The classification of education in surveys: a generalized framework for ex-post harmonization

Schneider, Silke L.

Quality & Quantity (2022) 56:1829–1866
https://doi.org/10.1007/s11135-021-01101-1

The classification of education in surveys: a generalized framework for ex-post harmonization

Silke L. Schneider[1]

Accepted: 30 January 2021 / Published online: 5 April 2021

Abstract
All social science (and many other) surveys measure respondents' educational attainment. However, most of them do it in different ways, resulting in incoherent education variables across surveys. This complicates the cumulation of different datasets and hampers survey data reuse. For cross-national surveys that are designed to be comparative from the outset, methods for ensuring comparability in the measurement of education across countries have improved substantially over the last decades, relying on *ex-ante* output harmonization. For *ex-post* harmonization, the situation is more difficult because the data have already been collected, with education measures that only partly overlap in the amount and kind of information they store about respondents' education. This results in aggregated measures when harmonizing data ex-post. Such aggregated measures may underestimate associations with education in multivariate analyses, leading to biased results. They also do not allow testing hypotheses on the effects of specific types of education, such as vocational programs. This paper presents a new framework for harmonizing education variables ex-post, building on the International Standard Classification of Education (ISCED) and experience from cross-national surveys using ex-ante harmonization. It includes a new coding scheme called 'generalized ISCED' or GISCED, and extension variables standardizing aspects of education not covered by ISCED. It proposes solutions for problems that specifically occur in ex-post harmonization, for example source categories spanning ISCED levels. The paper also shows how to apply the GISCED framework to existing data. An empirical illustration shows how detailed harmonized education measures may give insights for research and policy not possible with more aggregate measures.

Keywords Classification · Education · Survey · Data harmonization · ISCED

This paper is a part of the Quality & Quantity Special Issue on Cross-National Survey Data Harmonization, guest edited by Ilona Wysmułek, Joonghyun Kwak and Irina Tomescu-Dubrow.

✉ Silke L. Schneider
silke.schneider@gesis.org

[1] GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

1 The harmonization of education in survey research

Surveys are an attractive data source because of their standardization, resulting from the application of a highly structured, uniform questionnaire in a standardized interview. Some processes and data, however, cannot be standardized by the questionnaire, survey operations or standardized interview administration, but require further standardization to be achieved with data processing. The application of statistical definitions and classifications for deriving nationally or even internationally comparable standard variables is one such area. In such an endeavor, there is a delicate balance between standardization and construct validity (Hauser 2016).

"[...] All efforts that standardize inputs and outputs in comparative surveys" are referred to as 'harmonization' (Granda and Blasczyk 2010, p.1). The measurement of educational attainment[1] in cross-national surveys requires harmonization procedures in data processing: Educational systems and qualifications differ strongly across countries, and the names of educational qualifications usually cannot be reliably translated because across languages, similar terms are used to denote different levels of education. There is no international terminology that would be equally well understood by respondents across countries.[2] It is thus not advisable to design one 'source' measurement instrument and translate this into all languages needed for the survey (so-called 'input harmonization') when measuring levels of education. Therefore, comparative research usually relies on country-specific measures of individual's educational attainment. The resulting country-specific variables are harmonized by recoding them into a common standard after data collection. This process is called 'output harmonization' (for further details, see e.g. Schneider et al. 2016; Schneider 2016).

For truly cross-national surveys, i.e. surveys that are "deliberately designed for comparative research" (Harkness et al. 2010, p. 3), education harmonization is already part of the survey design phase and not limited to data processing, which is why it is called *ex-ante* output harmonization (Ehling 2003; Granda et al. 2010; Granda and Blasczyk 2010). These surveys design the harmonized target variable(s) before developing the national questionnaires, and design measurement instruments in the different countries as well as coding rules for harmonization before data collection. In this way, they make sure that every kind of education intended to be differentiated and coded internationally will also be identified in the country-specific questionnaire items. Ex-ante output harmonization requires a certain degree of organizational capacity, but then can be realized successfully.

The situation is rather different when surveys are not designed to be comparable from the outset, but researchers wish to combine data from different surveys (sometimes much) later to be able to examine a specific research question that requires increasing the variation on the country level or over time (Dubrow and Tomescu-Dubrow 2016; Slomczynski and Tomescu-Dubrow 2018), or increasing sample size to study specific groups (Doiron et al. 2012). Such an undertaking requires *ex-post* data harmonization (Ehling 2003; Granda et al. 2010; Granda and Blasczyk 2010), i.e. the adjustment of data that are to be

[1] Educational attainment refers to the highest level of formal education achieved by a person. This paper thus only covers formal education and excludes the measurement and harmonization of non-formal education. Work on the comparative measurement and harmonization of the latter is still in its infancy and it is also not measured in many surveys.

[2] For example, 'primary education' refers to the first four to six years of education in most countries, but in some Eastern European countries, it refers to the first eight to ten years of education. These are internationally known as 'basic' education, which is classified as lower secondary rather than primary education.

pooled, resulting in a single integrated dataset with coherent target variables (Dubrow and Tomescu-Dubrow 2016). This applies both to the combination of surveys conducted in different countries, but also surveys conducted in just one country, but designed independently of each other (i.e. not part of a time series or panel design). Then, the data need to be made comparable after data collection by recoding variables relating to the same underlying concept but resulting from different measurement instruments into a common standard. Here, flexibility is highly limited by the information collected in the different surveys (Wolf et al. 2016), and harmonization is therefore difficult (Dubrow and Tomescu-Dubrow 2016). With respect to education, for example, education categories are more detailed or better documented in some datasets than in others. This makes ex-post harmonization more challenging—and the results often more limiting—than ex-ante harmonization.

Education is a core social background variable covered by all surveys and used in many statistical analyses, but very difficult to harmonize. To facilitate this task and support ex-post harmonization projects that want to maintain as much information as possible from the original data, this paper proposes a new harmonization framework for harmonized educational attainment variables as target variables in harmonization projects. It builds on the International Standard Classification of Education (ISCED) 2011 (UNESCO 2012, see Sect. 3), but extends it for usage in ex-post harmonization of survey data. The framework firstly includes a new coding scheme called 'Generalized International Standard Classification of Education', or GISCED. The framework also builds on experience from ex-ante output harmonization of education for comparative surveys, especially the European Social Survey (ESS), complementing the concepts underlying ISCED (and thus GISCED) to better represent strongly stratified educational systems, as they exist in many European countries. The framework therefore secondly proposes a set of 'extension variables' that allow researchers to operationalize additional concepts to what ISCED traditionally covers, such as the stratification of secondary education or distinction of different types of vocational education and training.

The paper will proceed in Sect. 2 with a brief overview of existing harmonization schemes for education, and a discussion of the advantages and disadvantages of detailed vs. aggregated measurement and coding. Section 3 then presents ISCED 2011 as the foundation for the proposed coding framework. Section 4 introduces common obstacles in ex-post harmonization of education data and proposes specific solutions, leading to the proposal of the GISCED framework, consisting of the GISCED coding scheme and a number of extension variables. Then, in Sect. 5, I will explain how to apply the new framework to existing national education variables, and how to derive its codes from some other international education coding schemes. Section 6 provides an empirical illustration using different education schemes as independent variables, looking at intergenerational educational inequalities. The final section summarizes the paper, discusses the results, and gives a brief outlook.

2 Existing schemes for ex-post harmonization of education

There are three basic approaches for output-harmonizing information on educational attainment: levels of education, scaled levels of education, and years of education. They will be briefly presented here.

2.1 Years of education

The first approach is to convert national education variables into the corresponding years of education (following Duncan and Hodge 1963). The ISCED mappings provided by UIS (UNESCO 2021) include the required information on the cumulative duration of educational programmes across countries. Sometimes years of education are also derived from harmonized categorical education variables. The ex-post harmonized International Social Mobility File (ISMF, Ganzeboom and Treiman 2019), only accessible on request, for example provides codes for derived 'virtual' years of education (Ganzeboom 2019). In this derivation, vocational education is penalized by not counting fully the related number of years of education so that the resulting variable better reflects how much general education a respondent has obtained. The resulting information is easy to include in linear statistical models, and is particularly popular among economists analyzing returns to education (Mincer 1974; Flabbi et al. 2008).[3]

2.2 Levels of education

Regarding the second approach, levels of education, various coding schemes are used. For national data from different sources, harmonization can be supported by national education classifications (which not all countries have), while for international data, an international classification is needed.

At the most simple level, data harmonized using levels of education distinguish just a few broad, ordinal categories such as 'less than primary', 'primary', 'secondary' and 'higher' (or 'tertiary') education, which are often not explicitly related to an international standard. For example, the IPUMS Demographic and Health Surveys (DHS, Boyle et al. 2019) and IPUMS International (Minnesota Population Center 2019) harmonize education into just four broad categories, in addition to years of education. What specifically constitutes these broad categories unfortunately differs across projects, which is due to different levels of development and educational expansion across countries. Furthermore, the specific content of broad categories is not always documented, and if the source variables were translated into English, the link with the educational system of the country in question becomes obscure. The exact contents of categories is also often unclear because there is no universal understanding of terms like 'secondary' or 'higher' education. Therefore, with data using just broad education levels for measuring and/or harmonizing education data, interoperability with other data, including harmonization with other sources, is severely limited, unless they use the same broad education categories. Some projects use standard aggregations of ISCED levels to produce three broad levels (e.g. Barone and Ruggera 2018), which are also commonly used in official reports concerning education across countries (e.g. OECD 2017).

