with different sampling designs and final sampling units

survey comparison	*sampling design*		*final sampling unit*		
	simple design	complex design	individual	household	address/ dwelling
EU-LFS 2008	11	19	5	10	15
EU-LFS 2010	10	21	5	10	16
EU-LFS 2011	10	21	5	10	16
EU-LFS 2012	10	21	5	10	16
EU-SILC 2010	7	19	4	10	12
AES 2012	7	17	15	4	5
PIAAC 2011	6	6	12	0	0
EB 2010	0	29	29	0	0
EQLS 2012	2	27	29	0	0
EWCS 2010	2	27	29	0	0
ESS 2012	5	23	28	0	0
EVS 2008	3	25	28	0	0
ISSP 2012	6	18	24	0	0

Number of countries with differences in mandatory survey participation

survey comparison	*survey participation*	
	mandatory	voluntary
EU-LFS 2008	11	19
EU-LFS 2010	13	18
EU-LFS 2011	13	18
EU-LFS 2012	13	18
EU-SILC 2010	0	26
AES 2012	9	15
PIAAC 2011	0	12
EB 2010	0	29
EQLS 2012	0	29
EWCS 2010	0	29
ESS 2012	0	28
EVS 2008	0	28
ISSP 2012	0	24

Distribution of the sample size across countries

survey comparison	*sample size*			
	min	max	median	mean
EU-LFS 2008	7,752	355,771	60,369	80,716
EU-LFS 2010	8,915	348,224	48,296	83,179
EU-LFS 2011	8,915	340,214	46,903	78,700
EU-LFS 2012	8,966	313,244	48,094	84,118
EU-SILC 2010	4,498	26,129	9,246	11,162
AES 2012	2,404	22,522	5,246	6,910
PIAAC 2011	3,507	7,505	4,486	4,729
EB 2010	322	1,002	676	648
EQLS 2012	572	1,818	688	853
EWCS 2010	816	3,505	897	1,122
ESS 2012	481	1,888	1,285	1,237
EVS 2008	603	1,439	980	982
ISSP 2012	606	1,816	800	894

Distribution of the response rates across countries

survey comparison	*response rate (%)*			
	min	max	median	mean
EU-LFS 2008	32.0	97.1	80.9	79.7
EU-LFS 2010	31.4	97.5	82.0	78.6
EU-LFS 2011	32.7	97.9	80.6	78.2
EU-LFS 2012	28.2	98.2	78.7	77.7
EU-SILC 2010	57.3	97.1	85.9	82.3
AES 2012	43.4	94.5	65.0	68.1
PIAAC 2011	45.0	72.0	56.0	57.3
EB 2010			not available	
EQLS 2012	14.0	78.3	44.0	44.1
EWCS 2010	31.3	73.5	43.8	46.4
ESS 2012	33.8	77.1	58.3	60.3
EVS 2008	24.4	87.2	53.4	54.4
ISSP 2012	25.1	72.6	52.2	48.1

Distribution of the fieldwork duration across countries

survey comparison	*fieldwork duration (days)* min	max	median	mean
EU-LFS 2008	21	90	90	78
EU-LFS 2010	21	90	90	78
EU-LFS 2011	21	90	90	78
EU-LFS 2012	21	90	90	78
EU-SILC 2010	30	334	137	153
AES 2012	15	278	98	120
PIAAC 2011	181	285	244	237
EB 2010	11	19	17	16
EQLS 2012	41	136	80	83
EWCS 2010	28	216	86	91
ESS 2012	49	234	126	132
EVS 2008	7	244	95	111
ISSP 2012	8	265	68	90

Distribution of the index on age and gender and of Sodeur's index across countries

survey comparison	*Index on age and gender*		*Sodeur's Index*	
	min	max	equal	different
EU-LFS - EU-SILC	0.4	4.8	25	1
EU-LFS - AES	0.7	15.6	12	12
EU-LFS PIAAC	1.6	16.1	12	0
EU-LFS - EB	4.3	16.6	29	0
EU-LFS - EQLS	3.0	21.3	26	3
EU-LFS - EWCS	5.2	17.8	29	0
EU-LFS - ESS	1.6	11.4	28	0
EU-LFS - EVS	2.4	19.4	24	4
EU-LFS - ISSP	3.7	23.1	24	0

Number of countries with differences in the response categories of the education question

survey comparison	*similarity of the response categories of the education question* same	similar	different
EU-LFS - EU-SILC	7	8	11
EU-LFS - AES	8	8	8
EU-LFS PIAAC	1	3	8
EU-LFS - EB	0	3	26
EU-LFS - EQLS	0	2	27
EU-LFS - EWCS	0	2	27
EU-LFS - ESS	0	4	24
EU-LFS - EVS	0	4	24
EU-LFS - ISSP	1	2	21

Number of countries with differences in using proxy-reporting and register information

survey comparison	*proxy-reporting*		*using register information*	
	yes	no	yes	no
EU-LFS 2008	30	0	3	27
EU-LFS 2010	31	0	3	28
EU-LFS 2011	31	0	3	28
EU-LFS 2012	31	0	3	28
EU-SILC 2010	25	1	0	26
AES 2012	7	17	0	24
PIAAC 2011	0	12	0	12
EB 2010	0	29	0	29
EQLS 2012	0	29	0	29
EWCS 2010	0	29	0	29
ESS 2012	0	28	0	28
EVS 2008	0	28	0	28
ISSP 2012	0	24	0	24

Number of countries with differences in the centralisation of ISCED coding and applying official ISCED mappings

survey comparison	*centralisation of ISCED coding*			*applying official ISCED mappings*		
	decentralised	partly-central	entirely central	yes	accidental deviation	intended deviation
EU-LFS 2008	30	0	0	30	0	0
EU-LFS 2010	31	0	0	31	0	0
EU-LFS 2011	31	0	0	30	1	0
EU-LFS 2012	31	0	0	30	1	0
EU-SILC 2010	26	0	0	26	0	0
AES 2012	24	0	0	21	3	0
PIAAC 2011	12	0	0	12	0	0
EB 2010	0	0	29	22	7	0
EQLS 2012	29	0	0	22	6	1
EWCS 2010	29	0	0	21	7	1
ESS 2012	2	26	0	22	1	5
EVS 2008	28	0	0	23	4	1
ISSP 2012	24	0	0	13	11	0

Number of countries with different modes of data collection

survey comparison	*modes of data collection*			
	f2f	telephone	self-administered	mixed
EU-LFS 2008	21	7	0	2
EU-LFS 2010	20	7	0	4
EU-LFS 2011	20	7	0	4
EU-LFS 2012	20	7	0	4
EU-SILC 2010	17	4	1	4
AES 2012	14	2	1	7
PIAAC 2011	12	0	0	0
EB 2010	29	0	0	0
EQLS 2012	29	0	0	0
EWCS 2010	29	0	0	0
ESS 2012	28	0	0	0
EVS 2008	26	0	2	0
ISSP 2012	15	8	0	1

Number of countries with different fieldwork agencies

survey comparison	*fieldwork agency*		
	commercial institute	institute of public authority	university/ scientific institute
EU-LFS 2008	0	30	0
EU-LFS 2010	0	31	0
EU-LFS 2011	0	31	0
EU-LFS 2012	0	31	0
EU-SILC 2010	0	26	0
AES 2012	4	20	0
PIAAC 2011	5	7	0
EB 2010	29	0	0
EQLS 2012	29	0	0
EWCS 2010	28	1	0
ESS 2012	19	4	5
EVS 2008	22	3	3
ISSP 2012	16	2	6

Data sources:
EU-LFS 2008-2012, files from Eurostat, data file versions 2016, variable HATLEVEL, weighted using variable COEFF;
EU-SILC 2010, file from Eurostat, data file versions CROSS-2011-6, variable PE040, weighted using variable PB040;
AES 2011, file from Eurostat, data file version 1.0, variable HATLEVEL;
PIAAC 2011, file from OECD, data file version 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB);
Eurobarometer 73.2 & 73.3 (2010), file from Eurostat, data file versions 2.0.1 of 2012, variable v362, data weighted to correct regional oversampling for Germany and the UK;
EQLS 2012, data file version 3 of 2014, variable Y11_ISCEDsimple, weighted using variable w1;
EWCS 2010, data file version of 2011, variable ef1_isce, weighted using variable w1;
ESS 2012, data file version 2.3 of 2016, variable edulvlb, weighted using variable dweight; for Greece and Croatia data from 2010 were used (data file versions: 3.3., variable edulvlb) and for Latvia and Romania data from 2008 were used (data file version 4.4, variable edulvla);
EVS 2008, data file version 4.0.0 of 2016, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK;
ISSP 2012, data file version 4.0.0 of 2016, variable DEGREE, data weighted to correct regional oversampling for Germany; for Italy data from 2011 were used (data file version 3.0.0, variable DEGREE)

Annexe 4: Crosstabulations of selected survey characteristic and their association

Crosstabulation of sampling unit and sampling design

sampling design/ sampling unit	equal	unequal	total
equal	42	16	58
unequal	133	38	171
total	175	54	229

Cramér's V = -0.06

Crosstabulation of mandatory survey participation and sampling design

sampling design/ mandatory participation	equal	unequal	total
equal	95	40	135
unequal	80	14	94
total	175	54	229

Cramér's V = -0.17

Crosstabulation of Sodeur's index and sampling design

sampling design/ Sodeur's index	equal	unequal	total
equal	157	52	209
unequal	18	2	20
total	175	54	229

Cramér's V = -0.10

Crosstabulation of response rate and mandatory survey participation

mandatory participation/ response rate	equal	unequal	total
equal	13	2	15
higher, < 30%	19	3	22
lower, < 30%	52	33	85
lower, ≥ 30%	30	39	69
not available	21	17	38
total	135	94	229

Cramér's V = 0.29

Crosstabulation of proxy-reporting and mandatory survey participation

mandatory participation/ proxy-reporting	equal	unequal	total
equal	17	15	32
unequal	118	79	197
total	135	94	229

Cramér's V = -0.05

Crosstabulation of Sodeur's index and of mandatory survey participation

mandatory participation/ Sodeur's index	equal	unequal	total
equal	125	84	209
unequal	10	10	20
total	135	94	229

Cramér's V = 0.06

Crosstabulation of response rate and proxy-reporting

proxy-reporting/ response rate	equal	unequal	total
equal	7	8	15
higher, < 30%	11	11	22
lower, < 30%	8	77	85
lower, ≥ 30%	2	67	69
not available	4	34	38
total	32	197	229

Cramér's V = 0.45

Crosstabulation of response rate and fieldwork duration

fieldwork duration/ response rate	equal	lower, < 90 days	higher, < 90 days	higher, ≥ 90 days	total
equal	1	3	4	7	15
higher, < 30%	4	6	9	3	22
lower, < 30%	12	24	25	24	85
lower, ≥ 30%	5	23	30	11	69
not available	0	32	3	3	38
total	22	88	71	48	229

Cramér's V = 0.28

Crosstabulation of response rate and Sodeur's index

Sodeur's index/ response rate	equal	unequal	total
equal	13	2	15
higher, < 30%	18	4	22
lower, < 30%	76	9	85
lower, ≥ 30%	64	5	69
not available	38	0	38
total	209	20	229

Cramér's V = 0.18

Crosstabulation of response rate and mode

mode/ response rate	equal	unequal	total
equal	11	4	15
higher, < 30%	10	12	22
lower, < 30%	54	31	85
lower, ≥ 30%	48	21	69
not available	22	16	38
total	145	84	229

Cramér's V = 0.15

Crosstabulation of fieldwork duration and mode

mode/ fieldwork duration	equal	unequal	total
equal	17	5	22
lower, < 90 days	66	33	88
higher, < 90 days	39	32	71
higher, ≥ 90 days	23	25	48
total	145	84	229

Cramér's V = 0.25

Crosstabulation of register information and mode

mode/ register information	equal	unequal	total
equal	143	60	203
unequal	2	24	26
total	145	84	229

Cramér's V = 0.41

Crosstabulation of Sodeur's index and mode

mode/ Sodeur's index	equal	unequal	total
equal	130	79	209
unequal	15	5	20
total	145	84	229

Cramér's V = -0.08

Crosstabulation register information and applying ISCED mapping

applying ISCED mapping/ register information	equal	unequal	total
equal	161	42	203
unequal	18	8	26
total	179	50	229

Cramér's V = 0.08

Crosstabulation register information and centralisation

centralisation/ register information	equal	unequal	total
equal	155	48	203
unequal	19	7	26
total	174	55	229

Cramér's V = 0.02

Crosstabulation centralisation and applying ISCED mapping

applying ISCED mapping/ centralisation	equal	unequal	total
equal	138	36	174
unequal	41	14	55
total	179	50	229

Cramér's V = 0.05

Crosstabulation of education categories and applying ISCED mapping

applying ISCED mapping/ education categories	equal	unequal	total
equal	15	2	17
similar	32	4	36
different	132	44	176
total	179	50	229

Cramér's V = 0.14

Crosstabulation of education categories and proxy-reporting

proxy-reporting/ education categories	equal	unequal	total
equal	8	9	17
similar	11	25	36
different	13	163	176
total	32	197	229

Cramér's V = 0.36

Crosstabulation of sampling design and fieldwork agency

fieldwork agency/ sampling design	equal	unequal	total
equal	57	118	175
unequal	6	48	54
total	63	166	229

Cramér's V = 0.20

Crosstabulation of sampling unit and fieldwork agency

fieldwork agency/ sampling unit	equal	unequal	total
equal	37	21	58
unequal	26	145	171
total	63	166	229

Cramér's V = 0.47

Crosstabulation of mandatory survey participation and fieldwork agency

fieldwork agency/ mandatory participation	equal	unequal	total
equal	39	96	135
unequal	24	70	94
total	63	166	229

Cramér's V = 0.04

Crosstabulation of fieldwork duration and fieldwork agency

fieldwork agency/ fieldwork duration	equal	unequal	total
equal	9	13	22
lower, < 90 days	10	78	88
higher, < 90 days	18	53	71
higher, ≥ 90 days	26	22	48
total	63	166	229

Cramér's V = 0.37

Crosstabulation of response rate and fieldwork agency

fieldwork agency/ response rate	equal	unequal	total
equal	10	5	15
higher, < 30%	17	5	22
lower, < 30%	27	58	85
lower, ≥ 30%	5	64	69
not available	4	34	38
total	63	166	229

Cramér’s V = 0.51

Crosstabulation of Sodeur’s index and fieldwork agency

fieldwork agency/ Sodeur's index	equal	unequal	total
equal	51	158	209
unequal	12	8	20
total	63	166	229

Cramér’s V = -0.23

Crosstabulation of mode and fieldwork agency

fieldwork agency/ mode	equal	unequal	total
equal	33	112	145
unequal	30	54	84
total	63	166	229

Cramér’s V = -0.14

Crosstabulation of register information and fieldwork agency

fieldwork agency/ register information	equal	unequal	total
equal	50	153	203
unequal	13	13	26
total	63	166	229

Cramér’s V = -0.18

Crosstabulation of centralisation and fieldwork agency

fieldwork agency/ centralisation	equal	unequal	total
equal	59	115	174
unequal	4	51	55
total	63	166	229

Cramér’s V = 0.25

Crosstabulation of applying ISCED mapping and fieldwork agency

fieldwork agency/ applying ISCED mapping	equal	unequal	total
equal	55	124	179
unequal	8	42	50
total	63	166	229

Cramér’s V = 0.14

Crosstabulation of education categories and fieldwork agency

fieldwork agency/ education categories	equal	unequal	total
equal	15	2	17
similar	15	20	36
different	32	144	176
total	63	166	229

Cramer's V = 0.44

Data sources:
EU-LFS 2008-2012, files from Eurostat, data file versions 2016, variable HATLEVEL, weighted using variable COEFF;
EU-SILC 2010, file from Eurostat, data file versions CROSS-2011-6, variable PE040, weighted using variable PB040;
AES 2011, file from Eurostat, data file version 1.0, variable HATLEVEL;
PIAAC 2011, file from OECD, data file version 2013, variable edcat7, analysed using complex weights with the International Database Analyzer (IDB); Eurobarometer 73.2 & 73.3 (2010), file from Eurostat, data file versions 2.0.1 of 2012, variable v362, data weighted to correct regional oversampling for Germany and the UK;
EQLS 2012, data file version 3 of 2014, variable Y11_ISCEDsimple, weighted using variable w1;
EWCS 2010, data file version of 2011, variable ef1_isce, weighted using variable w1;
ESS 2012, data file version 2.3 of 2016, variable edulvlb, weighted using variable dweight; for Greece and Croatia data from 2010 were used (data file versions: 3.3., variable edulvlb) and for Latvia and Romania data from 2008 were used (data file version 4.4, variable edulvla);
EVS 2008, data file version 4.0.0 of 2016, variable v336, data weighted to correct regional oversampling for Belgium, Germany, and the UK;
ISSP 2012, data file version 4.0.0 of 2016, variable DEGREE, data weighted to correct regional oversampling for Germany; for Italy data from 2011 were used (data file version 3.0.0, variable DEGREE)

Paper IV

What Determines Immigrants' German Language Proficiency? A Panel Analysis of Established and Recently Arrived Immigrants in Germany

5 What Determines Immigrants' German Language Proficiency? A Panel Analysis of Established and Recently Arrived Immigrants in Germany

In 2015 and 2016 many immigrants arrived in Germany and a central factor in their successful integration is acquiring German language skills. Previous research identified three mechanisms that strongly affect immigrants' second language proficiency, namely language exposure, efficiency, and incentives. When analysing the effect of these mechanisms, most studies apply a static approach and focus on one point in time. However, we know that the process of learning a language is not linear and thus the effects of the mechanisms are likely to vary over time. This study aims to consider the whole process of learning the German language. The paper analyses the effects of these mechanisms for two groups of immigrants; for established immigrants, who have been living in Germany for at least five years, and recently arrived immigrants. Moreover, through using panel data, this study also analyses the effects of intra-individual changes within the mechanisms. Another feature of this study is the innovative operationalisation of immigrants' homeland education, a key indicator of the mechanism indicating efficiency. The results of this study indicate that all three mechanisms positively affect immigrants' German language skills; this finding is in line with previous research. Concerning intra-individual changes, these strongly affect language exposure for recently arrived immigrants, but they affect the mechanism of incentives for the established immigrants. The paper ends by discussing the results and drawing a conclusion.