At the most complex level, there are multi-digit coding schemes reflecting not just levels but also types of education within levels. The most common ones here are ISCED 2011 (UNESCO 2012, see further details in Sect. 3) and its predecessor ISCED 1997, developed for international official education statistics, and the CASMIN education scheme, which

[3] This approach is also used by some ex-ante harmonized surveys wishing to obtain a linear measure of education because there is a high level of measurement error in measures of 'years of education' or 'school-leaving age' when asked directly of respondents (Schröder and Ganzeboom 2014; Schneider 2010).

was developed in the academic project 'Comparative Analysis of Social Mobility in Industrial Nations' (König et al. 1988; Brauns et al. 2003). While the CASMIN scheme has been used mostly as a coding scheme for ex-post harmonization in selected European countries (Erikson and Goldthorpe 1992; Breen 2004; Breen et al. 2009), ISCED is by now widely used for ex-ante and ex-post harmonization (e.g. Barone and Ruggera 2018). For CASMIN, mappings linking national education categories with the international ones are only available for a limited set of countries. In the case of ISCED, such mappings are available for almost all countries in the world (UNESCO 2021) and in quite some detail specifically for Europe (Eurostat 2021).

To give some examples, the European Social Survey (ESS, 2020c), the European Values Study (EVS, 2020), the Survey of Health, Ageing and Retirement in Europe (SHARE, Börsch-Supan et al. 2013), the Programme for the International Assessment of Adult Competencies (PIAAC, OECD 2016), and the European Union Labour Force Surveys (EU-LFS, Eurostat 2020), harmonize education ex-ante and use fairly detailed categorical coding schemes based on ISCED. However, the specific ISCED-based schemes still differ from each other. The multi-digit education coding scheme used in the ESS since 2010, differentiating levels and various types of education within levels in one variable, called 'edulvlb', has fared quite well in empirical tests and was thus recommended for use in other comparative surveys (Schröder 2014, , chap. 3). It has since then gained some acceptance beyond the ESS community and has been implemented in the EVS 2017, SHARE, and—in slightly adapted form—for the upcoming cycle of OECD's PIAAC (Allen et al. 2017). Most of these datasets additionally have information on (actual or derived) years of education.

In between these two solutions using levels of education, there are datasets using a medium amount of differentiation, providing main education levels without differentiation of types of education. These main education levels are often defined with reference to ISCED, which is thereby being reduced to its 'bare bones' (Schröder 2014, p.76). The European Social Core variables project has specified ISCED main levels as the minimum requirement for official Eurostat surveys (Eurostat 2007). The Luxembourg Income Study (LIS, 2019) and the Survey Data Recycling project (SDR version1, Survey Data Harmonization Team 2017; Slomczynski and Tomescu-Dubrow 2018) use main ISCED levels, but the latter also provides additional information in separate so-called 'harmonization control' variables and thus in fact almost falls into the category of complex coding schemes: whether the respective education was not completed, whether it was vocational, and whether the category also includes higher levels of education. These harmonization control variables are intended to support researchers controlling for properties of the source survey items, and thus to enable researchers to 'recycle' rather than throw out data.

2.3 Scaling education

Finally turning to the third approach, scaled education variables, these are obtained by transforming categorical education variables into a linear metric using supporting variables. The ISMF project developed a linear education scale, constructed by Schröder and Ganzeboom (2014) through cause-and-effect-proportional scaling of national education variables—an approach building on prior work by Treiman (1975) and Smith and Garnier (1987)—, called International Standard Levels of Education (ISLED). The resulting scores

can in principle be applied to national education variables across surveys.[4] Given these national variables are usually not standardized within countries across surveys, this process is rather complex and introduces new errors.[5] ISLED scores can also be derived from the 'edulvlb' coding scheme used in the ESS (Schröder 2014, chap. 3),[6] as well as from 3-digit ISCED codes (Schröder 2014, chap. 5), making it more widely applicable but still relying on detailed harmonized measures of levels of education.

This summary of existing harmonization schemes reveals that both scaled education variables as well as derived years of education rely on harmonized measures of levels of education. The latter thus are not just important for data users who wish to analyze levels of education, but equally so for data users preferring other ways of analyzing educational attainment. However, as will be shown in Sect. 4, the existing international coding schemes for education are not easily applicable in ex-post harmonization, which motivates the development of the more generally applicable coding scheme GISCED.

2.4 Aggregate or detailed coding of education?

It can be debated whether survey data and harmonized datasets should rather provide detailed or aggregate education variables. Aggregate variables contain information on many different educational qualifications in very few summary categories, which often carry a more vague meaning. Detailed variables in contrast contain many categories that are more specific. An argument in favor of aggregate data is that mis-classifications will be less common than when providing detailed variables: some 'noise' will disappear due to the aggregation. Producing aggregated variables is also arguably less labor intensive than producing detailed variables. A lot of public reporting (see e.g. OECD's 'Education at a Glance'-series[7].) is also limited to aggregated measures, and it is easier to communicate about broad distinctions in public debate.

Cross-national survey data involving many European countries, however, notably chose the more complex coding schemes, because European educational systems tend to be more complex than the US-American system (Shavit and Müller 1998) or education systems in most low and middle income countries. When using years of education to measure educational attainment in complex educational systems, there is a concern that important distinctions between people who have followed different educational pathways with the same number of years of education would not be measured, leading to hidden inequalities and low validity of the data (Braun and Müller 1997; Schneider 2010; Kerckhoff and Dylan 1999; Kerckhoff et al. 2002; Schröder 2014). The same criticism may be applied to projects using a low number of broad education categories, and research shows that types of education within levels can make a difference for various outcomes (Triventi 2013a; van

[4] However, the author only found the conversion files for ISLED for the ESS on the ISMF website, not for other surveys included in the ISMF.

[5] See e.g. its application in the Generations and Gender Survey, GGS: "The descriptions of the educational levels in the GGS datasets do not always completely match with the descriptions of the educational levels in the tables of Schröder (2014) as derived from the ESS data. We always used the ESS ISLED scores from which the descriptions come closest to the descriptions of the GGS, which sometimes results in taken the average of multiple ISLED scores" (Brons and Mooyaart 2018, p. 6).

[6] For the coding information, see http://www.harryganzeboom.nl/isled/edulvlb_isled_R56.txt. This coding can be applied to later ESS rounds as well.

[7] https://www.oecd.org/education/education-at-a-glance/.

de Werfhorst 2017; Delaruelle et al. 2020). We can only study the effects of extreme disadvantage by adequately representing the periphery of the socio-structural concepts we are interested in. Avoiding loss of information, be it by losing entire cases because of incomplete data, by aggregating categories, or transforming a categorical measure of educational qualifications into hypothetical years of education, is therefore an important guiding principle of output harmonization (whether ex-ante or ex-post).

So, while both years of education and broad education categories are commonly used and may be perfectly adequate for data analysis in many research projects, they may be deficient as the sole harmonized education variable(s) for projects aiming for a high level of data reuse, since important categorical distinctions cannot be derived from years of education or broad education levels. Then, educational effects and inequalities may be underestimated, and hypotheses concerned with unveiling such inequalities cannot be tested. In contrast, when using a detailed categorical coding scheme, it is possible to derive years of education from these categories, or simplify the detailed scheme into broad categories in whichever way needed for a specific set of countries or a specific research question. Even for deriving years of education or linear education scores, the amount of detail covered in the source variables is crucial (Schröder and Ganzeboom 2014). Such derivations are not possible the other way round. The degree of 'interoperability' and 're-usability', core principles of FAIR data (Wilkinson et al. 2016), are thus highest when applying a detailed categorical scheme for harmonization, or (which is in terms of information content equivalent), harmonizing information into a whole set of harmonized variables. In contrast, especially when combining data from countries at different stages of development and different time points, the problem of harmonization becomes rather severe when only broad levels are available in the data, and the boundaries between the broad levels do not match across datasets. The GISCED framework therefore follows the strategy of coding at a high level of detail.

3 The international standard classification of education (ISCED) 2011

ISCED 2011 (UNESCO 2012; Schneider 2013) is the international statistical classification for education-related data and maintained by UNESCO Institute for Statistics in Montreal, Canada. Like all classifications, its aim is to "[...] group and organize information meaningfully and systematically, usually in exhaustive and structured sets of categories that are defined according to a set of criteria for similarity", and "to provide a simplification of the real world" (Hancock 2013, p. 3).

In contrast to its predecessor ISCED 1997 (UNESCO 2006), ISCED 2011 provides a three-digit numerical coding system for the sub-classification of levels of education (there is also a sub-classification for fields of education). There are two variants of the sub-classification of levels of education, one for the classification of educational *programs* (ISCED-P) and on for the classification of educational *attainment* (ISCED-A). Since surveys are mostly concerned with measuring the latter, ISCED-A forms the backbone of the proposed GISCED.[8] This section describe ISCED-A digit by digit. Table 1 gives an overview of all ISCED categories. The information on which national educational program and

[8] The difference is important in that sometimes a program at a specific level is considered too short to complete that level, so that the completion of this program is classified at the next lower level.

Table 1 ISCED 2011 codes and labels for educational attainment

ISCED-A	1st digit: main level		2nd digit: sub-level and orientation		3rd digit: level completion and access to higher level	
010	0	Less than primary	1	Never attended an education programme	0	Not further specified
020	0	Less than primary	2	Some early childhood education	0	Not further specified
030	0	Less than primary	3	Some primary education (no completion)	0	Not further specified
100	1	Primary	0	Not further specified	0	Not further specified
242	2	Lower secondary	4	General	2	Partial level completion, without direct access to upper secondary
243	2	Lower secondary	4	General	3	Level completion, without direct access to upper secondary
244	2	Lower secondary	4	General	4	Level completion, with direct access to upper secondary
252	2	Lower secondary	5	Vocational	2	Partial level completion, without direct access to upper secondary
253	2	Lower secondary	5	Vocational	3	Level completion, without direct access to upper secondary
254	2	Lower secondary	5	Vocational	4	Level completion, with direct access to upper secondary
342	3	Upper secondary	4	General	2	Partial level completion, without direct access to tertiary
343	3	Upper secondary	4	General	3	Level completion, without direct access to tertiary
344	3	Upper secondary	4	General	4	Level completion, with direct access to tertiary
352	3	Upper secondary	5	Vocational	2	Partial level completion, without direct access to tertiary
353	3	Upper secondary	5	Vocational	3	Level completion, without direct access to tertiary
354	3	Upper secondary	5	Vocational	4	Level completion, with direct access to tertiary
443	4	Post-secondary non-tertiary	4	General	3	Level completion, without direct access to tertiary
444	4	Post-secondary non-tertiary	4	General	4	Level completion, with direct access to tertiary
453	4	Post-secondary non-tertiary	5	Vocational	3	Level completion, without direct access to tertiary
454	4	Post-secondary non-tertiary	5	Vocational	4	Level completion, with direct access to tertiary
540	5	Short cycle tertiary	4	General	0	Not further specified
550	5	Short cycle tertiary	5	Vocational	0	Not further specified
560	5	Short cycle tertiary	6	Orientation unspecified	0	Not further specified
640	6	Bachelor's or equivalent level	4	Academic	0	Not further specified
650	6	Bachelor's or equivalent level	5	Professional	0	Not further specified
660	6	Bachelor's or equivalent level	6	Orientation unspecified	0	Not further specified

Table 1 (continued)

ISCED-A	1st digit: main level		2nd digit: sub-level and orientation		3rd digit: level completion and access to higher level	
740	7	Master's or equivalent level	4	Academic	0	Not further specified
750	7	Master's or equivalent level	5	Professional	0	Not further specified
760	7	Master's or equivalent level	6	Orientation unspecified	0	Not further specified
840	8	Doctoral or equivalent level	4	Academic	0	Not further specified
850	8	Doctoral or equivalent level	5	Professional	0	Not further specified
860	8	Doctoral or equivalent level	6	Orientation unspecified	0	Not further specified
999	9	Not elsewhere classified	9	Not elsewhere classified	9	Not elsewhere classified

qualification fulfills which criteria and is assigned to which ISCED code is found in the official ISCED mappings (Eurostat 2021; UNESCO 2021).

3.1 The first digit: ISCED main levels

The first digit of ISCED consists in an *ordered* set of categories going from no to the highest possible level of education. ISCED level 0 is used for individuals who have not completed primary education, which is defined as ISCED level 1. ISCED level 1 provides pupils with fundamental skills in reading, writing and arithmetic in four to six years. ISCED level 2 or lower secondary education is an intermediate level in which a broad range of subjects is being taught until approximately age 15 or 16. Completion of ISCED 2 is often regarded as the absolute minimum level of education, with some countries using the term 'primary education' for this level, which may lead to some confusion regarding the classification of such programs. While most countries offer one educational program for all students in ISCED 2, some countries already start sorting students into different programs by ability at the start of ISCED 2.

In upper secondary education (ISCED level 3), subject specialization increases substantially, especially in vocationally oriented programs. Program choice and achievements at this level predetermine to a large extent whether an individual will gain access and be able to successfully complete higher education. ISCED level 3 in most countries ends at age 18 or 19, after about 12 years of schooling. All individuals pass through ISCED levels 1 to 3, if they do not stop their education before the end of upper secondary education. ISCED level 4, post-secondary non-tertiary education, in contrast, is an 'optional' level. It mostly contains educational programs that allow individuals to change the specialization they had at level 3, in order to either gain access to tertiary education programs they would otherwise not be admitted to, or get vocational training to then enter the labor market. It also contains programs that may nationally be considered as 'tertiary', but that last less than two years and thus do not fulfill the requirement for classification in tertiary education.

ISCED level 5 is the first out of four levels summarized as 'tertiary education'. It contains all programs leading to qualifications below the level of a Bachelor's degree but that take at least two years of full-time education to complete. These are mostly vocational programs.[9] Higher education 'proper' starts with ISCED level 6, the Bachelor's level. Next to the prototypical Bachelor's degrees, it is also used for other qualifications in higher education that last 3 to 4 years since the end of upper secondary education, such as polytechnic diplomas. ISCED level 7, the Master's level, contains not just Master's degrees and other post-graduate qualifications, but also qualifications from long (5 to 6 years) first degree programs, which were common in many countries before the onset of the Bologna reforms and in many countries still exist in certain subject areas such as medicine or law. ISCED level 8 finally is reserved for doctoral programs. Code 9 is foreseen for education that cannot be classified in any of these levels.

These levels are also referred to as 'main' levels. Very often, nothing else is actually used, or not even that: official data often only use the three broad reporting categories 'low' (up to ISCED 2), 'medium' (ISCED 3 and 4) and 'high' education (ISCED 5 and upwards).

[9] In some countries, programs classified in ISCED 5 do not require a higher education entrance qualification so that people would often not consider them as higher education. It is therefore important not to confuse the terms 'tertiary' and 'higher' education when referring to ISCED.

However, since depending on the country, many respondents may accumulate in one or a few ISCED levels despite heterogeneous educational experiences, and specific types of education (such as vocational vs. general) are also of theoretical interest in research, it is worthwhile to also consider the next two digits of ISCED in data coding and analysis. Especially for data producers that provide harmonized data for researchers with diverging measurement requirements, a more detailed than first digit coding is advisable.

3.2 The 2nd digit: program orientation

The second digit has a different meaning depending on the level of education we are talking about. For ISCED 0, the second digit distinguishes (0) 'no education' (at all) from (1) 'some pre-primary' and (2) 'some primary education'—also a new feature of ISCED 2011. At ISCED levels 2 to 5, it distinguishes between (4) general and (5) vocational programs, without any ordering implied. A vocational orientation applies if a program prepares for a specific occupation or class of occupations. All other programs are considered to be 'general'. At ISCED levels 6 to 8, the terms 'academic' and 'professional' are sometimes used in a similar vein. However, since the distinction of programs preparing for a specific occupation is not of major importance here, and internationally agreed definitions for distinctions within higher education lacking, the second digit is rarely used in official data at these levels. This is why ISCED provides code (6) 'orientation unspecified' specifically for these levels.

3.3 The 3rd digit: level completion and access to a higher level

The third digit of ISCED 2011 for attainment (ISCED-A) is only defined for levels 2, 3 and 4, and identifies whether the educational program (at least partially) completes the ISCED level (code 2). This is usually not the case for programs of very short duration of e.g. just one year (which get code 1 on ISCED-P and are not considered for classification in ISCED-A). If the program in question completes the level, ISCED distinguishes whether it provides access to a higher level of education (code 4) or not (code 3). These codes are thus ordered and can be regarded as sub-levels. At ISCED levels 3 and 4, access refers to either of levels 5, 6 or 7. So if a vocational program at level 3 only gives access to a program at level 5 but not 6 or 7, it will be classified as providing access, even though it will likely be at a lower standard than a general program providing access to levels 6 and 7.

For the second and third digit of ISCED, code 0 is used when these digits are not specified at the respective ISCED level. For example, at ISCED 1, there is no distinction of vocational and general education and sub-levels don't apply either, and at ISCED 0, 1, and in tertiary education, all programs are assumed to complete the respective level and to give access to a higher level.

4 Extending the ISCED 2011 code scheme for ex-post harmonization

ISCED was developed for administrative rather than survey data. Only since the adoption of ISCED 2011, the classification offers some features specifically geared to measurement of education in surveys, distinguishing the classification of educational programs

(ISCED-P, for indicators related to enrollment) and educational qualifications (ISCED-A, for indicators related to educational attainment), as well as providing a 3-digit coding scheme. If data using this standard coding scheme were more widely available, ISCED could become a very useful tool in the standard coding of education in surveys, much like the International Standard Classification of Occupations (ISCO, International Labour Organisation 2007) for the derivation of occupation-related variables such as social status (Ganzeboom et al. 1992) or social class (e.g. Oesch 2006; Rose and Harrison 2010; Erikson et al. 1979).

Surveys using ex-ante output harmonization like the ESS can take the requirements of ISCED into account when developing measurement instruments. This is not the case with existing survey data on educational attainment that a researcher may want to harmonize ex-post, where measurement instruments and categories differ across countries (because of the different educational systems) and even across surveys within countries. Many education categories in actual surveys nevertheless can directly be coded into ISCED-A using the ISCED mappings (Eurostat 2021; UNESCO 2021). However, a few issues will repeatedly appear in any ex-post harmonization project:[10]

1. What if the ISCED main level of education is not identified in the source education variables, e.g. because an education category spans two ISCED levels?
2. What if information on program orientation, level completion and/or access to a higher ISCED level is unknown?
3. What if there is some potentially important piece of information covered in the source education variables, that is however not covered by ISCED, and that the harmonization project wants to keep visible in the harmonized data?

In order to offer standard rules and codes for ex-post harmonization taking these issues into account, this paper presents a generalized framework for the classification of education in surveys. The numerical code scheme underlying this framework is in short called GISCED, i.e. the *Generalized* International Standard Classification of Education. It is complemented by a set of separate extension variables that capture information not covered by ISCED in a standardized way. Both GISCED and the extension variables can be adapted to the requirements of a specific project with regards to the level of detail covered. This section describes the GISCED framework and classification rules in some detail.

A general idea when developing GISCED is that ISCED 2011 codes can be derived by just dropping newly specified digits, in order to avoid the need for complex recoding and to keep interoperability with data using official ISCED codes high. In this sense, ISCED forms the 'heart' of GISCED. The other general idea is that the framework should allow harmonization with minimal loss of information, also if the source education variables are richer in information than ISCED. Therefore, the generalized framework does not just offer new 'unspecified' categories at each ISCED digit, but also introduces new digits and so-called 'extension' variables to carry information not covered by ISCED but potentially relevant in comparative research.

[10] Granted, there are also other problems that hinder the harmonization of education variables across surveys, foremost lacking or low-quality documentation and processing error, leading to problems in linking source education categories with ISCED to start with. These problems cannot be dealt with at the level of classifications and are thus not dealt with in this paper. They require the improvement of documentation.