Keywords: immigrants, language proficiency, longitudinal perspective

5.1 Introduction

In 2015 and 2016 an unusually high number of people, predominantly from the Middle East and the Horn of Africa, immigrated to Germany and other European countries. Due to wars, political persecution and unrest, as well as forced labour and poverty in those countries, most immigrants will probably stay. This requires a renewed discussion on immigrants' integration in Germany. A central factor of immigrants' integration is acquiring German language skills. Speaking the language fluently is important when communicating with other people, for social integration when making contacts and establishing networks, including with natives, as well as for finding a job (Chiswick, 1991; Dustmann, 1994; Esser, 2006).

Previous research has developed and tested a model for the acquisition of immigrants' second language skills that considers three central mechanisms: language exposure, efficiency and incentives (Chiswick & Miller, 1995, 2001, 2007; Esser, 2006). Language exposure indicates the degree of confrontation with the language, which also implies opportunities for practising it. Efficiency refers to cognitive skills and ability for learning the language; and the third mechanism indicates the impact of incentives, which greatly depends on the expectation of a future in the destination country. When analysing the impact of these mechanisms, we must consider that the process of learning a language is not linear. In the beginning language skills increase rapidly and after some years the growth slows down and the achieved level of the language remains nearly static. (Chiswick & Miller, 2001; Esser, 2006; Stevens, 1999). Consequently, the effects of the mechanisms probably vary within this process, which is not reflected in the theoretical model. This paper fills this gap and adds the longitudinal perspective to this model.

So far, most empirical studies assessing the influence of these mechanisms on immigrants' second language proficiency apply a static approach that focuses on a single point in time (Kristen, 2019; van Tubergen & Kalmijn, 2009). Only single studies exist analysing pooled data to look at a longer time period (van Tubergen & Kalmijn, 2009) or panel data, which often focus on a short time interval in the early years after migration (Chiswick, Lee, & Miller, 2004; Hou & Beiser, 2006; Kristen, Mühlau, & Schacht, 2016). This study combines both empirical approaches and conducts an extensive analysis. I analyse two groups of first generation immigrants, which differ in their length of stay in Germany and therefore in their German language proficiency. This allows for reflecting on the development of immigrants' German language skills and the effect of the mechanisms in different stages, namely the crucial early years after migration and after some years in the country. Additionally, I analyse panel data for both groups of immigrants and thereby assess the effects on intra-individual changes of the mechanisms over time. Thus, I can consider nearly the whole process of learning the German language. In line with this, the study asks two research questions: How does the impact of the mechanisms that affect immigrants' German language proficiency change over time? Are the effects of the mechanisms the same for both groups of immigrants?

Section two provides the theoretical background of this paper. It gives a brief historical overview of migration to Germany and it discusses the central mechanisms determining immigrants' German language proficiency as well as their impact over time. In the third section, the data of the migration and refugee surveys of the German Socio-Economic Panel (SOEP), the analysis strategy, and the indicators related to the mechanisms and their operationalisation are described. A further asset of this study is that it considers indicators that are rarely included in previous studies, such as immigrants' proficiency of the mother tongue and their identification with Germany. Moreover, the indicator of immigrants' homeland education is operationalised, giving an innovative approach. In section four, the results are presented. The paper ends with a discussion of the results and the limitations of this study and draws a conclusion in section five.

5.2 Immigrants' Second Language Proficiency

5.2.1 Immigrants in Germany

This paper focuses on immigrants of the first generation, who immigrated to Germany and do not have a German passport. To better understand the relevance of this topic and to be aware of the different groups of immigrants living in Germany I briefly describe the largest waves of migration to Germany since the 1950s. The first group are migrant workers, so-called 'guest workers' (*Gastarbeiter*), who came seeking jobs. They migrated to Germany as a result of formal recruitment agreements with Italy (1955), Spain, Greece (both 1960), Turkey (1961), Morocco (1963), Portugal (1964), Tunisia (1965) and Yugoslavia (1968). Most guest workers came from Turkey and Yugoslavia (Kogan, 2007, 2011). By 1973 more than four million foreign-born people were living in Germany (Kalter & Granato, 2007; Rudolph, 1994). Due to the oil crisis in 1973 and the related economic stagnation, the migration of workers ended. However, migration did not stop, as many guest workers stayed and in the following years their families moved to Germany to be reunited. The next wave of migration was by ethnic Germans (*Aussiedler*), who were born abroad and returned to Germany (Kogan, 2007, 2011). In the 1980s most ethnic Germans came from Poland and Romania and in the 1990s from countries of the former Soviet Union. Overall, 2.5 million ethnic Germans returned between the mid-1980s and mid-1990s (Bundesverwaltungsamt, 2019). At the same time, roughly 350,000 people from countries of the former Yugoslavia fled to Germany due to war and ethnic cleanings (Lederer, 1997). Since the new law on the freedom of movement for EU citizens in 2005, it has been easier for people from other EU countries to settle down and work in Germany. Around 1.5 million people, mainly from Romania, Poland and Bulgaria, came to Germany (Bundesministerium des Innern für Bau und Heimat & Bundesamt für Migration und Flüchtlinge, 2020). Since the wars in Afghanistan and Iraq in the 2000s, and the 2011 war in Syria, as well as political unrest in other countries in the Middle East and the Horn of Africa, an increased number of people have been leaving their own countries. Since 2012, roughly two million migrants have arrived in Germany and applied for asylum. The peak was in 2015/ 2016 with 900,000 immigrants, since when the annual number has decreased (Bundesministerium des Innern für Bau und Heimat & Bundesamt für Migration und Flüchtlinge, 2020). Overall, as official data of 2018 from the Federal Office for Migration and Refugees (BAMF) show, around 8.4 million people (roughly 10%) living in Germany are immigrants of the first generation. The largest numbers of these first-generation immigrants are from Poland (13%), Turkey (10%), Russian Federation (8%), Kazakhstan (7%), Romania (6%), Syria (5%) and Italy (4%) (Bundesministerium des Innern für Bau und Heimat & Bundesamt für Migration und Flüchtlinge, 2020).

In this study, I focus on two groups of immigrants who differ greatly with regard to their length of stay in Germany and thus their German language proficiency. The first group are the so-called established immigrants, who migrated from countries of the former Yugoslavia or the Soviet Union or for family unification from other European countries. They mainly arrived in the 1990s, so they have been living in Germany for a considerable time (at least for five years), and therefore often have advanced German language skills. I can consider panel data covering an interval of five years (from 2013 to 2017) for these immigrants. Hence for this group, I analyse the long-term effects of the mechanisms on immigrants' German language proficiency. The second group covers recently arrived immigrants, who mainly come from the Middle East and the Horn of Africa. They migrated to Germany since 2013 and are living in Germany for less than five years, and therefore their German language acquisition is still in its early stages. For this group, I analyse panel data for two years (2016 and 2017). Looking at this group, I explain the improvement in immigrants' German skills shortly after their arrival and examine the short-term effects of the mechanisms.

5.2.2 Theory on Immigrants' Acquisition of Destination-Language Skills

Chiswick and Miller (2001, 2007) developed a 'standard model' that distinguishes three mechanisms of language acquisition for immigrants, namely language exposure, efficiency and incentives (Chiswick & Miller, 1995, 1998, 2001, 2007). The model is related to human capital theory (Becker, 1993). In general, a mechanism explains the observed regularities and specifies the causal process by which the outcome (here: immigrants' second language proficiency) is achieved (Hedström & Ylikoski, 2010). The mechanisms of language exposure, efficiency and incentives cannot be measured directly and therefore researchers analyse related indicators, which focus on important but delimited aspects of the mechanisms (Hedström & Ylikoski, 2010).

The three mechanisms, and in particular a large set of associated indicators affecting immigrants' second language skills, have been analysed in numerous studies for different destination countries, such as the US (Carliner, 2000; Chiswick & Miller, 1999; Espenshade & Fu, 1997), Canada (Chiswick & Miller, 2001; Hou & Beiser, 2006), Australia (Chiswick et al., 2004; Chiswick & Miller, 1995), Israel (Beenstock, Chiswick, & Repetto, 2001; Mesch, 2003), the UK (Dustmann & Fabbri, 2003), Norway (Hayfron, 2001), Belgium (van Tubergen & Wierenga, 2011), the Netherlands (van Tubergen, 2010) and Germany (Dustmann, 1994; Esser, 2006). These studies found almost equal empirical evidence of the effects of these mechanisms. Most studies analysed cross-sectional data, including census data, and focussed on a single point in time. However, as described, the process of acquiring a foreign language is not linear, and therefore the impact of the mechanisms varies over time. In the next sections, I will describe the mechanisms, their intra-individual changes, and indicate the differences in these mechanisms across the two groups of immigrants.

Language Exposure

The first mechanism affecting immigrants' second language proficiency is language exposure. This is defined as "the extent to which others, whether in person or through the media, use the destination language in one's presence and the extent to which the person himself or herself utilizes it" (Chiswick & Miller, 1995, p. 249). Empirical studies confirm that language exposure improves immigrants' second language skills, and that this mechanism is the major source determining their language proficiency (Chiswick & Miller, 2001, 2007; Esser, 2006; Kristen, 2019). Esser (2006) describes a relationship of increasing marginal returns between language exposure and the (practical) value of language. By practising the language, immigrants' proficiency will increase, but this investment will only pay off if it exceeds a certain threshold of language exposure that ensures a 'sustained use' (Esser, 2006, p. 86). Accordingly, premature break-offs or interruptions excessively reduce possible marginal returns. When studying language proficiency of first-generation immigrants, we can differentiate between indicators assessing language exposure before and after migration (Chiswick & Miller, 1995, 2001). Moreover, Chiswick and Miller (1995, 2001) suggest differentiating the effect of language exposure using indicators assessing the time unit (e.g. the years spent in the destination country), and the intensity (e.g. attendance at a language class, neighbourhood and family characteristics) (Chiswick & Miller, 2001).

I also consider the effects of intra-individual changes in language exposure over time. Negative changes can occur when the immigrant reduces, interrupts or completely breaks off exposure, through dropping out of a language class or moving into a neighbourhood in which the language of the destination country is rarely spoken. Thus, language skills do not further improve and often language skills already acquired diminish (Esser, 2006). Positive intra-individual changes are possible through extending language access, particularly when on a regular basis, for instance when attending a language class or establishing contacts with natives, such as in a sports club. This enhances immigrants' second language skills (van Tubergen, 2010).

Bearing in mind that the process of learning a foreign language is non-linear, the effect of language exposure and also of intra-individual changes will have a greater impact when they occur in the early stage of learning the language. Instead, when having achieved a high level of language proficiency, these effects decrease. Therefore, language exposure and any intra-individual changes are more crucial for recently arrived immigrants than for established ones.

From this, I establish the following hypotheses:

> H1: Language exposure has a positive effect on immigrants' German language skills.
>
> H2: Intra-individual changes in immigrants' language exposure affect their German language skills.
>
> H3: Language exposure and intra-individual changes in language exposure are more crucial for recently arrived immigrants than for established immigrants.

Efficiency

Efficiency is the second mechanism influencing the proficiency of second language skills. Empirical studies have found a positive effect of efficiency on immigrants' second language proficiency (Chiswick & Miller, 2001; Dustmann, 1994; Esser, 2006; Stevens, 1992; van Tubergen, 2010). Chiswick and Miller (2001) defined efficiency as "the extent of improvement in destination-language skills per unit of exposure" (p. 393). It refers to individuals' cognitive skills and ability, and indicates how easy or difficult it is for people to adapt a concept, learn a new one or deal with new grammatical structures and terminologies. Individual efficiency strongly depends on genetic factors as well as certain neurological and biological processes and also on age. The argument on age is that with increasing age at the time of migration, immigrants' ability to learn a foreign language reduces (Esser, 2006). Thus, immigrants who migrated at a higher age have to invest more time and effort to achieve the same level of the language as those who migrated at a younger age.

Individual efficiency in principle is quite stable and does not vary much over time (Esser, 2006) and so I will not formulate a hypothesis on the intra-individual changes related to efficiency. Regarding the two groups of immigrants, the effect of efficiency will be more crucial when starting to acquire language skills. Having achieved a sufficient level in the language of the destination country, the effect of efficiency will be smaller. Thus, I expect to find differences between the recently arrived and the established immigrants.

I derive the following hypotheses:

H4: Efficiency has a positive effect on immigrants' German language skills.

H5: Efficiency is more crucial for recently arrived immigrants than for established immigrants.

Incentives

Finally, immigrants' language proficiency also depends on incentives, which are partly determined by costs. Learning a foreign language is related to monetary cost, e.g. participation fees for language classes and related materials, as well as to opportunity costs, e.g. attendance at a language class reduces time available for work (Esser, 2006). Immigrants will bear these costs and try to become proficient in the language if they expect that their investment will pay off. This strongly depends on immigrants' expected length of stay, their future prospects in the destination country as well as their personal goals (Esser, 2006). Incentives for learning the language can be of an economic or non-economic nature. Economic incentives are particularly important for adult immigrants who want to work in the new country. Through acquiring language skills they increase their human capital as well as their opportunity for entering the labour market, achieving a higher occupational position and thus a higher income (Chiswick, 1991; Chiswick & Miller, 2001; Dustmann & Fabbri, 2003). In contrast, non-economic incentives reflect the motivation based on immigrants' social integration, their identification with the destination country and social contacts. These incentives are central for children or older immigrants, who will not (yet) look for work in the destination country (Esser, 2006; Kristen, 2019; van Tubergen & Mentjox, 2014). Overall, empirical studies indicate positive effects of economic and non-economic incentives (Chiswick & Miller, 2001; Kogan, 2016; van Tubergen & Mentjox, 2014).

I also expect intra-individual changes in the effect of incentives. Negative changes are likely to hamper immigrants' language proficiency, for instance when the immigrant is forced or decides to leave the country, or realises that the investment will not pay off because he/ she has not found a suitable job. In contrast, an upcoming chance of getting a job or the possibility of staying and developing long-term prospects in the destination country, provide positive changes in the incentives and will increase second language skills.

As with the other mechanisms, incentives are particularly important at the early stage of learning the language. They become less relevant when the immigrant has achieved a high level of language proficiency. This also applies to the effects of intra-individual changes. Therefore, incentives and intra-individual changes are expected to be more relevant for recently arrived immigrants than for established ones.

Concerning this mechanism, I develop the following hypotheses:

H6: Incentives have a positive effect on immigrants' German language skills.

H7: Intra-individual changes in immigrants' incentives affect their German language skills.

H8: Incentives and intra-individual changes in the incentives are more crucial for recently arrived immigrants than for established immigrants.

5.3 Data, Measures and Method

5.3.1 The SOEP Migration Surveys

This paper analyses the impact of mechanisms affecting immigrants' German language skills for two groups – established and recently arrived immigrants. For both groups, special surveys were conducted within the German Socio-Economic Panel (SOEP). The surveys interview immigrants on several biographical elements, their motivation and their route to Germany, their integration and on their attitudes and beliefs. Both surveys are household surveys, in which the immigrant and his/ her household are interviewed. Therefore, the samples also include children who have not migrated themselves or who migrated at a young age. Due to the focus of this study on first-generation immigrants, all respondents who were born in Germany are excluded. This also applies to immigrants who migrated to Germany below the age of ten because before that age learning a new language is easier (Esser, 2006). The SOEP surveys are repeated annually using computer-assisted face-to-face interviews. As this study also considers the effects of intra-individual changes, I also excluded respondents who participated in the survey only once.

The IAB-SOEP migration samples (M1 and M2) have been implemented through collaboration between the SOEP and the Institute for Employment Research (IAB) of the German Federal Employment Agency. This survey focuses on established immigrants, who arrived in Germany since 1994. It also includes roughly 1,000 immigrants who have been living in Germany for less than five years, whom I have excluded to make a clear separation between the two groups. The largest national groups of the established immigrants covered in this survey are from Russia (15.3%), Poland (12.1%), Kazakhstan (10.3%), Romania (8.2%) and Turkey (8.1%). The remaining 46% are immigrants from 87 other countries. The interviews are mostly conducted in German, but translation assistance is offered for English, Polish, Turkish, Romanian and Russian language speakers. These translations options are only used for around 5% of the respondents (Brücker et al., 2014; Kroh et al., 2016; Kühne & Kroh, 2017). The survey started in 2013 and since then nearly 2,700 households have been surveyed each year. I consider two to five observations for these immigrants between 2013 and 2017.

The second data source is the IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5), which is developed through cooperation between the SOEP, the IAB, and the Research Centre on Migration, Integration and Asylum of the Federal Office for Migration and Refugees (BAMF-FZ). This survey focuses on recently arrived immigrants who came to Germany since 2013. This survey also includes thirty immigrants who have been living in Germany for more than five years, and they have been excluded from the analysis. The final sample of recently arrived immigrants predominantly covers immigrants from Syria (57.5%) followed by those from Iraq (11.3%), Afghanistan (9.0%) and Eritrea (5.8%). The remaining 16% are from 45 other countries. Most respondents have only basic German language skills and therefore the questionnaire is offered in six further languages: Arabic, English, Farsi, Kurmanji, Pashtu and Urdu. Because most interviewers do not speak these languages, the interviews were conducted using audio files instructing the respondents and reading out the questions and answer categories, or with the support of third persons who helped with the translations. To a very small extent, professional interpreters were present during the interviews (Brücker, Rother, & Schupp, 2017; Brücker et al., 2016; Kroh et al., 2016). This survey was initiated in 2015, and in 2016 nearly 2,000 migrants were interviewed. We only have two observations for 2016 and 2017 per respondent that can be utilised in this study.

5.3.2 Measurement

Measuring Immigrants' German Language Skills

I examine immigrants' German language proficiency based on their self-reported speaking, reading and writing skills, which are measured annually. The items have a five-point answer scale ('very good' (1), 'good' (2), 'fair' (3), 'poor' (4) and 'not at all' (5)), which I recorded in reverse. The items are strongly correlated: Cronbach's α=.92 for the established immigrants and α=.94 for the recently arrived immigrants. Through a principal component analysis (PCA) the items are combined. The first dimension explains 85.7% of the variance for the established immigrants and 88.7% for the recently arrived immigrants, and has an eigenvalue above one for both groups. I use the extracted factor scores (z-standardised) of this dimension as the dependent variable.

Indicators for Language Exposure

As mentioned, I cannot directly measure the mechanisms affecting immigrants' German language proficiency and instead consider observable indicators. For each, I specify a bridge assumption indicating how it is related to the mechanism. Although some indicators can be related to more than one mechanism (Esser, 2006; van Tubergen, 2010), I assign most of them to only one for which I theoretically assume the closest link. Table 5.1 provides an overview of the mechanisms and their assigned indicators. For an indicator that varies over time, I also argue on the effect of intra-individual changes on immigrants' German language proficiency before describing the operationalisation of the indicator.

To assess the effect of language exposure, I consider the following indicators: length of stay, attending a German language class, having a German educational qualification, currently attending a formal educational programme in Germany, employment status, the household composition of the immigrant, especially if the immigrant is living together with a partner or children, as well as their contact with natives. All indicators focus on post-migration characteristics because data limitations prevent the inclusion of pre-migration indicators. I also do not differentiate between indicators assessing the time unit and the intensity of language exposure because empirically it is difficult to disentangle those (Kristen, 2019).

Length of stay is a key indicator for language exposure because while living in Germany, immigrants are regularly confronted with the German language and thereby learn and practise it. As mentioned, the effect of this indicator is not linear (Dustmann, 1994; Esser, 2006; Stevens, 1999). The length of stay variable at the first interview is time-invariant and derived from the variables for year of the first interview and year of migration to Germany. The related variable *survey year,* which increases with every year, instead indicates the time-varying effect since the first interview.

The indicator *attending a German language class* or integration courses reflects immigrants' systematic access to the language, which is beneficial for enhancing their German proficiency (Chiswick & Miller, 1995; van Tubergen, 2010). The effect of this indicator can change: while interruptions or break-offs hamper immigrants' proficiency of the German language, starting a language class is beneficial. The related variable indicates whether the immigrant has attended a German language class (1=yes, 0=no). Unfortunately, this indicator is only measured once for the established immigrants and thus is time-invariant, but it is time-varying for the recently arrived immigrants.

Table 5.1 The mechanisms of immigrants' second language proficiency and their related indicators

Indicators	**Language Exposure**	**Efficiency**	**Incentives**
length of stay*	+		
German language class*	+		
German education*	+		
currently attend education*	+		
employed*	+		
partner*	?		
children*	?		
social network*	+		
age at arrival		-	-
homeland education		+	
parents' education		+	
proficiency mother tongue		+	
region of origin		+	
health*		+	
settlement intention*			+
identification with Germany*			+
Gender			+

*time-varying indicators

Having a *German educational qualification* or *currently attending a formal educational* programme also indicates German language exposure. Most educational programmes are of some years' duration, and the majority of the lessons are given in the German language, thereby enabling immigrants to improve their language skills (Dustmann, 1997; Esser, 2006). However, the effects of these factors can also be reciprocal because attending an educational programme and having a German qualification requires some German language skills. Both indicators can vary over time. Through starting, interrupting, breaking off or finishing an educational programme, immigrants' language exposure changes and this affects their German language skills. For the indicator of having a German educational qualification, only positive changes are possible through successfully completing an educational programme. The related variable indicates whether the immigrant has a German qualification or currently attends an educational programme (1=yes, 0=no). Detailed differentiations by educational levels are not appropriate due to the low proportion of immigrants to whom these indicators apply. Moreover, for recently arrived immigrants I will not include the indicator of having a German qualification because hardly any had completed an education programme by then.

Through *being employed* immigrants come into contact and communicate with colleagues or customers and if those are natives, this will enhance their German language exposure (Chiswick, 1991; Chiswick & Miller, 1995; Dustmann & Fabbri, 2003; Esser, 2006). However, this indicator might also be reciprocal. The employment status can change over time when starting a job or becoming unemployed and this also determines immigrants' German language exposure. The variable of this indicator distinguishes whether the immigrant is employed (1=yes) or not (0=no).

Living together with a partner influences immigrants' language environment. Those who do not have a partner or are living alone in Germany are more likely go out and establish contacts, and thus increase their language exposure. In contrast, immigrants who are living together with their partner spend more time at home, where they communicate in their mother tongue or a third language, which hampers their learning German (Chiswick, Lee, & Miller, 2005; Chiswick & Miller, 2001; Stevens, 1992). There is one exception: if the immigrant's partner is a German native speaker, this will increase their language exposure (Chiswick & Miller, 1995). This indicator can change over time, for instance when an immigrant's partner arrives months later and then he/ she spends more time at home. Instead, breaking of a relationship or finding a German-speaking partner probably increases language exposure. The related variable indicates whether the immigrant is living with a partner in the same household (yes=1, no=0). It does not reflect whether the partner is a native because this applies to 17 established immigrants only.

The effect of *children* on their parents' language exposure is unclear. On the one hand, children can be door openers, who establish contact with other families to whom their parents need to speak in German. On the other hand, children can also become interpreters, who handle external communication, which reduces their parents' language exposure (Chiswick et al., 2005; Chiswick & Miller, 1995, 1999). An important intra-individual change happens through the birth of a child, when language exposure increases, through communication with doctors, nurses or nursery teachers. Instead, if a child leaves the household, it depends on his/ her role whether this positively or negatively affects parents' language exposure. Parents might then use their mother tongue more often or speak even more German to handle external communication. The related variable indicates whether a child is living in the same household (yes=1, no=0).

Through establishing a *social network* that includes natives, immigrants' language exposure increases (Drever & Hoffmeister, 2008; Haug, 2008; Kalter & Kogan, 2014). However, immigrants frequently establish those contacts only when they already have adequate language skills. This indicator also varies over time. Positive changes that increase language exposure occur when establishing a network, whereas negative changes, which reduce language exposure, occur through losing contacts or reducing the network. This indicator is operationalised differently for the two groups. For the established immigrants, I combine the items indicating whether in the last year the immigrant has visited a native or has been visited by a native at home. This final variable indicates immigrants' visits with natives (1=yes, 0=no) and is time-varying. For the recently arrived immigrants, I use the variable indicating how much time the immigrant spends with Germans: every day (1), several times per week (2), every week (3), every month (4), less often (5) and never (6). I reversed the scale and treat this variable as continuous. This item was only asked once and thus is time-invariant.

Indicators for Efficiency

To assess the impact of immigrants' efficiency I consider the following indicators: age at arrival, education, parents' educational level, proficiency in the mother tongue, region of origin and health status. Apart from health status, these indicators are pre-migration characteristics that do not vary over time.

Age at arrival reflects immigrants' cognitive ability to learn a new language, which decreases with age (Chiswick & Miller, 2001; Esser, 2006; Stevens, 1999). Moreover, it determines whether the immigrant will attend an educational programme in Germany, and researchers argue that age is the key factor for efficiency and not, as one would expect, educational attainment. The variable is derived from the immigrant's year of birth and year of migration to Germany.

Furthermore, immigrants' cognitive ability is important as this indicates the cognitive and conceptual skills that facilitate learning a new language. Usually, *education* is used as a proxy for measuring cognitive skills. In line with Dustman (1997) and van Tubergen (2010), this paper distinguishes between immigrants' German education indicating language exposure and immigrants' homeland education indicating efficiency. Measuring immigrants' *homeland education* is quite challenging as they come from different countries, with different educational systems, and the respective qualifications cannot be translated (Schneider, Joye, & Wolf, 2016; Smith, 1995).

In this paper, I apply an innovative approach of operationalising immigrants' homeland education to gain a high-quality and metric measure assessing the impact of cognitive ability. In the SOEP, homeland education is measured by different instruments. One instrument indicates immigrants' educational attainment and focuses on the qualitative aspect of education. It asks two questions – one on school education and another on post-school education, such as vocational training or university. From these I generate a variable that distinguishes ten categories: left school without graduating *and* no further training/ in-house/ other training (1), graduated from mandatory school *and* no further training/ in-house/ other training (2), left school without graduating/ graduated from mandatory school *and* extended apprenticeship at a company (3), left school without graduating/ graduated from mandatory school *and* vocational school (4), graduated from a higher-level secondary school *and* no further training/ in-house/ other training (5), graduated from a higher-level secondary school *and* extended apprenticeship at a company (6), graduated from a higher-level secondary school *and* vocational school (7), university/ college with a more practical orientation (8), university/ college with a more theoretical orientation (9), doctoral studies (10). Another instrument asks immigrants the number of years they spent in school, and thus indicates the quantitative aspect of education.

Although the categorical variable and the years of schooling variable are highly correlated (Spearman's rank correlation: ρ=.57 for the established and ρ=.71 for the recently arrived immigrants), they cover different aspects of education and each measure has different advantages and disadvantages.[11] To use the information of both variables and to generate a comprehensive education variable, I follow the idea of

[11] An advantage of the years of schooling variable is that this variable has a metric level of measurement and the related question in principal can be translated. The largest disadvantage of this instrument is that it is quite demanding for respondents to answer the question. The categorical variable measures respondents' qualification, which is often easier for respondents to remember than the time they spent in the education system. Instead using only descriptions of the different levels inspired by the German education system (as done in the SOEP) makes it difficult to answer the question for immigrants, who have foreign qualifications from different countries that also have different education systems. A larger discussion of the different education measures can be found in the in the additional analysis to this paper in section 6 and in the frame paper of this dissertation in section 1.4.1.

Schröder and Ganzeboom (2014) and combine the two measure to one education index.[12] Doing so, I have to consider the different levels of measurement (metric and categorical). Therefore, I apply a nonlinear principal component analysis using the CATPCA (Categorical Principal Component Analysis) programme (Linting, Meulman, Groenen, & van der Kooij, 2007; Meulman, van der Kooij, & Heiser, 2004). The first dimension, which has an eigenvalue above one[13], explains 75.95% of the variance for the established immigrants and 80.81% for the recently arrived immigrants. I use the z-standardised factor scores as an independent variable indicating immigrants' homeland education.

Immigrants' efficiency is also determined by their *parents' education, especially of the father*. Previous studies identified a direct effect of fathers' education, this indicates that not all factors related to families' cultural capital, social status and genetic influences are fully reflected when considering immigrants' education alone (Dustmann, 1997; van Tubergen, 2010). This is likely when immigrants have received only little or no education in their home countries, due to continuing wars. The effect of fathers' education might also be higher for women, since in many societies they lack the opportunity to go to school. Father's education is measured in less detail than the respondent's education, and the variable only distinguishes between low, medium/ high levels of education and don't know/ missing.

Through *being proficient in the mother tongue* immigrants have a better understanding of a linguistic system and this is helpful when learning another language. This indicator is rarely measured in surveys and thus not often considered in studies on immigrants' second language proficiency (Dustmann, 1994; van Tubergen & Wierenga, 2011). Theoretically, this indicator can also vary over time but it is measured only once and thus the indicator is time-invariant. The indicator is operationalised in the same way as the index on immigrants' German language proficiency.[14]

---

[12] I also estimated additional models that include the years of schooling variable and the categorical education variable (Models 3 (Table A5.6) for the established immigrants and Model 6 (Table A5.7) for the recently arrived immigrants in the Supplemental Material). Both models show statistically significant effects for both education variables.

[13] Although this nonlinear PCA is more specific than the explorative factor analysis, applying the Kaiser-Guttman criteria seems to be reasonable when deciding on the extraction of a dimension.

[14] Cronbach's α=.94 for the items on speaking, reading and writing skills of the mother tongue for the established immigrants and Cronbach's α=.77 for the recently arrived immigrants. The

Immigrants' *region of origin* is a proxy for the geographical, cultural and linguistic distance between the country of origin and Germany. People from a less distant region often have a lower linguistic distance to overcome, and therefore are more efficient in learning the German language (Esser, 2006). The indicator distinguishes between Europe (incl. Turkey and the Balkan countries), Russia (incl. all successor states of the former Soviet Union), the Arabic world, and other regions (incl. America, Latin America, Sub-Saharan Africa and East Asia). This distinction partly reflects the history of immigration to Germany.

Finally, I consider *immigrants' health status* because illnesses, whether physical or mental, hinder immigrants' learning of German (van Tubergen, 2010). This indicator is the only one of this mechanism that can vary over time. Positive changes will increase immigrants' efficiency because he/ she can better concentrate when learning German. In contrast, health problems reduce efficiency because the focus is shifted away from learning the language. The related variable distinguishes between five categories: very well (1), well (2), satisfactory (3), not very good (4), and poor (5). I reverse these categories and refer to this variable as metric in the analysis.

Indicators for Incentives

Most empirical studies contain few indicators assessing the effect of incentives on immigrants' second language proficiency. Although many indicators can theoretically be related to language exposure and to incentives, they are commonly assigned to exposure because it is difficult to disentangle the effects (Kristen, 2019). In this paper, I consider the following indicators: age at arrival, settlement intention, identification with Germany, and gender.

The indicator a*ge at arrival* has been described in the previous section. This also relates to the mechanism of incentives because for young adult immigrants in particular, finding a job, earning money and thus establishing a future life in Germany increases their motivation for learning German (Chiswick & Miller, 2001; Esser, 2006; Stevens, 1999). This effect will be smaller for immigrants who were older when they came to Germany.

---

first dimension of the PCA explains 90.0% of the variance for the established immigrants and 68.8% for the recently arrived immigrants and for both groups this dimension has an eigenvalue above one. Through calculating a PCA factor scores (z-standardised) are generated.

Settlement intention indicates higher incentives for immigrants who plan to stay in Germany and want to enter the labour market, earn money and establish contacts with natives (Chiswick & Miller, 1995; van Tubergen, 2010). However, immigrants' settlement intention strongly depends on their legal status. The indicator can change over time, for instance when the immigrant decides to leave Germany or the residency permission is declined, which would reduce the motivation for learning German. In contrast, deciding to stay, gaining permanent residency or having a job in prospect, positively affect immigrants' German skills. The related variable on immigrants' settlement intention simply distinguishes between yes (1) and no (0).

Identification with Germany is an incentive because immigrants who identify with Germany have a higher motivation for learning the language (Kristen, 2019). This factor also depends on residency permission, and immigrants' identification might change over time. Negative experiences, e.g. through discrimination, might hamper their motivation, whereas positive changes, such as getting permanent residency or finding a job, strengthen identification with Germany and increase the motivation to learn the language. This indicator is seldom considered in previous studies (Kristen, 2019; Kristen et al., 2016) because the respective question is rarely asked in surveys. The variable used here indicates how much the immigrant feels German, and distinguishes between completely (1), mostly (2), in some respects (3), barely (4) and not at all (5). The categories are recoded in reverse and the variable is treated as metric. This indicator, unfortunately, is only available for the established immigrants.

Lastly, *gender* is related to the mechanism of incentive because men often have a greater orientation to the labour market than women do, so they need to acquire German language skills in order to find a job and support the family (Dustmann, 1997; van Tubergen, 2010). Women, especially while they have small children, often stay away from the labour market to take care of them. The gender variable separates men (0) and women (1). An overview of the descriptive statistics of all indicators can be found in Tables A5.4 and A5.5 in the Supplemental Material.