4.1 Categories that span multiple main ISCED levels

A common harmonization problem is that sometimes, education categories in questionnaires do not correspond to a single main ISCED level but include a mix of levels (e.g. Luxembourg Income Study n.d.). For example, depending on compulsory schooling legislation in a country, countries differ in how much they differentiate education at very low levels: in the first four rounds of the ESS, Austria for example only distinguished completion of compulsory school, corresponding to ISCED level 2, and no educational qualification. As an effect, ISCED 0 (less than primary education) and ISCED 1 (primary education) cannot be distinguished for Austria in ESS rounds 1 to 4. While ISCED offers code 9 to classify educational programs and qualifications 'not elsewhere classified', this would result in severe loss of information since we *do* know that the respective category aggregates ISCED levels 0 and 1.

Therefore, another digit is added *after* the ISCED main level, which indicates the (included) upper bound of a category, while the main level indicates the (included) lower bound.[11] If there is no range, the code for the applicable level would simply be repeated. Code 9 on the first (and/or second) digit of GISCED should thus only be used if the applicable main ISCED levels are absolutely unknown. Table 2 shows how the first two digits of GISCED would then look like, in relationship to ISCED main levels.[12]

This leads to a classification where categories are not entirely mutually exclusive, because the 'spanning' categories include the non-spanning categories, which is usually undesirable (Hancock 2013). However, this is exactly needed when survey data are not collected or harmonized with the ISCED main levels as a coding target in mind. Education categories that can be coded into a non-spanning category should always be coded there and spanning categories only be used if otherwise missing data would be produced.

For harmonization projects for which the supplementary dimensions of ISCED—i.e. all information beyond main level—is not relevant, this scheme that turns the first ISCED digit into two digits may be entirely sufficient. It will be referred to as 'GISCED2'. Codes 02 (low), 34 (medium), and 58 (high) would represent the three broad ISCED levels commonly used in statistical reporting. GISCED2 can also be used to identify specific binary distinctions often used in the analysis of educational transitions, such as less than upper secondary (02) vs. upper secondary or more (38), or less than Bachelor's level (05) vs. Bachelor's level or more (68). For more demanding harmonization projects, ISCED second digit codes can be appended to these codes, too, e.g. when an education category clearly identifies vocational education, but mixes vocational education at two ISCED levels.[13]

GISCED2 can be approximately transformed into ISCED main levels by dropping the second digit. With this solution, there will be some systematic underestimation of education when simplifying data using the lower bound only, i.e. ignoring the upper bound.

[11] An alternative would be to use a flag variable indicating that the category includes higher levels of education for specific cases, like the harmonization controls in the SDR project (Survey Data Harmonization Team 2017). A disadvantage of this approach is that the precise levels included in the code are not clear for the user.

[12] Some possible codes are not shown because they will be of limited practical use (e.g. 26 - lower secondary to Bachelor's level, or 08 - no education to the doctoral level.)

[13] ISCED third digit codes referring to level completion and access should probably be regarded as unspecified for categories spanning levels (other than 3 and 4, where access refers to ISCED 5 for both) since this information is bound to identification of the precise ISCED level.

However, this is often preferable to the alternative of producing missing data. Sometimes a level between the lower and upper bound may be a more adequate simplification.

4.2 Unknown information for the second and third digit of ISCED

Sometimes we do not know whether a qualification is vocational or general because of incomplete information, or a response category contains both. For ISCED 0, the distinctions foreseen by ISCED of whether any pre-primary or primary education was attended are hardly ever available in survey data (although for developing countries, sometimes 'some primary education' is separately measured). In a similar vein, we may not know whether a qualification or program gives access to a higher level of education or not—or a response category mixes these different categories.

There are several possible solutions to this problem. Firstly, if we happen to know that one of the sub-categories is clearly dominant in the respective education category, i.e. much more common than the other, the respective sub-category could be coded, disregarding the measurement error for those respondents for whom another sub-category would be more adequate. Secondly, a specific code could be used to signify that we do not have this piece of information. This would be advisable if no specific sub-category dominates the education category found in the questionnaire. Following the solution for ISCED main levels, we will use code 9 across all digits and levels to identify unknown further specifications.[14]

Tables 3 and 4 show the codes available for the third and fourth digit of GISCED, corresponding to the second and third digit of ISCED. The only difference compared to the official ISCED coding (UNESCO 2012, Tables 2 and 3, p. 21–22) is then the changed label of code 9. Table 5 shows some exemplary GISCED codes and their labels to illustrate how the different digits come together, and how the lower and upper bounds (first and second digit) can be used to flexibly classify education categories that span ISCED levels.

4.3 Avoiding loss of information present in source education measures

There are (at least) five situations in which source education categories can be more differentiated than ISCED (see also Schneider and Kogan 2008): the existence of 1. external stratification or 'tracking' in secondary and 2. higher education, 3. different types of vocational education, 4. short higher level programs not considered for level completion, and 5. specific categories for incomplete education (dropout). In such situations, harmonizing data using ISCED leads to information loss. The GISCED framework therefore includes separate 'extension' variables, in order to allow this information to be retained in the harmonized data, without changing the GISCED code itself.[15] They carry substantive

[14] If the main level of education is unknown, including cases in which an education category spans several ISCED levels, sub-level, program orientation, level completion and access to a higher education level is regarded as unspecified rather than unknown in most cases since this information is often level-specific.

[15] The extension variables suggested here have some similarity with 'harmonization control variables' used in the Survey Data Recycling project (Survey Data Harmonization Team 2017), and can be used similarly in empirical analyses. However, the focus of extension variables is extra substantive information not covered by ISCED, therefore extending it conceptually, rather than documenting compromises that had to be made during the harmonization process given inconsistent source education variables. Therefore, harmonization controls can be additionally coded on aspects not covered by the GISCED framework, e.g. when a source measures does not cover vocational education.

Table 2 First two digits of GISCED: level codes including categories spanning ISCED main levels

Code	Lower bound	Upper bound	Label
00	=ISCED 0		Less than primary
01	ISCED 0	ISCED 1	Primary and less
02	ISCED 0	ISCED 2	Lower secondary and less (=low)
03	ISCED 0	ISCED 3	Upper secondary and less
04	ISCED 0	ISCED 4	Post-secondary non-tertiary and less
05	ISCED 0	ISCED 5	Less than Bachelor's level
06	ISCED 0	ISCED 6	Less than Master's level
07	ISCED 0	ISCED 7	Less than doctoral level
11	=ISCED 1		Primary
12	ISCED 1	ISCED 2	Primary and lower secondary
13	ISCED 1	ISCED 3	Primary to upper secondary
14	ISCED 1	ISCED 4	Primary to post-secondary
18	ISCED 1	ISCED 8	Any formal education
22	=ISCED 2		Lower secondary
23	ISCED 2	ISCED 3	Lower and upper secondary
28	ISCED 2	ISCED 8	More than primary
33	=ISCED 3		Upper secondary
34	ISCED 3	ISCED 4	Upper and post-secondary non-tertiary (=medium)
38	ISCED 3	ISCED 8	More than lower secondary
44	=ISCED 4		Post-secondary non-tertiary
45	ISCED 4	ISCED 5	Post-secondary and short cycle tertiary
46	ISCED 4	ISCED 6	Post-secondary to bachelor's level
47	ISCED 4	ISCED 7	Post-secondary to master's level
48	ISCED 4	ISCED 8	more than upper secondary
55	=ISCED 5		Short cycle tertiary
56	ISCED 5	ISCED 6	Short cycle tertiary and bachelor's level
57	ISCED 5	ISCED 7	Short cycle tertiary to master's level
58	ISCED 5	ISCED 8	Tertiary (=high)
66	=ISCED 6		Bachelor's level
67	ISCED 6	ISCED 7	Bachelors and master's level
68	ISCED 6	ISCED 8	Bachelor's level and higher
77	=ISCED 7		Master's level
78	ISCED 7	ISCED 8	Master's level and higher
88	=ISCED 8		Doctoral level
99	Unknown	Unknown	Unknown

information that goes beyond information harmonized in ISCED. These extensions will be useful for labor market and social stratification research, but are less relevant when education is just a background variable. Data users can use these variables to custom-build harmonized education variables that include the respective kind of information. More specifically, the extension variable 'edustrat' (see Sects. 4.3.1 and 4.3.2, and Table 6) allows researchers to complement the GISCED scheme with information about stratification in secondary and higher education. The extension variable 'vettype' (see Sect. 4.3.3 and

Table 3 Codes for the third digit of GISCED, corresponding to the second digit of ISCED-A: sub-level and orientation

Code	Concept	Label	Relevant for…
0		Not further specified	ISCED 1
1	Sub-level	Never attended an education program	ISCED 0
2		Some early childhood education	ISCED 0
3		Some primary education (no completion)	ISCED 0
4	Orientation	General / academic	ISCED 2–8
5		Vocational / professional	ISCED 2–8
6		Orientation unspecified	ISCED 6–8
9		Sub-level/orientation unknown	

Table 4 Codes for the fourth digit of GISCED, corresponding to the third digit of ISCED-A: level completion and access to higher level

Code	Label	Relevant for…
0	Not further specified	ISCED 0, 1, 5–8
2	Partial level completion without access to higher level	ISCED 2–3
3	Level completion without access to higher level	ISCED 2–4
4	Level completion with access to higher level	ISCED 2–4
9	Level completion and/or access unknown	

Table 7) complements GISCED by adding information about different types of vocational education and training.