5.3.3 Analysis Strategy

In the first part of the analysis, I describe the improvement of immigrants' German language proficiency together for both groups. For this analysis, I include all respondents of the both SOEP surveys, independent whether they fulfil the criteria of belonging the group of established or recently arrived immigrants. Here I also included immigrants, who participated only once in the survey, thus the case number for this analysis is quite large (number of respondents=10,071; number of observations=22,482).

In the second part, I strongly stick to the definitions for the established and recently arrived immigrants (see section 5.3.1) and I only included immigrant, who at least participated twice in the survey. Thus, the case number is lower for this analysis (2,049 established immigrants and 2,183 recently arrived immigrants). I run separate regression models for the two groups of immigrants to identify which mechanisms influence their German language skills at the different stages of the learning process. I also expect a large interaction of nearly all indicators with immigrants' length of stay, which can be reflected more easily by estimating separate models. Moreover, calculating different models allows for considering slightly different variables for the two groups.

I estimate the impact of the mechanisms as well as their changes over time on immigrants' German language skills through calculating a hybrid panel regression model (Allison, 2009; Firebaugh, Warner, & Massoglia, 2014; Halaby, 2004). This extends the standard regression model by additionally considering time-varying variables. Formally, the estimated model is: $y_{it} = (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\beta + \overline{\boldsymbol{x}}_i\gamma + \boldsymbol{z}'_i\boldsymbol{\delta} + \alpha_i + \varepsilon_{it}$ where the subscript i refers to the respondent and t to the time. The first term of this formula $(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\beta + \overline{\boldsymbol{x}}_i\gamma$ refers to time-varying variables, such as currently attending a German educational programme or health status. The effect of these variables is again split into the effect across respondents (between effect) focusing on using the mean of the respondents ($\overline{\boldsymbol{x}}_i\gamma$), and the deviation from this person-specific mean across time $\left((\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\beta\right)$ (within-effect). The second term $\boldsymbol{z}'_i\boldsymbol{\delta}$ indicates the effect of the time-invariant variables, such as age at arrival or homeland education. The formula includes two error terms, ε_{it} varies across respondents and time and α_i indicates the time-invariant individual error (Allison, 2009; Brüderl, 2010).

5.4 Results

5.4.1 The Development of Immigrants' German Language Proficiency

The first analysis step looks at the improvement in immigrants' German language skills by their length of stay in Germany. The horizontal axis of Figure 5.1 shows the number of years since the immigrant arrived in Germany and the vertical axis indicates the scores of their German language proficiency. The graph shows a rapid increase in immigrants' language skills in their first years after arrival. After five years of being in Germany immigrants achieve the average level of immigrants' German language proficiency. After eight years, immigrants' increase of Germany language skills slows down, meaning the achieved level of the German language remains quite constant. Further improvements become much smaller and often take much longer, so are less obvious in this graph. The shape of this curve is in line with our expectation.

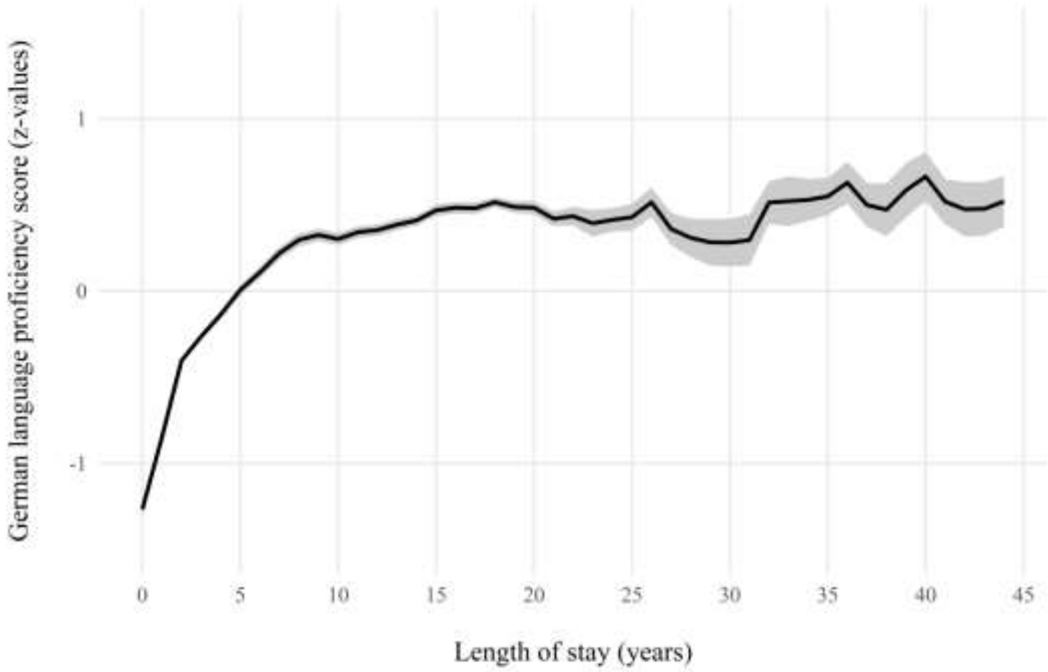


Figure 5.1 ***The development of immigrants' German language proficiency over time***

Note:
The graph only includes observations up to a length of stay of 44 years, because from 45 years onward there are fewer than 20 observations per year, leading to a very large standard error and thus a misleading graph.
Data sources:
IAB-SOEP migration sample (M1-M2) 2013-2017 and IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34

5.4.2 The Effect of the Mechanisms on Immigrants' German Language Skills

Before looking at the effects of the single indicators we look at the models' performance. Table 5.2 indicates the number of cases included in each model and the explained variance. Model 1 captures 2,049 established immigrants for which we can analyse 7,515 observations (455 respondents were interviewed two times, 390 three times, 305 four times and 843 five times). Model 2 covers 2,108 recently arrived immigrants and for each of them, we have two observations (N=4,216).

The overall R^2 in Model 1 is forty percent, indicating the explained variance of immigrants' German language proficiency through all included indicators, independent if they are time-invariant or vary over time. In Model 2, the overall R^2 is slightly larger with 44.5%. The between R^2 shows the explained variance of the time-invariant indicators and the between estimates of the time-varying variables. For the established immigrants the between R^2 is 46.1% and for the recently arrived immigrants it is 50.1%, which indicates that for both groups the largest part of the variance can be explained by the differences between the immigrants. The within R^2 shows the variance that can be explained by the within estimates of the time-varying indicators (intra-individual changes). For the established immigrants it is one percent and for the recently arrived immigrants, it is 28.3%. The substantial proportion of the within-variance for the recently arrived immigrants indicates that their German language proficiency varies greatly over time making it is worth considering intra-individual changes.

Table 5.2 Explained variance of the hybrid regression analyses

	Model 1 established immigrants	**Model 2 recently arrived immigrants**
number of cases	2,049	2,108
number of observations	7,515	4,216
R^2 overall	40.06	44.53
R^2 between	46.05	50.05
R^2 within	0.99	28.26

Data sources:
IAB-SOEP migration sample (M1-M2) 2013-2017 and IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34

Now I look at the impact of the mechanisms and their related indicators for the established immigrants[15] (Model 1 in Table 5.3). Starting with the mechanism of language exposure, we see that in line with our expectations, most indicators positively affect immigrants' German language proficiency. The effects are also statistically significant. Having a German educational qualification has the strongest effect size ($b=1.01$, $p<.001$) and enhances immigrants' German language skills by one standard deviation. Visiting with Germans also has a strong effect ($b=0.48$, $p<.001$) but all other indicators have smaller effects. In contrast, living together with a partner and/ or children hampers immigrants' German language skills, but only the effect of children is statistically significant ($b=-0.01$, $p<.01$). Overall, I do not reject hypothesis H1 indicating that language exposure has a positive effect on German language proficiency. Regarding the effects over time, I firstly look at the effects of the survey year. Compared to 2017, immigrants in 2013 have statistically significant better German skills ($b=0.05$, $p<.05$); thus over time their skills decreased, which is quite surprising[16]. The direction of the regression coefficients of the dummy for the years 2014, 2015 and 2016 differ but the effects are not statistically significant. Looking at the effects of the intra-individual changes of the other indicators, these are considerably smaller than the effect of the indicator itself, which reflects the differences between the immigrants. Of these, only two variables are statistically significant, the indicators on having a German educational qualification ($b=-0.07$, $p<.01$) and on currently attend an educational programme ($b=-0.10$, $p<.01$). The effects of currently attending an educational programme shows that changes, for instance through dropping out or starting an education programme, hamper immigrants' German language skills. From the descriptives we see that the number of immigrants who drop out of an educational programme is larger than the number of immigrants who started an educational programme. This might explain the negative effect. I reject hypothesis H2 because the intra-individual changes of the other indicators do not significantly affect immigrants' German language skills.

[15] I also did some robustness checks and calculate a model including migrants who are in Germany for at least eight years, and the effects are similar. Please see Model 5 in Table A5.6 in the Supplemental Material.

[16] This might be caused by the self-reporting of language skills, which more likely is prone to errors. This will be reflected in the discussion in section 5.5.

Table 5.3 Results of the hybrid regression analyses predicting immigrants' German language proficiency for the established and recently arrived immigrants

	Model 1 established immigrants				Model 2 recently arrived immigrants			
	b		SE	p	b		SE	p
Language exposure								
length of stay at first interview	0.01	**	0.00	.002	0.16	***	0.02	<.001
survey year (ref: 2017)								
2013	0.05	*	0.02	.019				
2014	-0.02		0.02	.309				
2015	0.01		0.02	.787				
2016	-0.01		0.02	.469	-0.48	***	0.02	<.001
German class	0.14	***	0.03	<.001	0.79	***	0.04	<.001
German language class (time-varying)					0.26	***	0.04	<.001
German education	1.01	***	0.11	<.001				
German education (time-varying)	-0.07	*	0.03	.034				
current education	0.26	*	0.11	.020	0.55	***	0.08	<.001
current education (time-varying)	-0.10	*	0.04	.011	0.07		0.06	.257
employed	0.23	***	0.04	<.001	0.27	***	0.05	<.001
employed (time-varying)	0.02		0.02	.343	0.07		0.04	.077
partner	-0.08		0.05	.098	-0.08	*	0.04	.059
partner (time-varying)	-0.02		0.04	.603	0.22	**	0.07	.001
children	-0.10	*	0.04	.008	-0.04		0.04	.315
children (time-varying)	-0.01		0.04	.838	-0.07		0.12	.533
visit with Germans at home	0.48	***	0.05	<.001				
visit with Germans at home (time-varying)	0.05		0.03	.102				
spend time with Germans					0.08	***	0.01	<.001
Efficiency								
age at arrival	-0.02	***	0.00	<.001	-0.02	***	0.00	<.001
homeland education	0.28	***	0.02	<.001	0.24	***	0.02	<.001
fathers' education (ref: low)								
medium, high	0.11	*	0.04	.004	0.07	*	0.03	.023
no answer	0.07		0.07	.293	0.01		0.04	.732
mother tongue	0.06	***	0.02	<.001	0.08	***	0.01	<.001
region of origin (ref: Europe)								
Russia	0.06		0.04	.095	-0.19		0.13	.127
Arabic world	-0.08		0.06	.208	-0.07		0.11	.496
other	-0.23	**	0.07	.001	-0.40	***	0.11	<.001
health	0.05	*	0.02	.009	0.07	***	0.02	<.001
health (time-varying)	0.01		0.01	.174	0.00		0.02	.859

Incentives								
settlement intention	-0.12	*	0.05	.017	0.03		0.06	0.60
settlement intention (time-varying)	0.01		0.03	.626	0.04		0.05	0.42
identification with Germany	0.25	***	0.02	<.001				
identification with Germany (time-varying)	0.04	***	0.01	<.001				
gender	0.14	***	0.03	<.001	-0.10	**	0.03	.002
constant	-1.05	***	0.15	<.001	-0.55	***	0.16	<.001

Data sources:
IAB-SOEP migration sample (M1-M2) 2013-2017 and IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34

The effects of the indicators related to efficiency in Model 1 are also in line with our expectations. Of these indicators, immigrants' homeland education has the largest impact[17] (b=0.28, p<0.001). For region of origin we recognise that learning German is significantly more difficult for immigrants from 'other' regions (b=-0.23, p<.01) compared to immigrants from European countries. For immigrants from the Arabic world or Russia there is no statistically significant effect on their German language proficiency. However, compared to the indicators of language exposure, the effects sizes of the indicators related to efficiency are rather small. Nevertheless, the effects of most indicators are statistically significant and I cannot reject hypothesis H4 stating that efficiency positively affects immigrants' German language skills. This mechanism contains only one time-varying variable, namely health status, which has a small positive effect but it is not statistically significant.

Studying the indicators related to incentives, all of them significantly influence immigrants' German language proficiency. The negative effect of age at arrival (b=-0.02, p<.001) and the positive effect of immigrants' identification with Germany (b=0.25, p<.001) follow the expectations. In contrast, the gender-effect (b=0.14, p<.001), indicating that women have higher German skills than men, contradicts our expectation. This also applies to immigrants' settlement intention (b=-0.12, p<.05), which hampers acquiring the German language. Although the results are mixed, the strong effect of identification with Germany leads us not to reject hypothesis H6 stating that incentives have a positive effect on immigrants' German language skills. Looking

[17] I also calculate a model using a simpler education measure that only uses the information of the categorical education variable and therefore contains more cases. The effects are quite similar to the described ones. Please see Model 4 in Table A5.6 for the established immigrants and Model 7 in Table A5.7 for the recently arrived immigrants in the Supplemental Material.

at the time-varying indicators, we see that the effect sizes of these variables are again smaller compared to the effect of the indicator itself. Only the effect of intra-individual changes in immigrants' identification with Germany is statistically significant (p<.001), which shows that this increases German language proficiency. From the descriptives we see that immigrants' identifaction with Germany slightly grows between 2013 and 2017 and this has a positive effect on their German language skills. Thus, I do not reject hypothesis H7 stating that intra-individual changes in immigrants' incentives change their German language skills.

To sum up, for the established immigrants most indicators related to language exposure, individual efficiency and incentives positively affect German language proficiency, and therefore I do not reject hypotheses H1, H4 and H6. I find mixed results for the time-varying variables and finally reject H2 on the impact of intra-individual changes in language exposure because most time-varying factors do not have a statistically significant effect. In contrast, I do not reject hypothesis H7 on the impact of intra-individual changes in incentives because of the highly significant effect of the intra-individual changes of immigrants' identification with Germany.

Now we look at the results for the recently arrived immigrants (Model 2 in Table 5.3). For the indicators related to language exposure the effects and the statistical significances are quite similar to the ones described for the established immigrants (in Model 1). In contrast to Model 1, the indicator of attending a German language class (b=0.79) has the largest impact, whereas the effect of spending time with Germans is rather small (b=0.08). Both effects are highly statistically significant (p<.001). In line with Model 1, German education strongly improves immigrants' German language skills. For the recently arrived immigrants, I cannot consider the indicator having a German qualification, and instead focus on the positive effect of currently attending an educational programme (b=0.55), which is also highly statistically significant (p<.001). As in Model 1, living together with a partner and/ or children negatively affects German language skills, but for recently arrived immigrants only the effect for the partner variable is statistically significant (p<0.05). Overall, also for the recently arrived immigrants, I do not reject hypothesis H1.

Looking at the effects over time, I firstly look at the variable indicating the survey year. It show that immigrants' German language skills increases significantly (p<.001) between 2016 and 2017 by a half standard deviation (b=0.48). This effect does not occur for the established immigrants because of their longer stay in Germany at the time of the first interview. Apart from the partner and children variables, the effect sizes for the variables on the intra-individual changes of the indicators are smaller than the effects of the indicator itself. The effect of changes in the indicator living together with the partner is statistically significant ($p<0.001$) showing that this improves immigrants’ German language skills. From the descriptives we see that in 2017 the number of immigrants living together with their partner increases by 150, probably because their family arrived later partly due to reunification progammes. In contrast to Model 1, the time-varying variable for currently attending an educational programme is not statistically significant for recently arrived immigrants. In Model 2, attending a German language class varies over time and this also significantly improves German language skills ($b=0.26$, $p<0.001$). From the descriptives we know that roughly half of the 930 immigrants who did not attended a German language class in 2016 do so in 2017 and this positive change is reflected in the model. Overall, the effects of intra-individual changes are mixed and thus, in contrast to the established immigrants, I cannot reject hypotheses H2 for the recently arrived immigrants. Comparing the results of the two groups, the indicators show larger effects for the recently arrived immigrants, and also the effects of intra-individual changes are significantly more important. Thus, language exposure is more crucial for the recently arrived immigrants than for the established immigrants and I cannot reject hypothesis H3.

Concerning the mechanism of efficiency, the effects of the indicators match our expectations. The directions of the effects and their statistical significances are similar to the results for the established immigrants. The only variable that differs between the models is the dummy variable of Russia as region of origin. For the recently arrived immigrants, coming from Russia hampers their German language proficiency whereas for established immigrants this facilitates learning German. This is not surprising because the latter group includes ethnic Germans, who might have learned some German from their relatives. However, for both groups, this dummy variable is not statistically significant. Overall, I do not reject hypothesis H4 indicating that efficiency has a positive effect on immigrants’ German language skills. Because of the large similarities in the results for both groups of immigrants, I reject hypothesis H5, which indicates that these groups might differ in their efficiency in acquiring German language skills.