4.3.1 Stratification in secondary education

While being comprehensive in most countries, some countries track students into different programs or school types already in lower secondary education. In these countries, some general education programs do not give access to all (and especially not to academically selective) programs at a higher level but only to vocational programs. Enrollment in such a pre-vocational program is consequential for individuals' educational careers and labor market outcomes (Bol and Van de Werfhorst 2011; Bol et al. 2014; van de Werfhorst and Mijs 2010; Allmendinger 1989), but not identifiable using ISCED. At the upper secondary level, different general tracks and resulting qualifications—which exist more rarely at this level—give unequal access to different types of higher education (see section 4.3.2). This is why the ESS opted for a more fine-grained measure of education, distinguishing whether a secondary qualification provides access to general or academic education at the higher level or not on the third (access) digit. The ESS education coding scheme contains this information at ISCED levels 2, 3 and 4. To keep the ISCED code intact, it is suggested here to code an extension variable for ISCED levels 2 – 4 using the following codes:

1. = track not giving access to any higher level education;
2. = track giving access only to vocational, professional or lower tier higher level education (i.e. limited access);

Table 5 Illustration of selected GISCED codes across all ISCED levels

Code	Description	Lower and	Upper bound	Orientation	Completion/access
0100	Less than lower secondary (=less than primary or primary)	ISCED 0	ISCED 1	0 = Not specified	0 = Not specified
0200	Less than upper secondary (=low)	ISCED 0	ISCED 2	0 = Not specified	0 = Not specified
0400	Less than tertiary (=low and medium)	ISCED 0	ISCED 4	0 = Not specified	0 = Not specified
1100	Primary education	ISCED 1	ISCED 1	0 = Not specified	0 = Not specified
1240	General schooling below upper secondary	ISCED 1	ISCED 2	4 = General / Academic	0 = Not specified
2244	General lower secondary, level completion with access to ISCED 3	ISCED 2	ISCED 2	4 = General / Academic	4 = Level completion with access to higher level
2253	Vocational lower secondary, level completion without access to ISCED 3	ISCED 2	ISCED 2	5 = Vocational / Professional	3 = Level completion without access to higher level
2340	General lower or upper secondary	ISCED 2	ISCED 3	4 = General / Academic	0 = Not specified
3353	Vocational upper secondary, level completion without access to tertiary	ISCED 3	ISCED 3	5 = Vocational / Professional	3 = Level completion without access to higher level
3344	General upper secondary, level completion with access to tertiary	ISCED 3	ISCED 3	4 = General / Academic	4 = Level completion with access to higher level
3444	General upper secondary or post-secondary non-tertiary, level completion with access to tertiary	ISCED 3	ISCED 4	4 = General / Academic	4 = Level completion with access to higher level
3499	Upper secondary or post-secondary non-tertiary (=medium)	ISCED 3	ISCED 4	9 = Sub-level/orientation unknown	9 = Level completion and/or access unknown
4453	Vocational post-secondary non-tertiary, level completion without access to tertiary	ISCED 4	ISCED 4	5 = Vocational / Professional	3 = Level completion without access to higher level
4890	Post-secondary and tertiary	ISCED 4	ISCED 8	9 = Sub-level/Orientation unknown	0 = Not specified
5550	Vocational short cycle tertiary	ISCED 5	ISCED 5	5 = Vocational / Professional	0 = Not specified
5840	Academic tertiary	ISCED 5	ISCED 8	4 = General / Academic	0 = Not specified
5890	Tertiary (=high)	ISCED 5	ISCED 8	9 = Sub-level/Orientation unknown	0 = Not specified
5640	Academic tertiary below Master's level	ISCED 5	ISCED 6	4 = General / Academic	0 = Not specified
7840	Academic Master's or higher level	ISCED 7	ISCED 8	4 = General / Academic	0 = Not specified
8860	Doctoral level, orientation unspecified	ISCED 8	ISCED 8	6 = Orientation unspecified	0 = Not specified

Table 6 Codes for the extension variable 'edustrat', reflecting stratification in educational systems

Code	Concept	Label	Relevant for…
0		Stratification of education not specified	0–1, 5
1	Stratification in secondary education	No access to any higher level education	2–4
2		Access only to vocational, professional or lower tier higher level education (i.e. limited access)	2–4
3		Access to academic, university or upper tier or all types of higher level education, including non-tracked (comprehensive) programs (i.e. full access)	2–4
4	Stratification in higher education	Lower tier (non-selective, polytechnic, applied, and other non-university institutions)	5–8
5		Upper or single tier (selective, traditional, research oriented universities awarding doctoral degrees)	5–8
9		Track in secondary or tier in higher education unknown	

Table 7 Codes for the extension variable 'vettype', reflecting different types of vocational education and training (VET)

Code	Label	Relevant for…
0	Type of VET not further specified	ISCED 0–1, 8
1	School-based VET (with full-time vocational schooling)	ISCED 2–7
2	'Dual' VET (combining in-company training and part-time vocational schooling)	ISCED 2–7
3	Work-based VET (without vocational schooling)	ISCED 2–4
9	Type of VET unknown	

3. = track giving access to academic, university or upper tier or all types of higher level education, including non-tracked (comprehensive) programs (i.e. full access).

These codes of the extension variable 'edustrat' are consistent with the third digit of the 'edulvlb' coding scheme used in the ESS (see section 5.2).

4.3.2 Stratification in higher education

Vertical stratification in tertiary and higher education is well captured in ISCED by the distinction of four tertiary education levels. *Horizontal* stratification in higher education takes on different forms in different countries though, hampering comparative measurement (Marginson 2016; Triventi 2013b, a). Again, this institutional differentiation is expected to be related to social background (Triventi 2013a) and to have consequences for individual's educational and labor market outcomes (Shavit et al. 2007). Educational expansion then does not necessarily lead to equalization of opportunities because it may generate new inequalities through institutional differentiation (Lucas 2001). In countries with little institutional diversification in higher education, field of study is a good indicator of qualitative (horizontal) differences

in education, but it is only rarely measured. In countries with a diversified higher education system, at the most basic level, there is often a differentiation between traditional or elite universities and institutions without university or 'elite' status (Bourdieu 1988).

One way of operationalizing this idea is that the former are more academic and research-oriented and are thus the prime (or sole) institutions that award PhDs. The latter are more practically-oriented and usually do not award PhDs. This differentiation is not made in ISCED though. The ESS at levels 6 and 7 therefore distinguishes (commonly less selective and more professional) qualifications from polytechnics, universities of applied science and lower tier colleges from (commonly more selective and academic) traditional university degrees. The ESS education coding scheme contains this information on the second digit (1 signifying lower tier/non-university programs, and 2 signifying higher or single tier/traditional university programs). The variable 'edustrat' already used for stratification in secondary education is therefore extended to higher education, using the distinctions already made in the ESS:

4. lower tier (non-selective, polytechnic, applied, and other non-university institutions);
5. upper tier (selective, traditional, research oriented universities awarding doctoral degrees) or single tier.

For ISCED 5 qualifications that are not part of the higher education but rather vocational education and training system in a country, edustrat should be coded 0 (e.g. master crafts in Germany). For ISCED 5 qualifications awarded at universities or polytechnics, it should be coded in the respective higher education tier.

Code 9 on this extension variable can again be used if the information on track in secondary or tier in higher education is not available or the different types are mixed in a single education category. Code 0 can be used for cases for which this extension variable is not applicable (e.g. less primary education).

4.3.3 Different vocational education and training types

Vocational education and training (VET) is organized differently across countries (Dieckhoff 2008). ISCED does not distinguish between school-based, work-based or the 'dual system' of combined school- and work-based VET. However, when comparatively studying skill production, success in higher education or labor market outcomes, the institutional context may make a difference (e.g. Hanushek et al. 2017; Forster et al. 2016; Saar et al. 2017): different types of VET relate to different degrees of employer involvement, occupational specificity, labor market linkage, standardization, skill transparency, state regulation, and development of general skills (Shavit and Müller 1998; Dieckhoff 2008; Bol and Van de Werfhorst 2016; Andersen and van de Werfhorst 2010). The organization of VET is also relevant for education and labor market policy. Due to lacking data, this is often only measured at the aggregate level, but often, various types of VET are available in one country. For projects interested in respondent level effects of different types of VET, a separate variable 'vettype' is introduced to the GISCED framework, distinguishing

1. School-based VET (with full-time vocational schooling);
2. 'Dual' VET (combining in-company training and part-time vocational schooling);
3. Work-based VET (without vocational schooling). This latter type may be dubious with respect to its recognition as formal education, and could be used to downgrade the

affected respondents' attainment level if education is intended to e.g. proxy general skills in the project in question.

This variable needs to be specified separately from the stratification variable introduced previously because both apply to the same levels and categories of education. Since some countries have started offering 'dual' VET in tertiary education (see e.g. degree level apprenticeships in the UK and Berufsakademie or 'duale Hochschule' in Germany), this variable is not only relevant for secondary but also tertiary education. Code 9 can again be used for cases for which the information is not available or mixed in a single education category, and category 0 for cases for which type of VET is not relevant (e.g. primary education).

4.3.4 Short higher level programs

Educational programs that are too short to complete an ISCED level, a situation that only occurs at ISCED levels 2 and 3, are classified as 2X1 and 3X1 respectively in ISCED-P. In ISCED-A, which is more relevant for survey research, qualifications from such a program would be classified as 100 and 2X4 respectively, which usually corresponds to the code of the qualification completed *prior* to entering the short program. Therefore, data coded using ISCED omit the information that a short higher-level program was completed. I thus firstly suggest to include the orientation of the higher level program in the coding of attainment at the lower level.[16] Secondly, a flag variable should be added to signal that in addition to the highest level of attainment, a short program from a higher level was also completed.[17] Alternatively, if the extension variable on educational stratification 'edustrat' has been coded, code 6 on that variable, so far empty, can also be used. This is useful because the track of the previously attended education would often be unknown for respondents who completed such short higher level programs anyway, so they would effectively be recoded from code 9 to 6 on 'edustrat'.

4.3.5 Dropout categories

Some countries employ education categories reflecting that an educational program was not successfully completed in some surveys. However, then the actually highest level of education completed is not known. It is thus usually advised not to use such categories. If faced with this information in ex-post harmonization, it is not clear how to code it in ISCED. It is suggested here to try to infer the highest successfully *completed* program and code this as the respondents' educational attainment in GISCED, which would be comparable to the information collected in other countries (i.e. excluding dropout-categories). Then, a separate variable can be produced containing the information about additional but incomplete education (again coded in GISCED), which would however only be available for those countries and surveys where this information is collected.

[16] This leads to orientation also being introduced at ISCED level 1 in GISCED, resulting from the completion of short lower secondary programs that are not considered as level completion but where a distinction of general and vocational programs may already take place.

[17] The ESS uses code 9 on the third digit of 'edulvlb' for this, which in GISCED already signifies 'specification unknown' (see Section 4.2) so that this ESS solution cannot be adopted.

5 Applying GISCED

When harmonizing education categories using the GISCED framework, the best approach is to start from detailed, country-specific education variables, since these contain most information. This can be quite a labor-intensive process though. In order to establish which country-specific education category to map to which GISCED code, the official ISCED mappings (UNESCO 2021; Eurostat 2021) should primarily be used. These will usually give all information needed for coding into GISCED. A good rule to follow in order to maximize cross-national comparability is to only take into account *completed* education levels, and downgrade all education levels that were attended but not completed to the highest level that was successfully completed (see also Luxembourg Income Study n.d.).[18]

For extension variables, depending on the specific extension, this information will also be available in the ISCED mappings, but often in a non-standard format (e.g. included in a textual note rather than standardized column). Otherwise, this will need to be established via separate channels such as the scholarly literature, existing harmonized data such as the ESS, information on educational systems on the Internet (such as Eurydice[19] or Surveycodings[20]), or expert consultations. The same applies to educational qualifications not covered by the ISCED mappings, especially outdated qualifications (which are successively being added to the ISCED mappings though), which are covered in Surveycodings though.

Table 8 shows an example, where the GISCED scheme and extension variable on educational stratification are applied to the country-specific education variable for Austria in ESS round 2 (vettype would be 9 for the vocational categories 3 and 5 and is thus not included). The Austrian education measure can neither be harmonized into ISCED main levels nor broad levels (low/medium/high).[21] GISCED allows a coding of categories 1, 5 and 6, which span ISCED levels, that avoids any loss of information due to harmonization. The example also shows how 'edustrat' differentiates between two categories at ISCED level 3 which both include vocational education with access to tertiary education, but where one category gives access to university studies and the other does not.

If the data to be harmonized only contain already harmonized education categories (but with a different harmonization target), you can still use GISCED, which is (at least in the case of cross-national data) less labor intensive, but will also potentially result in less differentiated data (see next section for details). If using country-specific education variables, harmonized education variables using ISCED already present in the data can substantially speed up the process since often these will already contain most of the information needed to code the national education variables into GISCED. For example, the ISCED level could be derived from those already harmonized variables, and only for the coding of the second and third digit of ISCED as well as potential extension variables the country-specific variables would be needed. Whichever approach is chosen, it is important that the process is

[18] The SDR project chose the alternative to code also incomplete levels of education, but flag, in a harmonization control variable, cases where the level of education was not completed (Survey Data Harmonization Team 2017, p. 27). For GISCED it is recommended to code only completed levels of education in order to better align with official data on educational attainment and for higher consistency across countries, which mostly only measure completed levels of education, and code a separate variable with information on incomplete education (see Sect. 4.3.

[19] https://eacea.ec.europa.eu/national-policies/eurydice.

[20] http://www.surveycodings.org/levels-education, see also Schneider and Ortmanns (2019).

[21] Measures such as these were one reason for changing the ex-ante output harmonization procedures in the ESS from round 5 onwards.

Table 8 Coding the Austrian education variable in ESS round 2 (2004) to GISCED and edustrat

Country-specific categories		Description in English	GISCED	Edustrat
1	Pflichtschule nicht abgeschlossen	Compulsory school not completed	0100	0
2	Pflichtschule	Compulsory school	2244	2
3	Abschluss einer weiterbildenden Schule, BMS, Berufslehre, Berufsschule	Vocational education and training at upper secondary level, including apprenticeship	3354	2
4	Höhere Schule mit Matura (AHS, BHS)	High school with maturity exam (general, vocational)	3394	3
5	Ausbildung nach Abschluss einer Höheren Schule, Bakkalaureat, hochschulverwandte Lehranstalt (berufsbildende, pädagogische Akademie), Kolleg	Post high school education including apprenticeship, vocational college and Bachelor's degree	4699	9
6	Akademischer Grad, (Fach-) Hochschulabschluss oder äquivalent	Academic degree above Bachelor's degree	7890	9

fully documented, and documentation (including mapping tables, Wysmulek et al. 2015) made available to later users of the harmonized data (Granda et al. 2010).

Analysts may also wish to harmonize data ex-post that already contain harmonized education variables, but not the same ones (e.g. ESS and PIAAC). This can be achieved by mapping GISCED with coding schemes that are used in these datasets. This paper provides the mapping between four other harmonized education coding schemes and GISCED: years of education, the ESS-coding scheme 'edulvlb', the European Survey Version of ISCED (ES-ISCED), and CASMIN.

5.1 Linking GISCED with years of education

Like ISCED, GISCED normally assumes source variables representing levels of education. With some pragmatic inference, variables including information on completed years of education can, however, be recoded into ISCED levels and thus GISCED as well. In this case, the correspondences provided in Table 9 should be used, which use the cumulative years of education by which ISCED levels are generally defined.[22] In such a case, GISCED does not add anything to ISCED main levels apart from the distinction of 'some primary education', which forms the second digit of ISCED-A at ISCED level 0 (and is thus included for ISCED-A in Table 9). Then, it may well be better to use derived years of education as the harmonized target variable anyway in order not to throw out any information.

[22] Of course this will lead to anomalies across countries especially for countries that have different cumulative durations for specific levels, e.g. in Russia, upper secondary education with access to higher education normally ends after 11 years of education already, while ISCED indicates a typical duration of 12 – 13 years. Since primary education in Russia only starts at age 7, ISCED 3 graduates are still of the same age as in most countries then, but this is not reflected in a conversion from years of education to ISCED levels. Another anomaly will be the overestimation of levels of education given that actual educational careers often take longer than the nominal cumulative duration of all educational programs completed by an individual.

Table 9 Correspondence between years of education, ISCED-A, and GISCED

Years completed	Assumed completed ISCED-A level	GISCED	Description
0	00	0090	Never attended primary education
1	03	0030	Some primary education
2	03	0030	Some primary education
3	**03**	0030	Some primary education
4	03	0030	Some primary education
5	03	0030	Some primary education
6	**1**	1100	Primary education completed
7	1	1100	Primary education completed
8	1	1100	Primary education completed
9	**2**	2299	Lower secondary education
10	2	2299	Power secondary education
11	2	2299	Power secondary education
12	**3**	3399	Upper secondary education
13	3	3399	Upper secondary education
14	**5**	5590	Short-cycle tertiary education
15	6	6690	Bachelor-level education
16	**6**	6690	Bachelor-level education
17	7	7790	Master-level education
18	**7**	7790	Master-level education
19	7	7790	Master-level education
20	8	8890	Doctoral-level education
21	**8**	8890	Doctoral-level education
22 or more	8	8890	Doctoral-level education

For recoding ISCED levels into years of education, following the Luxembourg Income Study (2019), the years shown in bold here should be used.[23]

5.2 Linking GISCED with 'edulvlb', ISCED-A and ES-ISCED

Next, information for links between variables more directly relating to ISCED are presented together. The harmonized variable 'edulvlb' is used by the ESS since round 5 (2010). This coding scheme was implemented in the ESS while ISCED 2011 was still in development, but the most important features were already known. Therefore, the relationship between ISCED 2011 and the ESS education scheme is very close, even though the exact codes used at the second and third digits differ. However, the ESS education scheme adds information that is not covered in ISCED, along the lines of the differentiations mentioned here for the extension variable 'edustrat' (see Sect. 4.3). The mapping of country-specific education categories to ESS education codes is documented in Appendix A1 of the Data Documentation report of each survey round (European Social Survey 2018a, b, c,

[23] ISCED 4, which cannot be clearly identified by years of education, should also be derived into 14 years when going from ISCED levels to years of education.

Table 10 Correspondence between 'edulvlb', ISCED-A, ES-ISCED and GISCED with stratification extension

Edulvlb code and label		ISCED-A	ES-ISCED	GISCED	Edustrat
0	Less than primary completed	0	I	0090	0
113	ISCED 1, general	100	I	1100	0
119	Short general ISCED 2	100	I	1140	6
129	Short vocational ISCED 2	100	I	1150	6
211	General ISCED 2 without access to ISCED 3	243	II	2243	1
212	General/pre-vocational ISCED 2 with access to 32x	244	II	2244	2
213	General ISCED 2 with access to all ISCED 3	244	II	2244	3
219	Short general ISCED 3	244	II	2244	6
221	Vocational ISCED 2 without access to ISCED 3	253	II	2253	1
222	Vocational ISCED 2 with access to 32x	254	II	2254	2
223	Vocational ISCED 2 with access to all ISCED 3	254	II	2254	3
229	Short vocational ISCED 3	254	II	2254	6
311	General ISCED 3 without access to tertiary	343	IIIb	3343	1
312	General ISCED 3 with access to voc/lower tier tertiary	344	IIIa	3344	2
313	General ISCED 3 with access to upper/single tier higher education	344	IIIa	3344	3
321	Vocational ISCED 3 without access to tertiary	353	IIIb	3353	1
322	Vocational ISCED 3 with access to voc/lower tier tertiary	354	IIIb	3354	2
323	Vocational ISCED 3 with access to academic tertiary	354	IIIa	3354	3
411	General ISCED 4 without access to tertiary	443	IV	4443	1
412	General ISCED 4, with access to voc/lower tier tertiary	444	IV	4444	2
413	General ISCED 4 with access to academic tertiary	444	IV	4444	3
421	Vocational ISCED 4 without access to tertiary	453	IV	4453	1
422	Vocational ISCED 4, with access to voc/lower tier tertiary	454	IV	4454	2
423	Vocational ISCED 4, with access to academic tertiary	454	IV	4454	3
510	General/academic tertiary qualification below bachelor's degree	540	IV	5540	0
520	Vocational/professional tertiary qualification below bachelor's degree	550	IV	5550	0

Table 10 (continued)