Analysing the impact of incentives for recently arrived immigrants, the effect of age at arrival is negative and statistically significant ($b=-0.02$, $p<0.001$), which is in line with the expectation. The gender-effect is also statistically significant ($b=-0.10$, $p<0.01$) but the direction of the effect is different from that in Model 1. For the recently arrived immigrants, men have better German language skills than women, mainly because they are more oriented to the labour market; this follows our expectation. In contrast to Model 1, the effect of settlement intention is positive ($b=0.03$) but not statistically significant. The indicator of identification with Germany is not available for recently arrived immigrants. Overall, I do not reject hypothesis H6 indicating that incentives positively affect immigrants' German language skills. Assessing the effects of intra-individual changes, we only consider the time-varying variable settlement intention. The descriptives show that the number of immigrants telling that they want to stay in German increases between 2016 and 2017 but the effect is not statistically significant. Thus, I reject hypothesis H7 indicating that intra-individual changes in the incentives have a positive effect on German skills. Comparing the results for both groups, I recognise that according to this analysis, incentives are a stronger mechanism for the established immigrants than for the recently arrived immigrants, and thus I reject hypothesis H8.

To sum up, I find that the effects of most indicators are in line with our expectations and overall I found positive effects of language exposure, efficiency and incentives on immigrants' German language skills. Thus, for both groups of immigrants, I do not reject hypotheses H1, H4 and H6. For the time-varying indicators, the effects differ between groups. For the established immigrants, intra-individual changes in language exposure do not affect their German language skills, so I reject hypothesis H2, but changes in the incentives do have an effect and thus I do not reject hypothesis H7. For the recently arrived immigrants, it is vice versa. Comparing the two groups, the mechanism of language exposure is more important for the recently arrived than for established immigrants, which is in line with hypothesis H3. For the mechanism of incentives, it is the other way round and I reject hypothesis H8. For the mechanism of efficiency, I did not find substantial differences across the two groups and thus I also reject hypothesis H5.

5.5 Conclusion and Discussion

This paper analysed the impact of the well-studied mechanisms of language exposure, efficiency and incentives on immigrants' German language proficiency. The important benefit of this study is that, in contrast to most previous studies, it reflects the non-linearity of the learning process. Therefore, I extended the 'standard theoretical model' of the acquisition of immigrants' second language skills. In the analysis, I examined two groups of immigrants, which differ in the time they have spent in Germany and in the stage of acquiring German language skills. Moreover, through analysing panel data I also estimate the effects of intra-individual changes on the mechanisms. Thereby I consider nearly the whole process of learning the German language when analysing the impact of the three mechanisms. In line with previous research, I find that all three mechanisms enhance immigrants' German language proficiency. Two mechanisms, language exposure and incentives can vary over time whereas efficiency is time-invariant. Looking at the effects over time, I observe meaningful changes in the language exposure for the recently arrived immigrants and in the incentives for the established immigrants. Overall, the results indicate that the effect of language exposure and intra-individual changes in language exposure are central for German language proficiency for recently arrived immigrants. In contrast, for the established immigrants the mechanism of incentives, in particular changes in the identification with Germany, improves their German language skills. Concerning the factor of identification with Germany, we have to keep in mind that this is not measured for the recently arrived immigrants. Thus the mechanism of incentives captures different factors across the groups. Overall, this mechanism has the lowest number of indicators, and although many of the indicators of the mechanism of language exposure can partly also be related to incentives, as mentioned, this is difficult to disentangle (Kristen, 2019).

The results of this study indicate that for improving immigrants' German language skills and thus for strengthening their integration, immigrants themselves as well as politicians and the whole society are called to contribute. Offering language exposure especially through German language classes is central in the first years after migration. These classes seem to be quite successful and provide the basic German language skills that lay the foundation and prepare immigrants to enter the labour market or attend another educational programme. In the later stage, immigrants' identification with Germany and their contact with natives are essential. Therefore, creating opportunities where immigrants and natives come into contact with each other, learn more about the other culture and have new experiences are important. Thereby, hopefully, prejudices from both sides can be reduced.

I also want to emphasise a secondary finding on the effect of the region of origin. For both groups, I find a statistically significant negative effect for immigrants coming from 'other regions', including America, Latin American, Sub-Saharan African and Southeast and East Asian countries. Immigrants from these regions have lower German language skills than immigrants from Europe, Russia or the Arabic world. This finding indicates that Germany has good offers for the main groups of immigrants living in or entering the country, but fewer or inappropriate offers for those from other countries. To increase their German language skills, we could take better care of this group and create additional provision, such as a native mentor or bringing them together with Germans more deliberately. Another possible explanation of this effect is that these immigrants, especially those of America and Southeast and East Asian countries feel less pressure to learn German. They have proficient English skills that allow them to enter the labour market as highly skilled professionals and they can handle everyday communication often also in English.

An important methodological asset of this study is the innovative operationalisation of immigrants' homeland education. The education index considers two different aspects of immigrants' education – their educational attainment and their years spent in school. Due to the disadvantages of both measures and their different foci, combining them to generate a powerful indicator was desirable and feasible. The results confirm this because the effect of this education index is in line with previous studies (Chiswick & Miller, 2001; Dustman, 1997; Esser, 2006; van Tubergen, 2010), which also identify a positive and highly significant effect of immigrants' homeland education on their second language proficiency.

This study also faces some limitations. The main limitation is related to the sample composition of the two SOEP surveys. The separation of established and recently arrived immigrants follows the conceptualisation of the SOEP samples;

nevertheless I lost about 1,000 respondents[18] who had been living for less than five years in Germany but who are part of the survey of the established immigrants. For the analysis, it would have been good to add them to the recently arrived immigrants. This is possible but it comes with a major shortcoming, in that the SOEP uses different questionnaires for the two groups. Thus, the indicators are operationalised differently (e.g. the indicator of the social network), or the indicator does not exist for one of the groups (e.g. the indicator on identification with Germany), or the indicator is time-invariant for one group but time-varying for the other, as with attending a German language class.

Another limitation of this study concerns the measurement of immigrants' German language, being based on self-reports. This increases the chance of misreporting, which can lead to measurement errors. Misreports can be time-consistent through continuous under- or over-estimation of language proficiency or they may vary over time. The latter, for instance, happens when shortly after their arrival, immigrants overestimate their skills and later realise that they actually have a lower level. Both kinds of reporting errors also correlate with immigrants' cognitive ability and thus they likely also systematically over- or underestimate their language skills by their education (Edele, Seuring, Kristen, & Stanat, 2015). Misreports, of course, can also contain random errors (Dustmann & van Soest, 2001).

We also have to consider the differences in the survey languages. For the established immigrants the survey is in German and only some translation assistance is offered through printed booklets of the questionnaire. These are only available for a few languages and not systematically offered. Although most immigrants of this group have a sufficient or high level of German, we know from other studies that responding to a survey in a language that is not the mother tongue is often more challenging and this likely also reduces data quality (Kleiner, Lipps, & Ferrez, 2015; Wenz, Al Baghal, & Gaia, 2020 online first). In contrast, the questionnaire for the survey on the recently arrived immigrants, who often have no or few German language skills, is systematically translated into seven languages. Thus, most of these immigrants can answer the questionnaire in their mother tongue, which can positively affect data quality.

[18] I also calculate a model including these immigrants with the recently arrived immigrants and only include variables that are similar in both surveys, and it shows a similar pattern (see Model 8 in Table A5.7 in the Supplemental Material)

As mentioned, the major advantage of this study is its longitudinal perspective. However, the time series covered by the SOEP surveys are short, in particular for adequately assessing the effects of intra-individual changes, which usually happen slowly. Therefore, it would be beneficial to extend this analysis in a few years and include further survey waves. For the recently arrived immigrants, this would allow us to identify more precisely when the increase in German language skills slows down. We again could compare the development for both groups and see if the effects of the mechanisms and the changes over time are the same then or if there are still differences between these groups.

Another idea for further research is to perform similar analyses for immigrants living in other countries. In the context of the so-called 'European refugee crisis' in 2015, other European countries such as Austria, France, Italy and Sweden accepted a large number of immigrants. It would be useful to see if the results, especially for the recently arrived immigrants, are similar across the destination countries. Additionally, we could then include contextual or country characteristics, such as the attitude towards immigrants in the population, the labour market situation, and different regulations for immigrants, e.g. concerning family reunification, status of residency and access to labour. Thereby we could also identify the effects of different integration politics across countries. Such research might be relevant for better assessing the integration of recently arrived and also of established immigrants across countries.

5.6 Acknowledgments

I especially thank Silke Schneider, Christof Wolf, Matthias Sand, Ranjit Singh, Joss Roßmann, Michael Braun, Katharina Meitinger, Patricia Hadler and Roberto Briceno-Rosas for their helpful inputs and suggestions on this study as well as their feedback on earlier versions of this paper. I also want to thank the colleagues at the DIW Berlin, who provide the SOEP data, for their support, in particular Elisabeth Liebau, Lisa Pagel, Jana Nebelin and Florian Griese. I also thank Jane Roberts for proofreading.

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 programme under grant agreement No 654221.

5.7 References

Allison, P. D. (2009). *Fixed Effects Regression Models. SAGE Series: Quantitative Applications in the Social Sciences No. 160.* Thousand Oaks: SAGE Publications.

Becker, G. S. (1993). *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education.* (3rd ed.). Chicago: The University of Chicago Press.

Beenstock, M., Chiswick, B. R., & Repetto, G. L. (2001). The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel. *International Migration*, *39*(3), 33–60. doi: 10.1111/1468-2435.00155

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 216: Series C. Berlin: DIW/SOEP. Retrieved from: http://doku.iab.de/fdz/iab_soep/surveypaper_en.pdf (accessed April 2020)

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.* Politikberatung kompakt Nr. 123. Berlin: DIW/SOEP. Retrieved from: https://www.diw.de/documents/publikationen/73/diw_01.c.563710.de/diwkompakt_2017-123.pdf (accessed April 2020)

Brücker, H., Rother, N., Schupp, J., von Gostomski, C. B., Böhm, A., Fendel, T., … Vallizadeh, E. (2016). *Flucht, Ankunft in Deutschland und erste Schritte der Integration.* DIW Wochenbericht Nr. 46. Berlin: DIW/SOEP. Retrieved from: https://www.diw.de/documents/publikationen/73/diw_01.c.546854.de/16-46-4.pdf (accessed April 2020)

Brüderl, J. (2010). Kausalanalyse mit Paneldaten. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 963–994). Wiesbaden: VS Verlag für Sozialwissenschaften, Springer Fachmedien.

Bundesministerium des Innern für Bau und Heimat, & Bundesamt für Migration und Flüchtlinge. (2020). *Migrationsbericht 2018*. Retrieved from: https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Migrationsberichte/migrationsbericht-2018.html?nn=403964 (accessed April 2020)

Bundesverwaltungsamt. (2019). *Aussiedler. Spätaussiedler und ihre Angehörigen. Zeitreihe 1950-2018*. Retrieved from: https://www.bva.bund.de/SharedDocs/Downloads/DE/Buerger/Migration-Integration/Spaetaussiedler/Statistik/Zeitreihe_1950_2018.pdf?__blob=publicationFile&v=2 (accessed April 2020)

Carliner, G. (2000). The Language Ability of U.S. Immigrants : Assimilation and Cohort Effects. *International Migration Review*, *34*(1), 158–182. doi:10.1177/019791830003400107

Chiswick, B. R. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. *Journal of Labor Economics*, *9*(2), 149–170. doi:10.1086/298263

Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2004). Immigrant's Language Skills: The Australian Experience in a Longitudinal Survey. *International Migration Review*, *38*(2), 611–654. doi:10.1111/j.1747-7379.2004.tb00211.x

Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2005). Family Matters: The Role of the Family in Immigrants' Destination Language Acquisition. *Journal of Population Economics*, *18*(4), 631–647. doi: 10.1007/s00148-005-0008-0

Chiswick, B. R., & Miller, P. W. (1995). The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics*, *13*(2), 246–288. doi:10.1086/298374

Chiswick, B. R., & Miller, P. W. (1998). Language Skill Definition : A Study of Legalized Aliens. *International Migration Review*, *32*(4), 877–900. doi:10.1177/019791839803200402

Chiswick, B. R., & Miller, P. W. (1999). Language Skills and Earnings among Legalized Aliens. *Journal of Population Economics*, *12*, 63–89. doi:10.1007/s001480050091

Chiswick, B. R., & Miller, P. W. (2001). A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada. *Demography*, *38*(3), 391–409. doi:10.2307/3088354

Chiswick, B. R., & Miller, P. W. (2007). *The Economics of Language. International Analyses*. Oxfordshire: Routledge.

Drever, A. I., & Hoffmeister, O. (2008). Immigrants and Social Networks in a Job-Scarce Environment: The Case of Germany. *International Migration Review*, *42*(2), 425–448. doi:10.1111/j.1747-7379.2008.00130.x

Dustmann, C. (1994). Speaking Fluency, Writing Fluency and Earnings of Migrants. *Journal of Population Economics*, *7*, 133–156. doi:10.1007/BF00173616

Dustmann, C. (1997). The Effects of Education, Parental Background and Ethnic Concentration on Language. *The Quarterly Review of Economics and Finances*, *37*(1), 245–262. doi:10.1016/S1062-9769(97)90068-5

Dustmann, C., & Fabbri, F. (2003). Language Proficiency and Labour Market Performance of Immigrants in the UK. *The Economic Journal*, *113*(489), 695–717. doi:10.1111/1468-0297.t01-1-00151

Dustmann, C., & van Soest, A. (2001). Language Fluency and Earnings: Estimations with Misclassified Language Indicators. *The Review of Economics and Statistics*, *83*(4), 663–674.

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why Bother with Testing? The Validity of Immigrants' Self-Assessed Language Proficiency. *Social Science Research*, *52*, 99–123. doi:10.1016/j.ssresearch.2014.12.017

Espenshade, T. J., & Fu, H. (1997). An Analysis of English-Language Proficiency among U.S. Immigrants. *American Sociological Review*, *62*(2), 288–305. doi:10.2307/2657305

Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt: Campus Verlag.

Firebaugh, G., Warner, C., & Massoglia, M. (2014). Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Reseach* (pp. 113–132). Dordrecht, Heidelberger, New York, London: Springer.

Halaby, C. N. (2004). Panel Models in Sociological Research: Theory into Practice. *Annual Review of Sociology*, *30*, 507–544.

doi:10.1146/annurev.soc.30.012703.110629

Haug, S. (2008). Migration Networks and Migration Decision-Making. *Journal of Ethnic and Migration Studies*, *34*(4), 585–605. doi:10.1080/13691830801961605

Hayfron, J. E. (2001). Language Training, Language Proficiency and Earnings of Immigrants in Norway. *Applied Economics*, *33*(15), 1971–1979. doi:10.1080/00036840010018630

Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, *36*, 49–67. doi:10.1146/annurev.soc.012809.102632

Hou, F., & Beiser, M. (2006). Learning the Language of a New Country: A Ten-Year Study of English Acquisition by South-East Asian Refugees in Canada. *International Migration*, *44*(1), 135–165. doi:10.1111/j.1468-2435.2006.00358.x

Kalter, F., & Granato, N. (2007). Educational Hurdles on the Way to Structural Assimilation in Germany. In A. F. Heath & S. Y. Cheung (Eds.), *Unequal Chances: Ethnic Minorities in Western Labour Markets* (pp. 271–319). Oxord: Oxford University Press.

Kalter, F., & Kogan, I. (2014). Migrant Networks and Labor Market Integration of Immigrants from the Former Soviet Union in Germany. *Social Forces*, *92*(4), 1435–1456. doi:10.1093/sf/sot155

Kleiner, B., Lipps, O., & Ferrez, E. (2015). Language Ability and Motivation among Foreigners in Survey Responding. *Journal of Survey Statistics and Methodology*, *3*(3), 339–360. doi:10.1093/jssam/smv015

Kogan, I. (2007). A Study of Immigrants' Employment Careers in West Germany Using the Sequence Analysis Technique. *Social Science Research*, *36*(2), 491–511. doi:10.1016/j.ssresearch.2006.03.004

Kogan, I. (2011). New Immigrants - Old Disadvantage Patterns? Labour Market Integration of Recent Immigrants into Germany. *International Migration*, *49*(1), 91–117. doi:10.1111/j.1468-2435.2010.00609.x

Kogan, I. (2016). Integration Policies and Immigrants' Labor Market Outcomes in Europe. *Sociological Science*, *3*, 335–358. doi:10.15195/v3.a16

Kristen, C. (2019). Language Assimilation and the Education of Immigant Students. In R. Becker (Ed.), *Research Handbook on the Sociology of Education* (pp. 519–534). Cheltenham: Edward Elgar Publishing.