Edulvlb code and label		ISCED-A	ES-ISCED	GISCED	Edustrat
610	Lower tier bachelor's degree	660	V1	6690	4
620	Upper/single tier bachelor's degree	660	V1	6690	5
710	Lower tier master's degree	760	V2	7790	4
720	Upper/single tier master's degree	760	V2	7790	5
800	Doctoral degree	860	V2	8890	0

Table 11 Correspondence between CASMIN and GISCED with stratification extension

Code	Label	Edustrat	GISCED
1a	Inadequately completed general education	0	0100
1b	General elementary education	2	2244
1c	Basic vocational qualification or general elementary education and vocational qualification	2	2359
2a	Intermediate vocational qualification or intermediate general qualification and vocational qualification	3	2359
2b	Intermediate general qualification	3	2244
2c_gen	General maturity certificate	3	3344
2c_voc	Vocational maturity certificate/General maturity certificate and vocational qualification	3	3454
3a	Lower tertiary education	4	5690
3a_gen	Lower tertiary education - general diplomas	4	5640
3a_voc	Lower tertiary education - diplomas with vocational emphasis	4	5650
3b	Higher tertiary education	5	6890
3b_low	Higher tertiary education - lower level	5	6690
3b_high	Higher tertiary education - higher level	5	7890

2020a, b). These may thus also help for the harmonized coding of the extension variable 'edustrat' in other datasets. Table 10 shows how 'edulvlb', ISCED-A and GISCED correspond, using 'edustrat' as supporting variable. The close relationship between edulvlb and ISCED-A is very apparent, resulting in GISCED codes that do not actually need the new digit for categories spanning ISCED levels (see Sect. 4.1). New information is located almost exclusively in the extension variable 'edustrat'.

From 'edulvlb', and thus also GISCED with the extension variable 'edustrat', it is also possible to derive the so-called European Survey-Version of ISCED (ES-ISCED). This variable has a much lower number of categories than 'edulvlb' and is thus suitable for statistical analysis. It has been shown to produce more valid and comparable results than ISCED main levels (Schneider 2010; Schröder 2014), and is thus an attractive alternative to ISCED main levels.

5.3 Linking GISCED with CASMIN

The final comparative education coding scheme to link with GISCED is the CASMIN education scheme, which has its roots in social stratification research. It is not possible to code CASMIN into ISCED directly because some CASMIN categories span across ISCED levels, which GISCED, however, has a solution for (see Sect. 4.1). The mapping of CASMIN to GISCED is more difficult than of the ESS education scheme because in contrast to ISCED, CASMIN is constructed on a relative education scale in order to capture social selectivity effects of education (Brauns et al. 2003; König et al. 1988; Brauns and Steinmann 1999). It thereby identifies both class-specific barriers in educational systems and labor market signals of educational qualifications. This means that the mapping of qualifications to CASMIN is somewhat a 'moving target' since selectivity changes over time as educational systems expand. The duration of education as well as vocational vs. general

education play a considerable role in both CASMIN and GISCED (as well as ISCED) though, allowing the construction of a rough correspondence.

CASMIN category 1b ('general elementary education') is defined as the (non-selective) 'social minimum' of education in any given country. To achieve a correspondence, I assume ISCED level 2 to correspond to this 'social minimum', even though this may differ across countries and time. Also, CASMIN differentiates general elementary education and intermediate general education, which, historically, maps to primary education (ISCED 1) and lower secondary education (ISCED 2) respectively, but today rather corresponds to differentiations *within* ISCED 2. This distinction is achieved by combining the extension variable on educational stratification with ISCED. Similarly, when general elementary or intermediate general education is combined with vocational training (categories 1c - Basic vocational qualification or general elementary education and vocational qualification and 2a - Intermediate vocational qualification or intermediate general qualification and vocational qualification), this can only be distinguished using 'edustrat', using the same values as with the qualifications without subsequent VET. Additionally, these two CASMIN categories span across ISCED levels. We therefore map both with vocational education at ISCED 2 to 3, resulting in code 2359[24]. Finally, while CASMIN 3a corresponds to short tertiary education in most countries, which can be mapped to ISCED level 5, for Germany, it refers to the lower tier of higher education, classified as ISCED level 6. As a consequence, CASMIN 3a will be mapped to ISCED 5 and 6 lower tier, and CASMIN 3b to ISCED 6 – 8 upper tier. Table 11 shows how CASMIN coded data can be coded into GISCED, including 'edustrat'. It is apparent that here, the harmonization using GISCED is much less straightforward than for the ESS education coding scheme, but nevertheless, it can be done (unlike with just ISCED).

6 Empirical illustration

Beyond the fact that GISCED improves the interoperability and re-usability of data and allows studying more specific education-related hypotheses than years of education or broad education levels, another criterion for its usefulness is what can be gained for substantive and policy-oriented research when using GISCED, possibly with extension variables, rather than other ways of harmonizing educational attainment. For illustration, this section presents some empirical analyses, using educational inequality, i.e. the impact of parental education on respondents' education, as an example. It uses ESS round 5 to 9 data (ESS ERIC 2018a, b, c,2020a, b).

Respondents' education is reduced to levels of education, i.e. the ordinal ISCED main levels, which are treated as a metric variable.[25] Parental education is harmonized in a number of different ways: Model 1 uses three broad education levels derived from ISCED

[24] For Germany, the GISCED code derived from CASMIN 1c and 2a should be 2559, because the master crafts qualifications classified in ISCED level 5 are included in CASMIN categories 1c and 2a rather than 3a. Applying the 'majority rule' to German qualifications would result in GISCED 2359 as well though.

[25] While there is a debate on whether this is justified (Winship and Mare 1984; Kim 1975), others point to the difficulties of correctly interpreting the outcomes of alternative non-linear probability models (Breen et al. 2018). I choose a linear model because error produced by this strategy is usually very small (Labovitz 1970), and to ease interpretation. It also appears justified especially when the number of levels is as high as in this case.

levels, Model 2 uses ISCED 2011 levels, and Model 3 uses the detailed ESS education variable 'edulvlb', which corresponds to the combination of GISCED codes and the extension variable edustrat (see Table 10).[26] Parental education is thus conceptualized as both levels and types of education, i.e. education as a categorical variable using GISCED and extension variables. In this way, we can see how different levels *and* types of education influence the educational attainment in the filial generation. This model shows how using a high level of detail in harmonization, including stratification in education, provides further insights into educational inequalities across generations than ISCED levels alone. Model 4 uses ES-ISCED as an alternative aggregation to main ISCED levels, because a detailed coding scheme like the one used in Model 3 is not very practical in empirical analyses. As has been suggested elsewhere (Schröder 2014), ISCED 0 is differentiated from ISCED 1 here because in the parental generation, less than primary education is actually still a rather common category. Table 12 shows results from these four random intercept multilevel linear regression models of respondents' educational attainment on the education of their most highly educated parent. For each model, R-squared following Snijders and Bosker 1994 is reported for both level 1 (respondents) and level 2 (countries), using the Stata package 'Multilevel tools' (Möhring and Schmidt 2012). All models use design weights, control for age and sex, and use respondents aged 26 to 64 only.

The results show that the broader the education measure, the more heterogeneity of effects is hidden from view. This is particularly visible when comparing Model 1 with Model 2, and Model 2 with Model 3. Looking at Model 1, compared to respondents with parents with a low level of education (ISCED 0 – 2), respondents with highly educated parents (ISCED 5 – 8) obtain two more levels of education, and those with medium educated parents one more level of education. Model 2 reveals that each additional level of education of parents leads to approximately half an extra level of education for their children, with two exceptions: (1) respondents with parents that have not even completed primary education (ISCED 1) are severely disadvantaged and get three quarters of an education level less than those whose parents have at least completed ISCED 1 (granted, the number of affected respondents is low, but in terms of social policy, this is a highly relevant result). (2) having parents educated at ISCED level 5 only gives a quarter extra level compared to parents educated at ISCED level 4, and ISCED 6 another quarter extra level compared to ISCED 5. The differentiations introduced by ISCED 2011 compared to ISCED 1997 (where ISCED levels 5 – 7 were all in one level) appear to have been fruitful, with ISCED 5 having lower effects than ISCED 6, ISCED 6 lower effects than ISCED 7, and ISCED 7 lower effects than ISCED 8 (all effects are statistically significantly different from each other). Having one parent with a PhD gives an advantage of almost three education levels compared to having parents with primary education at most. Also these results show that even though ISCED 4 is rare, it does lead to effects that are significantly higher than those of ISCED 3 (the level it is commonly aggregated with).

Then looking at the highly detailed results of Model 3, we find a number of interesting differentiations: Firstly, children of parents with vocational lower secondary education (ISCED 2), which has a rather bad reputation, achieve a *higher* level of education than children of parents with just general lower secondary education, *if the vocational program*

[26] Both variables, GISCED and edustrat, are best combined into one variable for analysis because 'edustrat' is only relevant for a number of GISCED codes. The same applies to the flag for completion of a short higher level program. There are no theoretical reasons to expect that type of VET would differentially affect educational inequality so that the extension variable 'vettype' is not used here.