Kristen, C., Mühlau, P., & Schacht, D. (2016). Language Acquisition of Recently Arrived Immigrants in England, Germany, Ireland, and the Netherlands. *Ethnicities*, *16*(2), 180–212. doi:10.1177/1468796815616157

Kroh, M., Brücker, H., Kühne, S., Liebau, E., Schupp, J., Siegert, M., & Trübswetter, P. (2016). *Das Studiendesign der IAB-BAMF-SOEP-Befragung von Geflüchteten.* SOEP Survey Papers 365: Series C. Berlin: DIW/SOEP. *SOEP Survey Papers*. Retrieved from: http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0365.pdf (accessed April 2020)

Kühne, S., & Kroh, M. (2017). *The 2015 IAB-SOEP Migration Study M2: Sampling Design, Nonresponse, and Weighting Adjustment.* Survey Papers 473: Series C. Berlin: DIW/SOEP. Retrieved from: https://www.diw.de/documents/publikationen/73/diw_01.c.571195.de/diw_ssp0473.pdf (accessed April 2020)

Lederer, H. W. (1997). *Migration und Integration in Zahlen. Ein Handbuch*. Retrieved from: http://www.efms.uni-bamberg.de/pubfor4d.htm (accessed April 2020)

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Nonlinear Principal Components Analysis: Introduction and Application. *Psychological Methods*, *12*(3), 336–358. doi:10.1037/1082-989X.12.3.336

Mesch, G. S. (2003). Language Proficiency among New Immigrants: The Role of Human Capital and Societal Conditions. The Case of Immigrants from the FSU in Israel. *Sociological Perspectives, 46*(1), 41–58. doi:10.1525/sop.2003.46.1.41

Meulman, J. J., van der Kooij, A. J., & Heiser, W. J. (2004). Principal Components Analysis With Nonlinear Optimal Scaling Transformations for Ordinal and Nominal Data. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 49–70). Thousand Oaks: SAGE Publications.

Rudolph, H. (1994). Dynamics of Immigration in a Nonimmigrant Country: Germany. In H. Fassmann & R. Münz (Eds.), *European Migration in the Late Twentieth Century. Historical Patterns, Actual Trends and Social Implications* (pp. 113–126). Aldershot: Edward Elgar Publishing.

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. Fu (Eds.), *The SAGE Handbook of Survey Methodology* (pp. 288–307). London: Sage Publications.

Schröder, H., & Ganzeboom, H. B. G. (2014). Measuring and Modelling Level of Education in European Societies. *European Sociological Review, 30*(1), 119–136. doi:10.1093/esr/jct026

Smith, T. W. (1995). Some Aspects of Measuring Education. *Social Science Research*, *24*(3), 215–242. doi.org:10.1006/ssre.1995.1008

Stevens, G. (1992). The Social and Demographic Context of Language Use in the United States. *American Sociological Review*, *57*(2), 171–185. doi:10.2307/2096203

Stevens, G. (1999). Age at Immigration and Second Language Proficiency among Foreign-Born Adults. *Language in Society*, *28*(4), 555–578. doi:10.1017/S0047404599004030

van Tubergen, F. (2010). Determinants of Second Language Proficiency among Refugees in the Netherlands. *Social Forces*, *89*(2), 515–534. doi:10.1353/sof.2010.0092

van Tubergen, F., & Kalmijn, M. (2009). A Dynamic Approach to the Determinants of Immigrants' Language Proficiency: The United States, 1980-2000. *International Migration Review*, *43*(3), 519–543. doi: 10.1111/j.1747-7379.2009.00776.x

van Tubergen, F., & Mentjox, T. (2014). Minority Language Proficiency of Adolescent Immigrant Children in England, Germany, the Netherlands, and Sweden. *Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *66*, 241–262. doi:10.1007/s11577-014-0273-6

van Tubergen, F., & Wierenga, M. (2011). The Language Acquisition of Male Immigrants in a Multilingual Destination: Turks and Moroccans in Belgium. *Journal of Ethnic and Migration Studies*, *37*(7), 1039–1057. doi:10.1080/1369183X.2011.572476

Wenz, A., Al Baghal, T., & Gaia, A. (2020 online first). Language Proficiency among Respondents: Implications for Data Quality in a Longitudinal Face-to-Face Survey. *Journal of Survey Statistics and Methodology*, 1–21. doi:10.1093/jssam/smz045

Data Reference

Liebig, S., Goebel, J., Schröder, C., Schupp, J., Bartels, C., & Alexandra Fedorets, Andreas Franken, Marco Giesselmann, Markus Grabka, Jannes Jacobsen, Selin Kara, Peter Krause, Hannes Kröger, Martin Kroh, Maria Metzing, Janine Napieraj, Jana Nebelin, David Richter, Diana Schacht, Paul Schmelzer, Christian Schmitt, Daniel Schnitzlein, Rainer Siegers, Knut Wenzig, Stefan Zimmermann. (2019). *Socio-Economic Panel (SOEP), data from 1984-2017*. doi:10.5684/soep.v34

5.8 Supplemental Material

Table A5.4 Basic descriptive of the variables for the established immigrants

Variable	overall				between			within			
	abs freq	freq in %	mean	std dev	abs freq	freq in %	std dev	%	std dev	min	max
dependent variable:											
German language skills (z-standardised)			0.00	1.00			1.00			-2.87	1.51
time-invariant variables:											
Length of stay at the first interview			13.22	6.56			6.62			5	47
German language class:			0.65	0.48			0.48			0	1
no	2606	34.68			717	34.99					
yes	4909	65.32			1332	65.01					
age at arrival			30.37	9.73			9.73			10	76
homeland education (z-standardised)			0.00	1.00			1.00			-2.99	2.93
fathers' level of education			2.11	1.86			1.90			1	3
low	2523	33.57			702	34.26					
medium/ high	4510	60.01			1210	59.05					
don't know/ not answer	482	6.41			137	6.69					
mother tongue skills (z-standardised)			0.00	1.00			1.00			-7.29	0.40
region:			1.66	0.85			0.86			1	4
Europe	3975	52.89			1087	53.05					
Russia	2559	34.05			686	33.48					
Arabic world	529	7.04			147	7.17					
Other	452	6.01			129	6.30					
sex:			0.56	0.50			0.50			0	1
male	3305	43.98			915	44.66					
Female	4210	56.02			1134	55.34					
time varying variables:											

Variable	overall				between			within			
	abs freq	freq in %	mean	std dev	abs freq	freq in %	std dev	%	std dev	min	max
Time			2.82	1.39			0.73		1.25	1	5
2013	1730	23.02			1730	84.43		29.54			
2014	1689	22.48			1689	82.43		29.15			
2015	1486	19.77			1486	72.52		26.47			
2016	1440	19.16			1440	70.28		25.7			
2017	1170	15.57			1170	57.10		24.11			
Current education:			0.04	0.20			0.15		0.14	0	1
no	7214	95.99			2030	99.07		96.82			
yes	301	4.01			191	9.32		43.78			
German education:			0.05	0.21			0.13		0.18	0	1
no	7155	95.21			2044	99.76		95.17			
yes	360	4.79			321	15.67		32.32			
Employed:			0.68	0.47			0.41		0.24	0	1
no	2442	32.5			946	46.17		70.42			
yes	5073	67.5			1620	79.06		85.36			
Partner living in household:			0.82	0.38			0.35		0.16	0	1
no	1330	17.7			505	24.65		72.83			
yes	6185	82.3			1780	86.87		94.45			
Child living in household:			0.66	0.48			0.45		0.16	0	1
no	2585	34.4			855	41.73		85.24			
yes	4930	65.6			1414	69.01		93.37			
Visit with natives:			0.83	0.37			0.31		0.22	0	1
no	1248	16.61			588	28.70		60.01			
yes	6267	83.39			1880	91.75		90.22			
Health:			3.42	1.05			0.89		0.60	1	5
poor	356	4.74			228	11.13		43.70			
not very good	1145	15.24			686	33.48		44.70			
satisfactory	1998	26.59			1139	55.59		46.94			
well	2989	39.77			1429	69.74		57.17			
very well	1027	13.67			589	28.75		49.42			

Variable	overall abs freq	overall freq in %	overall mean	overall std dev	between abs freq	between freq in %	between std dev	within %	within std dev	min	max
Settlement intention:			0.81	0.39			0.34		0.21	0	1
no	1,422	18.92			597	29.14		67.32			
yes	6093	81.08			1821	88.87		90.45			
Feel as German:			3.25	1.12			0.96		0.61	1	5
not at all	639	8.5			353	17.23		51.71			
barely	971	12.92			511	24.94		51.33			
in some respects	2856	38			1249	60.96		62.98			
completely	1086	14.45			502	24.50		58.73			
mostly	1963	26.12			928	45.29		56.32			

Data sources: IAB-SOEP migration sample (M1-M2) 2013-2017, Data file version 34

Table A5.5 Basic descriptive of the variables for the recently arrived immigrants

	overall				**between**			**within**			
Variable	**abs freq**	**freq in %**	**mean**	**std dev**	**abs freq**	**freq in %**	**std dev**	**%**	**std dev**	**min**	**max**
dependent variable:											
German language skills (z-standardised)			0.00	1.00			1.00			-1.99	2.20
time-invariant variables:											
Length of stay at the first interview			1.40	0.70			0.70			0	3
less than 1 year	180	4.27			90	4.27					
1 year	2542	60.28			1271	60.29					
2 years	1132	26.86			566	26.85					
3 years	362	8.59			181	8.59					
Contact with natives:			4.01	1.84			1.84			1	6
never	696	16.51			348	16.51					
less often	494	11.72			247	11.72					
every month	244	5.79			122	5.79					
every week	664	15.75			332	15.75					
several times per week	890	21.12			445	21.11					
every day	1228	29.11			614	29.13					
Age at arrival:			32.63	9.90			9.90			16	74
Homeland education (z-standardised)			0.00	1.00			1.00			-2.17	2.67
Fathers' level of education			2.60	2.81			2.81			1	3
low	2135	50.63			1067	50.62					
medium/ high	1416	33.59			708	33.59					
don't know/ not answer	665	15.78			333	15.80					
Mother tongue skills (z-standardised)			0.00	1.00			1.00			-5.23	0.51
			3.03	0.46			0.46			1	4

	overall				between			within			
Variable	**abs freq**	**freq in %**	**mean**	**std dev**	**abs freq**	**freq in %**	**std dev**	**%**	**std dev**	**min**	**max**
Region of origin:											
Europe	82	1.95			41	1.94					
Russia	148	3.51			74	3.51					
Arabic world	3548	84.15			1774	84.16					
Other	438	10.39			219	10.39					
Sex:			0.35	0.48			0.48			0	1
male	2736	64.89			1368	64.90					
female	1480	35.11			740	35.10					
time varying variables:											
Time			1.50	0.50			0.01		0.50	1	2
2016	2108	50.00			2108	100.00		50.00			
2017	2108	50.00			2108	100.00		50.00			
German language class:			0.76	0.43			0.34		0.25	0	1
no	1003	23.8			768	36.43		65.30			
yes	3213	76.2			1873	88.85		85.77			
Current education:			0.06	0.24			0.18		0.15	0	1
no	3961	93.95			2080	98.67		95.22			
yes	255	6.05			227	10.77		56.17			
Employed:			0.15	0.36			0.28		0.22	0	1
no	3571	84.72			1994	94.59		89.57			
yes	645	15.28			530	25.14		60.75			
Partner living in household:			0.61	0.49			0.47		0.13	0	1
no	1623	38.51			881	41.79		92.11			
yes	2593	61.49			1366	64.80		94.91			
Child living in household:			0.60	0.49			0.48		0.09	0	1
no	1678	39.81			872	41.37		96.22			
yes	2538	60.19			1302	61.76		97.47			
Health:			3.98	1.06			0.90		0.55	1	5
poor	100	2.37			94	4.46		53.19			
not very good	389	9.23			333	15.80		58.41			

Variable	overall abs freq	overall freq in %	overall mean	overall std dev	between abs freq	between freq in %	between std dev	within %	within std dev	min	max
satisfactory	612	14.52			524	24.86		58.40			
well	1490	35.35			1173	55.65		63.55			
very well	1625	38.53			1147	54.41		70.79			
Settlement intention:			0.90	0.30			0.24		0.18	0	1
no	434	10.3			354	16.79		61.30			
yes	3782	89.7			2028	96.20		93.24			

Data sources: IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34

Table A5.6 Results from further regression analyses for the established immigrants

	Model 3 **incl. 2 measures of homeland education (years of schooling and categorical levels)**				**Model 4** **incl. the categorical education measure only, thus also some more cases**				**Model 5** **incl. immigrants who are living in Germany at least for 8 years**			
	b		**SE**	**p**	**b**		**SE**	**p**	**b**		**SE**	**p**
Language exposure												
length of stay at first interview	0.01	**	0.00	.001	0.01	**	0.00	.007	0.00		0.00	.407
survey year (ref: 2017)												
2013	0.05	*	0.02	.019	0.05	*	0.02	.027	0.09	***	0.02	<.001
2014	-0.02		0.02	.311	-0.02		0.02	.214	0.01		0.02	.706
2015	0.01		0.02	.786	0.00		0.02	.837	0.05		0.02	.049
2016	-0.01		0.02	.465	-0.01		0.02	.464	0.03		0.02	.275
German class	0.14	***	0.03	<.001	0.14	***	0.03	<.001	0.10	**	0.04	.008
German education	0.98	***	0.11	<.001	0.87	***	0.11	<.001	1.09	***	0.13	<.001
German education (time-varying)	-0.07		0.03	.034	-0.05		0.03	.062	-0.09	*	0.04	.017
current education	0.25	*	0.11	.024	0.21	*	0.10	.037	0.27	*	0.11	.014
current education (time-varying)	-0.10	*	0.04	.011	-0.08	*	0.04	.022	-0.07		0.04	.110
employed	0.22	***	0.04	<.001	0.23	***	0.04	<.001	0.23	***	0.05	<.001
employed (time-varying)	0.02		0.02	.344	0.02		0.02	.404	0.01		0.03	.673
partner	-0.08		0.05	.079	-0.07		0.04	.105	-0.04		0.05	.384
partner (time-varying)	-0.02		0.04	.603	-0.03		0.03	.441	-0.01		0.04	.768
children	-0.10	*	0.04	.011	-0.09	*	0.04	.011	-0.12	**	0.04	.005
children (time-varying)	-0.01		0.04	.837	0.00		0.04	.944	0.00		0.04	.942
visit with Germans at home	0.47	***	0.05	<.001	0.49	***	0.05	<.001	0.49	***	0.06	<.001
visit with Germans at home (time-varying)	0.05		0.03	.103	0.05		0.03	.107	0.06		0.03	.088
Efficiency												
age at arrival	-0.02	***	0.00	<.001	-0.02	***	0.00	<.001	-0.02	***	0.00	<.001
homeland education (index)	not included				not included				0.27	***	0.02	<.001
homeland education: (ref: no graduation and no further training)									not included			
mandatory school *and* no further training	0.16	*	0.07	.016	0.28	***	0.06	<.001				
no graduation/ mandatory school *and* apprenticeship	0.24	**	0.09	.006	0.38	***	0.08	<.001				

	Model 3 incl. 2 measures of homeland education (years of schooling and categorical levels)				Model 4 incl. the categorical education measure only, thus also some more cases				Model 5 incl. immigrants who are living in Germany at least for 8 years			
	b		**SE**	**p**	**b**		**SE**	**p**	**b**		**SE**	**p**
no graduation/ mandatory school *and* vocational school	0.11		0.08	.157	0.23	**	0.08	.004				
higher-level secondary *and* no further training	0.36	***	0.08	<.001	0.61	***	0.07	<.001				
higher-level secondary *and* apprenticeship	0.32	***	0.09	<.001	0.58	***	0.08	<.001				
higher-level secondary *and* vocational school	0.28	**	0.09	.002	0.53	***	0.08	<.001				
university/ college - practical orientation	0.41	***	0.09	<.001	0.65	***	0.08	<.001				
university/ college - theoretical orientation	0.54	***	0.08	<.001	0.78	***	0.07	<.001				
doctoral studies	0.66	***	0.14	<.001	0.93	***	0.12	<.001				
years of schooling	0.07	***	0.01	<.001	not included				not included			
fathers' education (ref: low)												
medium, high	0.11	**	0.04	.003	0.13	***	0.03	<.001	0.09	*	0.04	.023
no answer	0.07		0.07	.317	0.07		0.07	.274	0.08		0.08	.305
mother tongue	0.05	***	0.02	<.001	0.08	***	0.02	<.001	0.04	**	0.02	.005
region of origin (ref: Europe)												
Russia	0.06		0.04	.082	0.04		0.03	.288	0.02		0.04	.703
Arabic world	-0.07		0.06	.242	-0.02		0.06	.701	-0.11		0.07	.083
other	-0.22	**	0.07	.001	-0.19	**	0.07	.006	-0.18	*	0.08	.037
health	0.05	*	0.02	.013	0.06	**	0.02	.002	0.06	*	0.02	.012
health (time-varying)	0.01		0.01	.175	0.01		0.01	.134	0.01		0.01	.222
Incentives												
settlement intention	-0.11	*	0.05	.020	-0.13	**	0.05	.005	-0.12	*	0.06	.038
settlement intention (time-varying)	0.01		0.03	.625	0.02		0.03	.449	0.06		0.03	.099
feel as German	0.25	***	0.02	<.001	0.25	***	0.02	<.001	0.26	***	0.02	<.001
feel as German (time-varying)	0.04	***	0.01	<.001	0.04	***	0.01	<.001	0.04	***	0.01	<.001
gender	0.14	***	0.03	<.001	0.14	***	0.03	<.001	0.12	**	0.04	.001
constant	-2.04	***	0.16	<.001	-1.69	***	0.15	<.001	-0.99	***	0.17	<.001
R^2 overall	40.50				39.46				40.73			
R^2 between	46.52				45.03				47.05			

	Model 3 incl. 2 measures of homeland education (years of schooling and categorical levels)			**Model 4 incl. the categorical education measure only, thus also some more cases**			**Model 5 incl. immigrants who are living in Germany at least for 8 years**		
	b	**SE**	**p**	**b**	**SE**	**p**	**b**	**SE**	**p**
R^2 within	0.99			0.99			1.33		
number of cases	2,049			2,113			1,594		
number of observations	7,515			7,751			6,011		

Data sources: IAB-SOEP migration sample (M1-M2) 2013-2017 and IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34

Table A5.7 Results from further regression analyses for the recently arrived immigrants

	Model 6 **incl. 2 measures of homeland education (years of schooling and categorical levels)**				**Model 7** **incl. the categorical education measure only, thus also some more cases**				**Model 8** **incl. recently and establsihed immigrants who are living in Germany for less than 5 years in Germany**			
	b		**SE**	**p**	**b**		**SE**	**p**	**b**		**SE**	**p**
Language exposure												
length of stay at first interview	0.16	***	0.02	<.001	0.15	***	0.02	<.001	0.13	***	0.02	<.001
survey year (ref: 2017)												
2013									-0.33	***	0.05	<.001
2014									-0.29	***	0.05	<.001
2015									-0.32	***	0.02	<.001
2016	-0.48	***	0.02	<.001	-0.47	***	0.02	<.001	-0.42	***	0.02	<.001
German class	0.79	***	0.04	<.001	0.77	***	0.04	<.001	0.52	***	0.04	<.001
German language class (time-varying)	0.26	***	0.04	<.001	0.26	***	0.04	<.001				
German education									0.27		0.16	.099
German education (time-varying)									0.04		0.06	.523
current education	0.54	***	0.08	<.001	0.53	***	0.08	<.001	0.64	***	0.06	<.001
current education (time-varying)	0.07	***	0.06	.257	0.07		0.06	.260	0.03		0.04	.500
employed	0.27	***	0.05	<.001	0.26	***	0.05	<.001	0.34	***	0.04	<.001
employed (time-varying)	0.07		0.04	.078	0.08		0.04	.060	0.04		0.03	.143
partner	-0.09	*	0.04	.045	-0.09	*	0.04	.025	-0.12	**	0.04	.001
partner (time-varying)	0.22	**	0.07	.001	0.22	**	0.06	.001	0.12	*	0.05	.010
children	-0.04		0.04	.319	-0.05		0.04	.201	-0.03		0.04	.379
children (time-varying)	-0.07		0.12	.533	-0.07		0.11	.537	-0.02		0.06	.794
spend time with Germans	0.08	***	0.01	<.001	0.08	***	0.01	<.001				
Efficiency												
age at arrival	-0.02	***	0.00	<.001	-0.02	***	0.00	<.001	-0.02	***	0.00	<.001
homeland education (index)									0.25	***	0.01	<.001
homeland education: (ref: no graduation and no further training)												

mandatory school *and* no further training	0.08		0.04	.055	0.17	***	0.04	<.001				
no graduation/ mandatory school *and* apprenticeship	0.24		0.15	.118	0.31		0.16	.043				
no graduation/ mandatory school *and* vocational school	0.13		0.16	.426	0.20		0.15	.197				
higher-level secondary *and* no further training	0.12	*	0.05	.013	0.32	***	0.04	<.001				
higher-level secondary *and* apprenticeship	0.25	*	0.11	.029	0.45	***	0.11	<.001				
higher-level secondary *and* vocational school	0.31	**	0.12	.008	0.45	***	0.12	<.001				
university/ college - practical orientation	0.28	***	0.07	<.001	0.49	***	0.06	<.001				
university/ college - theoretical orientation	0.40	***	0.06	<.001	0.58	***	0.05	<.001				
doctoral studies	0.94	***	0.14	<.001	1.08	***	0.13	<.001				
years of schooling	0.05	***	0.01	<.001								
fathers' education (ref: low)												
medium, high	0.07	*	0.03	.031	0.07	*	0.03	.017	0.08	*	0.03	.010
no answer	0.01		0.04	.735	0.01		0.04	.852	0.02		0.04	.687
mother tongue	0.08	***	0.01	<.001	0.11	***	0.02	<.001	0.09	***	0.01	<.001
region of origin (ref: Europe)												
Russia	-0.19		0.13	.131	-0.18		0.13	.145	0.01		0.06	.842
Arabic world	-0.06		0.11	.584	-0.04		0.11	.697	0.04		0.05	.404
other	-0.39	**	0.12	.001	-0.36	**	0.12	.002	-0.32	***	0.06	<.001
health	0.06	***	0.02	<.001	0.07	***	0.02	<.001	0.08	***	0.02	<.001
health (time-varying)	0.00		0.02	.860	0.00		0.02	.846	0.00		0.01	.890
Incentives												
settlement intention	0.03		0.06	.634	0.02		0.05	.738	0.03		0.05	.561
settlement intention (time-varying)	0.04		0.05	.423	0.03		0.05	.483	0.07		0.04	.089
gender	-0.09	**	0.03	.003	-0.07	**	0.03	.004	-0.04		0.03	.149
constant	-1.10	***	0.16	<.001	-0.73	***	0.15	<.001	-0.36	**	0.12	.002
R^2 overall	45.02				43.64				34.83			
R^2 between	50.70				48.85				37.84			
R^2 within	28.26				28.27				20.02			
number of cases	2,108				2,140				3,110			
number of observations	4,216				4,278				6,578			

Data sources: IAB-SOEP migration sample (M1-M2) 2013-2017 and IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34

6 Additional Analysis to Paper IV: A Construct Validation of Different Education Measures in the SOEP Migration Surveys[19]

6.1 Introduction

As illustrated in section 1.4.2 of this thesis, there are three approaches to measuring immigrants' homeland education in migration surveys (Schneider, 2018). The IAB-SOEP migration samples and the IAB-BAMF-SOEP survey of refugees in Germany, which I used for the analysis in paper IV, implemented different measurement instruments and thus offer a larger number of education variables. This includes the years of schooling variable and different variables that can be derived from the SOEP 'generic' standard education measure, which asks immigrants about their homeland education using an adapted German instrument. In these surveys, the CAMCES tool has been implemented that asked immigrants on their education using country-specific education measures and from this instrument also different education variables are derived. This additional analysis looks at these different education measures in more detail. To identify which education measure should be best used when analysing the impact of immigrants' educational attainment on their German language proficiency, we conduct a construct validation.

In the next section, we present the data and the validation strategy. In section 3, we describe the different measurement instruments for immigrants' education employed in these studies, and the variables used in the validation. The results of the validation are presented in section 4. We conclude with a brief summary of results and a discussion and we will motivate the usage of the index for the analysis in paper IV.

---

[19] This additional analysis is part of the manuscript "Measuring Migrants' Homeland Education: A Validation Study of Competing Measures". This manuscript is co-authored by Silke L. Schneider who is the first author. The manuscript is submitted as contribution to the book "Empirische Sozialforschung in Zeiten der Digitalisierung", edited by F. Faulbaum Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute, Wiesbaden: Springer VS.

6.2 Data and Validation Strategy

The data we used for this validation analysis has been introduced in paper IV (see section 5.3.1). In this analysis, we validate measures of immigrants' homeland education and therefore, we select only respondents who received their education abroad. We excluded respondents who have a German educational qualification or who currently attend a German educational programme (mostly very young respondents). Unfortunately, we cannot identify and exclude respondents who have started a German educational programme but did not complete it. We also decided to remove respondents who were below the age of 18 when arriving in Germany from our analysis sample because they would have been obliged to attend education in Germany. Thereby we exclude all immigrants who (likely) have received parts of their education in Germany. In addition, we excluded respondents stating that they have never visited a school and those not responding to the education question(s).[20]

In this analysis, we look at the construct validity of different measures of education, using immigrants' German language proficiency as dependent variable and adjusted R^2 to measure explanatory power. We run separate analyses for the two groups of immigrants because the impact of homeland education likely differs by length of stay in Germany. Here, we will not only look at data coded into ISCED but also at other education variables.

6.3 Education Measures and Derived Variables

In this section, we firstly describe the questionnaire instruments used in the SOEP migration surveys to obtain information on respondents' foreign education, including reflections on some potential measurement and comparability problems. Then, we present the education variables we derive from this information, and on which the validation analyses are based.

[20] The selection criteria for this construction validation analysis differ slightly from the criteria described in paper IV. In the latter I, for instance, control for the impact of having a German qualification, which is not suitable for this validation analysis. The criteria for the defining the groups of recently arrived and established immigrants also differ slightly from those described in paper IV. In the latter, the selection criteria to the groups are based more theoretically, especially concerning variables indicating the years of migration and years spent in Germany. For the validation analysis, the definitions of the two groups are less strict and more in analogy with the SOEP samples M1-M2 of the established and M3-M5 of the recently arrived immigrants. Moreover, the age cut differs between the paper IV and the construct validation.

6.3.1 The Different Measurement Instruments in the SOEP Migration Surveys

In the SOEP surveys, three different instruments for measuring immigrants' homeland education are implemented. The first instrument asks respondents on the number of years they spent in school outside of Germany.[21] The second instrument, which we will refer to as the SOEP standard instrument asks the same two (generic) categorical questions to every respondent who indicates that he/ she has last received education abroad. The first question asks for their highest school leaving certificate (left school without graduating, graduated from a mandatory school, graduated from a higher-level secondary school), and the second for the highest post-school qualification (in-house training, extended apprenticeship at a company, vocational school, university/ college with a more practical orientation, university/ college with a more theoretical orientation, doctoral studies, or other post-school education).

The third instrument offers respondents culturally adapted response options reflecting the educational qualifications of the country in which the immigrant was educated by implementing the CAMCES tool in the CAPI system used for SOEP data collection. The CAMCES tool was employed in the migrant sample in 2015, 2016 and 2017 and in the refugee sample 2017, in the earliest re-interview of every sample member.[22] All respondents who report to have foreign educational qualifications are routed into the CAMCES tool. Then, firstly, the country where the respondent received his/ her education is identified. Secondly, respondents are asked for their highest educational qualification. For their response, they can search their educational qualification in the CAMCES database using text string matching, or they select it from a country-specific list of educational qualifications, which is also stored in the CAMCES database.[23] After data collection, the codes of these country-specific categories are recoded into ISCED, again relying on information included in the CAMCES database.

Asking two questions is common when measuring educational attainment in Germany but not in other countries. It may be problematic if the term 'school' refers to different parts of educational systems in other languages. In contrast to the SOEP

[21] Question text: How many years did you attend school? in M1-M2, and in M3-M4: How many years did you attend school in total?
[22] We thus lose a number of cases due to panel attrition.
[23] We thus lose some further cases because they were educated in a country not (at that point in time) covered by the CAMCES database.

instrument, the question module in which the CAMCES tool is embedded asks only one question on the highest foreign educational qualification (no matter whether from 'school' or other institutions). The answer categories of the SOEP questions are also inspired by the German educational system, revealing the difficulty of phrasing universally applicable education categories. For example, the category “Graduated from mandatory schooling with school-leaving certificate” on the first item will refer to different levels of education in different countries of origin, since the length of mandatory school differs across countries. Regarding the CAMCES tool, this places a burden on the respondent to remember and report his/ her foreign education in the respective language, which may be difficult especially for older respondents who completed their education many years ago, and who may have lived in Germany for many years. More details on the CAMCES tool and its implementation in the SOEP migration surveys can be found in Briceno-Rosas et al. (2018) and Schneider et al. (2018).

6.3.2 Education Measures and Dependent Variable Used in the Validation

We compare a number of different education measures. The simplest one refers to years of schooling, which is asked directly (see above). We top-coded this variable at 13 because school in all countries stops/ ends after 13 years at most, but a substantial number of respondents, especially in the refugee sample, reported more years of education (possibly because the term 'school' is in some languages understood in a broader sense than in German). The most detailed measure derived from the SOEP 'standard instrument' combines the two original variables into one with ten categories. The most detailed variable derived from the CAMCES measurement is a three-digit ISCED 2011 code (UNESCO-UIS, 2012) with 17 categories. Moreover, we derive two categorical variables that can be compared across the SOEP and CAMCES measures, one with eight and one with three categories, as shown in Table 6.1.

Lastly, we generate four education indices using a scoring approach, which allows us to combine the information included in different education variables (Braun & Müller, 1997; Schröder & Ganzeboom, 2014). To generate the index, we a conduct non-linear principal component analysis (PCA) which allows scoring of variables at different levels of measurement (Linting, Meulman, Groenen, & van der Kooij, 2007; Meulman, van der Kooij, & Heiser, 2004). Thereby we can combine the most detailed categorical SOEP and CAMCES variables and the metric years of schooling variable described above. In one index we combine the information of all three variables and in the other three indices we use all combinations of two education variables following the same approach.

The dependent variable for the construct validation is an index measuring immigrants' German language skills. This has been described in paper IV in section 5.3.2.

Table 6.1 Overview of categorical education variables (note that detailed SOEP and CAMCES categories on the same row do not always match)

SOEP standard measure			Comparable variables		Harmonized CAMCES measure	
item school education	**item post-school education**	**Code**	**8 categories**	**3 categories (broad ISCED)**	**(ISCED 2011, 3 digits)**	
left school without graduating	no further training OR in-house OR other training	(1)	ISCED 0-1	low	ISCED 0	primary education not completed
					ISCED 100	primary education
graduated from mandatory school	no further training OR in-house OR other training	(2)	ISCED 2		ISCED 250	vocational lower secondary
					ISCED 240	general lower secondary
left school without graduating OR graduated from mandatory school	(not covered)		ISCED 3 vocational	medium	ISCED 352	partial vocational upper secondary
					ISCED 353	vocational upper secondary without access to tertiary
	extended apprenticeship at a company	(3)			ISCED 354	vocational upper secondary with access to tertiary
	vocational school	(4)				
	(not covered)		ISCED 3 general		ISCED 343	general upper secondary without access to tertiary
graduated from a higher-level secondary school	no further training OR in-house OR other training	(5)			ISCED 344	general upper secondary with access to tertiary
	(not covered)		ISCED 4 vocational		ISCED 453	vocational post-secondary non-tertiary without access to tertiary
	extended apprenticeship at a company	(6)			ISCED 454	vocational post-secondary non-tertiary with access to tertiary
	vocational school	(7)				
any	(not covered)		ISCED 5 (vocational)	high	ISCED 550	vocational short-cycle tertiary
					ISCED 540	general short-cycle tertiary
	university/ college with a more practical orientation	(8)	ISCED 6-7		ISCED 600	Bachelor's level or equivalent
	university/ college with a more theoretical orientation	(9)			ISCED 700	Master's level or equivalent
	doctoral studies	(10)	ISCED 8		ISCED 800	Doctoral level or equivalent

6.4 Results

We run a construct validation to identify which education measure has the highest predictive power. The dependent variable is the index of immigrants' German language skills. We calculate one model for each education measure, and in all models, we control for age and sex. We run the analyses separately for the established and the recent immigrants because the measures may work somewhat differently across these groups. We only include cases for which we have valid data in all three instruments and thus the case number reduces to 1157 respondents in the immigrant sample and 1054 respondents in the refugee sample.

Comparing the two immigrant groups, the explanatory power of homeland education on second language skills is considerably (about five percentage points) lower amongst the established immigrants than recent immigrants, no matter which education variable we look at (see Figure 6.1). It is very plausible that homeland education loses its relevance the longer an immigrant resides in the destination country, and other factors such as work experience or social networks with members of the majority population will prevail.

Within the two immigrant groups, the most detailed versions of the SOEP and CAMCES measures have pretty much the same explanatory power (12 and 11% respectively in the refugee samples, and both measures 7% in the samples of established immigrants). This is a strong indication that the SOEP measure, being much simpler than the CAMCES measure, does not miss any crucial information. In contrast, years of education fare worse, especially in the sample of established immigrants where it explains less than 2% of the variance. For recent immigrants, years of education work relatively well though (adjusted R^2=9.3%), which is quite remarkable given this variable only covers school education. Among the established immigrants, we may see a memory effect in addition to the loss of importance of homeland education mentioned above: it is probably more difficult to remember the years of education than the educational qualification if the completion of education was long ago, as is often the case amongst established immigrants educated abroad. This question may also be interpreted differently by people from different countries of origin or speaking different languages, and years of education may correlate differently with other variables - e.g. cognitive skills - across countries. The sample of recent immigrants is more homogeneous than the sample of established immigrants, and years of education may be a better proxy for cognitive skills in the former than in the latter (e.g. because compared to Middle Eastern countries, the Soviet Union has long had effective compulsory schooling). When collapsing the detailed into more aggregate categorical measures, the CAMCES-based variables lose more explanatory power than the SOEP-based variables (2 vs. 1 percentage point in the 3-category version). This is not much in the case of recent immigrants, but for established immigrants, it amounts to more than 10% of the original explanatory power. But even the three-category measures work quite well in these models.

Looking at the indices, we again find strong differences across the immigrant groups. For the recently arrived immigrants the index combining the years of schooling variable and the SOEP measure has the highest predictive power. For the same index we observe the lowest predictive power of all indices for the established immigrants. This is quite surprising. However, we have to keep in mind that the years of schooling variable alone has a much lower predictive power for the established immigrants, which also decreases the predictive power of the indices in which this variable is included. Therefore, the index combining the SOEP and the CAMCES measure has a higher predictive power for the established immigrants compared to the indices combining one of these measures with the years of schooling variable. Overall, the indices do not have a much higher predictive power than the SOEP measure alone. However, since these variables have a metric scale, they may be preferable for data users, especially if they are not interested in the signalling effects of educational qualifications, which are less meaningful when looking at immigrants' homeland education anyway (Friedberg, 2000; Weins, 2010).

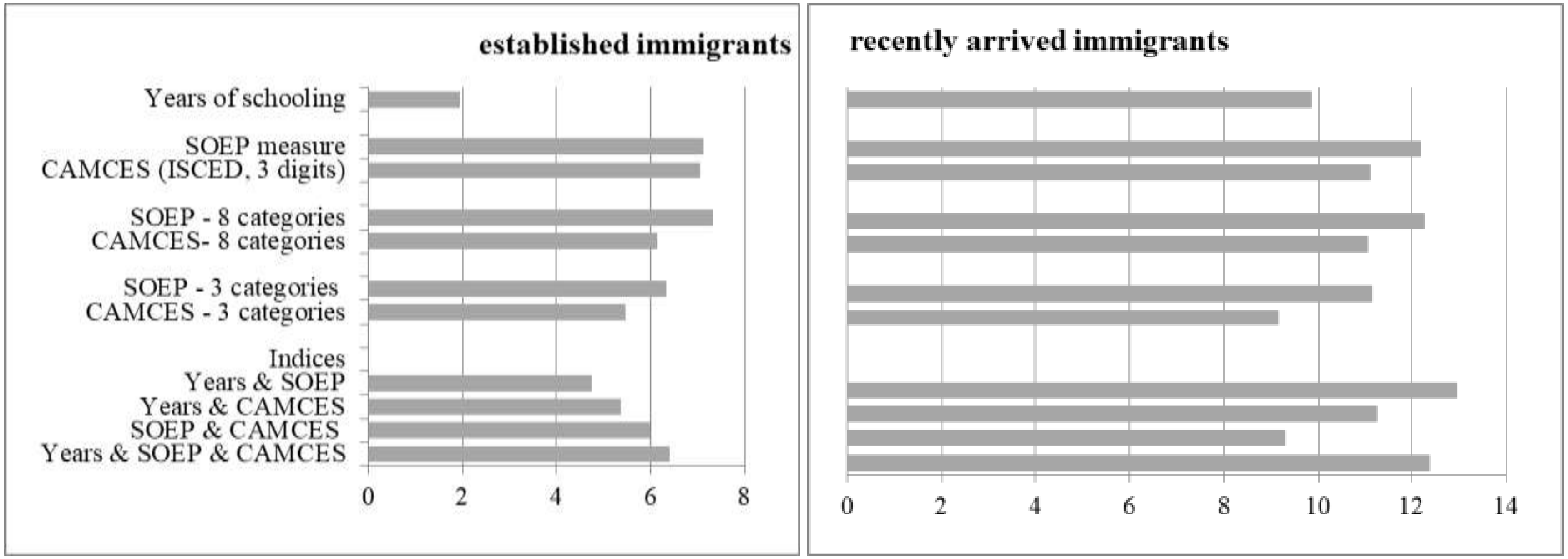


Figure 6.1 Adjusted R^2 (in %) for different education variables predicting immigrants' German language proficiency

N=1157 (established immigrants) and 1054 (recent immigrants). Analysis controls for age and sex.
Data sources: IAB-SOEP migration sample (M1-M2) 2013-2017 and IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34.

Next let's look in some more detail at the coefficients of some of these models to check how comparable the effects of different education categories are across the categorical measurement instruments. Here, we only take the refugee sample since the relationship between homeland education and second language skills is closer in this group, and focus on the comparable education variables with 8 and 3 categories. As Table 6.2 reveals, for the 8-category measures, the same education categories are statistically significantly related to German language skills across the SOEP and CAMCES instruments, and the standard errors are also similar. Bachelor's and Master's level education have highly similar effects across both measures. These are very good signs for the validity of both measures. Only doctoral education – even though it is rare – has a strong effect according to the SOEP measure, while the CAMCES measure captures too few respondents with PhDs to reliably estimate an effect. Therefore, it appears that the SOEP measure is better able to capture PhDs. The substantial effect of a PhD likely leads to a slightly higher explanatory power of the SOEP measure compared to the CAMCES measure. Below tertiary education, the effects are slightly stronger in the CAMCES than the SOEP measure. Vocational education, including short cycle tertiary education (ISCED 5), does not have any effects (with the exception of post-secondary non-tertiary education, ISCED 4). This supports the idea that ISCED 5 should not be regarded as higher education (Schneider, 2008). Even though the SOEP measure is less able to capture vocational education than the CAMCES measure, this shortcoming is empirically inconsequential in this validation analysis. This is because vocational education hardly pays off in terms of second language skills - no matter which measure we look at. In sum, the different measurement instruments lead to highly consistent results, with the exception of doctoral education, which is generally a very small category though.

Table 6.2 Education effects of 8 category SOEP and CAMCES education measures in the SOEP refugee samples

	SOEP 8 categories					CAMCES 8 categories				
	n	b		SE	p	n	b		SE	p
ISCED 0-1	189	-0.18	*	0.09	0.043	119	-0.30	**	0.10	0.004
ISCED 2	235	ref.				217	ref.			
ISCED 3 vocational	19	0.09		0.22	0.681	112	0.16		0.11	0.126
ISCED 3 general	323	0.31	***	0.08	0.000	225	0.40	***	0.09	0.000
ISCED 4 vocational	30	0.39	*	0.18	0.026	56	0.45	**	0.14	0.001
ISCED 5 (vocational)	-					84	0.15		0.12	0.205
ISCED 6-7	240	0.55	***	0.08	0.000	236	0.55	***	0.09	0.000
ISCED 8	18	1.31	***	0.22	0.000	5	0.08		0.41	0.850
adj. R^2	12.26					11.05				

N= 1054 (recent immigrants only). Analysis controls for age and sex not shown.
Data sources: IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34.

When reducing the categories to three (see Table 6.3), the order of magnitude of the effects is similar, even though upper secondary is more effective when looking at the CAMCES rather than the SOEP measure, and for tertiary education, it is the other way around. The effects of medium and high education are as a result more differentiated in the SOEP than the CAMCES measure. The explanatory power of the 3-category measure derived from CAMCES increases when ISCED 5 is aggregated with ISCED 3 and 4 rather than 6 and 7 though, and then the effects also become more similar between the SOEP and CAMCES measures. Compared to earlier research (Schneider, 2010), the three-category measure works reasonably well here because in this refugee sample, the heterogeneity within the broad categories is relatively low, since most cases accumulate in a few paradigmatic categories that mostly spread across broad levels: ISCED 0/ 1, ISCED 2, ISCED 3 general, and Bachelor's level education. The minor losses of explanatory power are to about a third driven by the aggregation of primary education or less with lower secondary education, a distinction that is relevant amongst the population of recent immigrants and important for host country language acquisition. This is a reminder that the aggregation of education categories for analysis should take the specific sample into account.

Table 6.3 Education effects of 3-category SOEP and CAMCES education measures in the SOEP refugee samples

	SOEP 3 categories					CAMCES 3 categories				
	n	b		SE	p	n	b		SE	p
low	424	Ref.				336	Ref.			
medium	372	0.38	***	0.06	0.000	393	0.45	***	0.07	0.000
high	258	0.68	***	0.07	0.000	325	0.55	***	0.07	0.000
adj. R^2	11.14					9.15				

N= 1054 (recent immigrants only). Analysis controls for age and sex not shown.
Data sources: IAB-BAMF-SOEP survey of refugees in Germany (M3, M4 and M5) 2016 and 2017, Data file version 34.

To sum up, apart from the years of schooling variable and the education indices involving years of schooling in the migration sample, the predictive power of the different education variables is rather similar, despite their highly different underlying measurement instruments. Thus, for analysing immigrants' German language proficiency, almost all measures of homeland education can be used without facing strong biases due to the measurement of immigrants' homeland education. The generic SOEP measure works remarkably well as a predictor of second language skills.

6.5 Conclusion and Discussion

This analysis examined different ways of measuring immigrants' homeland education in surveys, using SOEP migration and refugee sample data. We specifically compared the popular 'years of schooling', education measures based on 'generic' questionnaire items, and measures based on country-specific items, which were administered by implementing the CAMCES tool in the SOEP survey. We also constructed education scales by combining these different types of measures.

Analysing the construct or predictive validity of the different education measures we recognise that the SOEP and the CAMCES measures perform equally well when looking at the detailed variables. The generic SOEP measure even showed to be less sensitive to aggregation error when simplifying the detailed measures to 8 and then 3 categories, most likely just *because* vocational education is not included and thus the education of those with vocational education is coded as one level lower than what ISCED (and thus the CAMCES measure) would code. Years of education fare somewhat worse, but this strongly differs by immigrant group. This also affects the performance of the indices including years of schooling for the established immigrants. For the sample of refugees, all indices perform almost equally well and thus combining different variables to generate a metric index seems to be quite feasible.

In conclusion, depending on the purpose of the measurement, i.e. the theoretical meaning and interpretation to be attached to educational attainment in a specific study and the outcomes of education to be studied, several solutions to the challenge of measuring immigrants' homeland education can be envisaged (Schneider, 2018). If the purpose is to know respondents' absolute level of education (e.g. whether lower or upper secondary education was completed) the survey, strictly speaking, needs to measure the specific educational qualifications available in the country of origin and recode these into ISCED after data collection. This would be advisable e.g. to produce official statistics on the education of (especially recent) immigrants in a country. The resulting data can be compared across countries and can also be transformed to an international education classification such as ISCED. This is surely the most demanding approach in terms of both effort and costs, especially if a survey does not focus on a few countries of origin but the whole migrant population.

If the aim is rather to know the respondents' approximate position in the education distribution to e.g. proxy cognitive skills in order to correlate this with outcomes of education or skills, it may be sufficient to measure education in less specific terms. However, this approach will only allow deriving ISCED based on the application of ISCED criteria to these general response options rather than with reference to specific foreign qualifications and their 'real' classification according to the ISCED mappings. This may be sufficient for some survey projects or research questions though. This approach may work well for migrant surveys because foreign qualifications do not have the same (if any) signalling character in the labour market as domestic qualifications (Weins, 2010), so that the symbolic meaning attached to a specific foreign qualification in the country of origin will not matter much in the destination country. Also, institutional specificities that would remain invisible when using this approach may not matter much in migration research, unless they are strongly linked with factors that do matter in the destination country, such as respondents' cognitive competences or a privileged social background. This 'generic' approach may however be less promising when there is a specific interest in vocational education and training. This has shown to be under-identified with the generic questionnaire items used in the SOEP. Our validation variable, language skills, is certainly more strongly linked with the kinds of cognitive skills that are best developed in general education, so that in this specific validation, the generic measure works very well. With a different dependent variable in focus, this may look somewhat differently.

The years of schooling variable in general is not without problems (Braun & Müller, 1997; Schneider, 2010), this also holds for migration surveys. This measure challenges comparative cross-national as well as cross-cultural research. The term ‘school’ used in this item can be understood differently, depending on the educational system. Years of schooling may also correlate differently with cognitive skills (or other concepts one may want to measure with education) across countries. This might explain that this measure actually works quite well in the survey on the recently arrived immigrants, but not for the established immigrants. Using this measure as the only predictor for education in an analysis thus introduces the risk of underestimating the effect of education.

As seen, the predictive powers of the different education measures are rather similar for the established and the recently arrived immigrants. Therefore, we decide by theoretical reasons which measure to include in the analysis in paper IV in which we estimate the impact of several indicators on immigrants’ German language proficiency. In this paper, we are not particularly interested in the signalling effect of a specific certificate or the effect of a single year spent in the educational system. Instead, the education variable is used as a proxy for immigrants’ cognitive ability and competencies and therefore we favour an education variable of a metric level of measurement. Moreover, we want to cover as much information on immigrants’ education as possible. Thus, we selected the index combining the years of schooling measure and the SOEP measure, which for both groups has a quite large predictive power (12.9% for the recently arrived immigrants and 4.7% for the established immigrants). As indicated, for many immigrants we do not have a CAMCES measure and to nevertheless include these cases in the analysis of paper IV, we decided on the index combing only two measures. However, the results of the validation indicate that the index has a higher predictive power of all tested education variables for the recently arrived immigrants. In comparison, for the established immigrants this index has the lower predictive power of all tested indices. This is due to the years of schooling variable, which has a very low predictive power for the established immigrants and by including this in the index, its predictive power also decreases. Concerning the interpretation of the education effect in the analysis of paper IV, this indicates that the effect of homeland education for the established immigrants might be slightly underestimated.

To sum up, it is good news that generic questionnaire items work well in a multivariate analysis involving immigrants' homeland education. Developing country-specific education instruments do not seem to be necessary for a wide range of study contexts. While this may seem disappointing after a lot of work was put into the development of the CAMCES tools, only this development made such a comparison possible in the first place. For survey research, it may thus be worth investing more efforts in testing and potentially improving generic education questions for migrant surveys, especially the identification of vocational education and training. One advantage of this approach is also that it can be implemented in telephone surveys in addition to personal interviews or web surveys. A disadvantage is that translation has to be handled very carefully, to make sure that the 'universal' meaning of the response categories remains intact across languages, which may be challenging.

6.6 References

Braun, M., & Müller, W. (1997). Measurement of Education in Comparative Research. *Comparative Social Research*, *16*, 163–201.

Briceno-Rosas, R., Liebau, E., Ortmanns, V., Pagel, L., & Schneider, S. L. (2018). *Documentation on ISCED Generation Using the CAMCES Tool in the IAB-SOEP Migration Samples M1/M2. SOEP Survey Papers 579: Series D. Berlin: DIW/SOEP*. Retrieved from https://www.diw.de/documents/publikationen/73/diw_01.c.608161.de/diw_ssp0579.pdf (accessed April 2020)

Friedberg, R. M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics*, *18*(2), 221–251. doi:10.1086/209957

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Nonlinear Principal Components Analysis: Introduction and Application. *Psychological Methods, 12*(3), 336–358. doi:10.1037/1082-989X.12.3.336

Meulman, J. J., van der Kooij, A. J., & Heiser, W. J. (2004). Principal Components Analysis With Nonlinear Optimal Scaling Transformations for Ordinal and Nominal Data. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 49–70). Thousand Oaks: SAGE Publications.

Schneider, S. L. (2008). Suggestions for the Cross-National Measurement of Educational Attainment: Refining the ISCED-97 and Improving Data Collection and Coding Procedures. In S. L. Schneider (Ed.), *The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries* (pp. 311–330). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES).

Schneider, S. L. (2010). Nominal Comparability is not Enough: (In-)equivalence of Construct Validity of Cross-National Measures of Educational Attainment in the European Social Survey. *Research in Social Stratification and Mobility*, *28*(3), 343–357. doi:10.1016/j.rssm.2010.03.001

Schneider, S. L. (2018). Das Bildungsniveau von Migrantinnen und Migranten: Herausforderung in Erfassung und Vergleich. In D. B. Maehler, A. Shajek, & H. U. Brinkmann (Eds.), *Diagnostik bei Migrantinnen und Migranten - Ein Handbuch* (pp. 47–56). Göttingen: Hogrefe Verlag.

Schneider, S. L., Briceno-Rosas, R., Herzing, J., & Ortmanns, V. (2018). *Computer-Assisted Measurement of Education in Surveys (CAMCES) - Final Project Report*. Mannheim.

Schröder, H., & Ganzeboom, H. B. G. (2014). Measuring and Modelling Level of Education in European Societies. *European Sociological Review, 30*(1), 119–136. doi:10.1093/esr/jct026

UNESCO-UIS. (2012). *International Standard Classification of Education 2011. Montreal: UNESCO-UIS*. Retrieved from http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf (accessed April 2020)

Weins, C. (2010). Kompetenzen Oder Zertifikate? Die Entwertung Ausländischer

Bildungsabschlüsse auf dem Schweizer Arbeitsmarkt. *Zeitschrift Fur Soziologie*, *39*(2), 124–139. doi:10.1515/zfsoz-2010-0203

Data Reference

Liebig, S., Goebel, J., Schröder, C., Schupp, J., Bartels, C., & Alexandra Fedorets, Andreas Franken, Marco Giesselmann, Markus Grabka, Jannes Jacobsen, Selin Kara, Peter Krause, Hannes Kröger, Martin Kroh, Maria Metzing, Janine Napieraj, Jana Nebelin, David Richter, Diana Schacht, Paul Schmelzer, Christian Schmitt, Daniel Schnitzlein, Rainer Siegers, Knut Wenzig, Stefan Zimmermann. (2019). *Socio-Economic Panel (SOEP), data from 1984-2017*. doi:10.5684/soep.v34