Table 12 Multilevel models for validation using various education codings (N L1=145350; N L2=35)

Model 1			Model 2			Model 3							Model 4		
ISCED broad levels			ISCED 2011			GISCED with edustrat, corresponding to 'edulvlb'							ES-ISCED		
Lev	b	p	Lev	b	p	GISCED Category		Edustrat/flag short higher program		Edulvlb	b	p	Lev	b	p
Low	Ref.		0	-0.75	0,000	0090	Not completed ISCED level 1	0	Not further specified	0	-0.74	0.000	0	-0.75	0.000
			1	Ref.		1100	ISCED 1, completed primary education	0	Not further specified	113	ref.		I	ref.	
						1100	ISCED 1, completed primary education	6	With short voc program	129	0.28	0.003			
			2	0.48	0.000	2253	Voc ISCED 2>= 2 years, no access ISCED 3	1	No access	221	0.42	0.000	II	0.48	0.000
						2244	eneral ISCED 2, access ISCED 3	2	To vocational/Lower tier	212	0.27	0.008			
						2244	General ISCED 2, access ISCED 3	3	To academic/Upper/single tier	213	0.47	0.000			
						2254	Voc ISCED 2, access ISCED 3	2	To vocational/Lower tier	222	0.59	0.000			
						2254	Voc ISCED 2, access ISCED 3	3	To academic/Upper/single tier	223	0.66	0.000			
						2254	Voc ISCED 2, access ISCED 3	6	With short voc program	229	0.64	0.000			
Med.	0.90	0.000	3	1.02	0.000	3343	Gen ISCED 3>= 2 years, no access ISCED 5	1	No access	311	1.07	0.000	IIIb	0.81	0.000
						3353	Voc ISCED 3>= 2 years, no access ISCED 5	1	No access	321	0.82	0.000			
						3354	Voc ISCED 3, access ISCED 5	2	To vocational/Lower tier	322	0.86	0.000			
						3344	Gen ISCED 3, access ISCED 5	2	To vocational/Lower tier	312	1.31	0.000	IIIa	1.21	0.000
						3344	Gen ISCED 3, access ISCED 5	3	To academic/Upper/Single tier	313	1.12	0.000			
						3354	Voc ISCED 3, access ISCED 5	3	To academic/Upper/Single tier	323	1.25	0.000			

Table 12 (continued)

Model 1			Model 2			Model 3							Model 4		
ISCED broad levels			ISCED 2011			GISCED with edustrat, corresponding to 'edulvlb'							ES-ISCED		
Lev	b	p	Lev	b	p	GISCED Category		Edustrat/flag short higher program		Edulvlb	b	p	Lev	b	p
			4	1.45	0.000	4453	Voc ISCED 4 no access ISCED 5	1	No access	421	1.28	0.000	IV	1.60	0.000
						4444	Gen ISCED 4, access ISCED 5	2	To vocational/Lower tier	412	1.99	0.000			
						4444	Gen ISCED 4, access ISCED 5	3	To academic/Upper/single tier	413	1.47	0.000			
						4454	Voc ISCED 4, access ISCED 5	2	To vocational/Lower tier	422	1.47	0.000			
						4454	Voc ISCED 4, access ISCED 5	3	To academic/Uupper/single tier	423	1.51	0.000			
High	2.04	0.000	5	1.73	0.000	5540	Gen ISCED 5	0	Not further specified	510	1.69	0.000			
						5550	Voc ISCED 5	0	Not further specified	520	1.73	0.000			
			6	2.05	0.000	6690	ISCED 6	4	Lower tier	610	1.93	0.000	V	2.04	0.000
						6690	ISCED 6	5	Upper/Single tier	620	2.15	0.000			
			7	2.57	0.000	7790	ISCED 7	4	Lower tier	710	2.29	0.000	VI	2.60	0.000
						7790	ISCED 7	5	Upper/Single tier	720	2.61	0.000			
			8	2.92	0.000	8890	ISCED 8	0	Not further specified	800	2.90	0.000			
Const.	3.45	0.000		3.15	0.000						3.13			3.13	0.000
R2 L1	21.79			24.74							25.17			24.97	
R2 L2	59.25			61.36							62.31			61.67	

gives access to upper secondary education (GISCED 2254). If it does not (GISCED 2253), there is no statistically significant disadvantage though. Within general lower secondary education, parents who completed the higher or a single track (edustrat code 3) more positively influence their children's education than those who completed a lower (pre-vocational) track (edustrat code 2), but the difference between these effects is only marginally statistically significant (p=0.0684).

Secondly, within upper secondary education (ISCED 3), there are also marked differences between the effects of generally versus vocationally educated parents, here pointing to advantages of children of *generally* educated parents. The effects of different kinds of upper secondary *general* education are not statistically significantly different from each other. However, parents with vocational upper secondary education that gives access to academic higher education (GISCED 3354, edustrat code 3) give their offspring the same advantage as those with general upper secondary education (GISCED 3344). In line with previous research (Schneider 2010), vocational upper secondary education giving access to all kinds of higher education is not a disadvantage compared to general education. If the access is limited to only the lower tier or vocational tertiary education though, there is a statistically significant disadvantage compared to those with full access. This is likely due to the different content of vocational upper secondary education programs giving access to university studies, and resulting differences in skills of its graduates. Here we find that the extension variable 'edustrat' indeed carries important information that is not represented by ISCED alone.

Thirdly, while ISCED level 4 is rather uncommon, we still find differences in effects here, notably that parents who graduated from ISCED 4 programs that do not give access to tertiary education do not convey any advantage to their children beyond the advantage they would have provided when completing upper secondary education with access to tertiary. Still, since this type of ISCED 4 education often follows ISCED 3 programs that do not give access to tertiary education, compared to that, parents with ISCED 4 still give some advantage to their children.

Finally, in tertiary education, while distinctions of orientation of education are irrelevant at ISCED level 5, i.e. for qualifications below the Bachelor's degree, types of education as identified by the extension variable 'edustrat' are relevant in the levels belonging to higher education strictly speaking (ISCED 6 – 8). Both in ISCED levels 6 and 7, parents who completed lower tier (i.e. college or polytechnic) programs (edustrat code 4) convey less educational advantage to their offspring than parents who completed higher or single tier (i.e. traditional university) programs (edustrat code 5). These results are very much in line with expectations from educational stratification research, and again point to the validity of the extension variable 'edustrat', and that this is a useful extension of ISCED.

Model 4 shows results when aggregating the detailed education variable in an alternative way, not using ISCED main levels. Using ES-ISCED reveals how different types of upper secondary education of the parents differ in their effects on the educational attainment of their children. The explained variance is between the one obtained with ISCED main levels and edulvlb, even though the variable has one category less than ISCED has main levels. It is thus a more parsimonious alternative to ISCED levels, which however cannot be derived from data coded using ISCED levels alone.

7 Discussion and outlook

The combination of different datasets is increasingly popular. Survey variables on educational attainment are, however, often not coherently measured and coded: Different national and cross-national surveys use different measurement instruments and coding schemes. In order to support researchers, data harmonization projects and data producers with the task of education harmonization, this paper described a generalized coding framework for ex-post harmonization of educational attainment variables in surveys, called GISCED. It is inspired by the education coding scheme 'edulvlb' used in the ESS, but adapts it for the purpose of ex-post harmonization and sets up a closer link with ISCED 2011. By extending the official three-digit ISCED code to four digits, it should be possible to code almost every education category in any dataset to an internationally intelligible code—if it is well enough documented to understand its relationship with ISCED main levels. For retaining information available in source data that is not taken into account by ISCED, it is suggested to code separate 'extension' variables recording this information in standard format, e.g., the secondary education track, type of higher education, or type of vocational education and training.

Compared to other education coding schemes, the suggested framework results in data with a maximum possibility for reuse and interoperability if coded at the detailed four-digit level. GISCED can be transformed into ISCED 2011 and 1997 (in detail and main levels) and derived years of education. Data coded at this level of detail can be further processed to satisfy many different research needs and allows testing of innovative hypotheses that cannot be tested using measures based on years or main levels of education alone, especially when looking at differential effects of different types of education within main levels. It can also help coding aggregated education variables in a way that allows better modeling of specific education effects, as the above empirical illustration showed when using ES-ISCED distinguishing university-preparatory from other programmes at the upper secondary level. Depending on the specific research questions, there are many other ways in which the detailed variable can be simplified. The fact that the framework has four digits makes this look more complex than it actually is since only few codes are actually used at the third and fourth digits, and not all combinations are possible.

Granted, these fine-grained distinctions will not be relevant for all research using education as an independent variable, but the above illustration shows that they do carry meaning. Depending on the purpose of the data harmonization, more or less effort may be justified in producing such harmonized extra information. For many ex-post harmonization projects, the variant 'GISCED2' that only improves the harmonization into main levels of education without considering program orientation or sub-levels, will be sufficient. For research interested in social and educational inequalities, further details will generate richer research opportunities.

The empirical illustration is of course rather limited and the results may not be generalizable to other relationships. Sor further research might further investigate which kinds of research a detailed harmonization is most fruitful for, and for which kinds of research it is not worth the effort. One could, for example, additionally look at how track placement amongst respondents is influenced by parental education (where parental tracks may be even more important), or how respondents' occupational attainment is influenced by their education (where type of education, including type of VET as measured by the extension variable 'vettype' may be highly relevant), but this goes beyond the scope of this paper.

The illustration is also based on ESS data that are already available with detailed education codes obtained via *ex-ante* harmonization. In a next step, it would be highly relevant to employ the GISCED scheme for a real *ex-post* harmonization project to further study its usability and empirical value. The generalized coding framework may be most useful as a coding framework for international ex-post harmonization projects. However, national projects, especially longitudinal surveys, may find it equally useful to structure their harmonization over time given national education classifications suitable for survey use do not exist in all countries. For data producers and harmonization projects that aim to produce data that are suitable for a range of research questions, including labor market and social stratification research, the framework could be a helpful tool to standardize the harmonization of education categories across diverse sources that were not designed with cross-national or over time comparability in mind.

If the source variables already differ in their scope, e.g. one dataset only measures vocational education if this is provided as full-time schooling (like the ISSP), and another one measures all vocational education irrespective of whether the school-based element is full-time or part-time (like many surveys using ISCED), some degree of non-comparability cannot be overcome by a common code-frame that otherwise harmonizes the categories of the education variables in the respective surveys. GISCED can do nothing to remedy this but recommend the usage of harmonization controls (Slomczynski and Tomescu-Dubrow 2018) to data users, as well as adherence to the definition and guidelines for the measurement of educational attainment in official statistics, such as OECD, Eurostat (2014), to data producers.

While GISCED will allow almost every education category appearing in surveys using measures of levels of education to be harmonized with very little loss of information, it will still result in heterogeneous data if the variables to be harmonized are measured at very different levels of detail. Therefore, the general recommendation for surveys to not aggregate education categories too much already at the stage of data collection (Schneider 2008) is still highly relevant: information not gathered cannot be coded and harmonized—whether using ISCED, GISCED, derived years of education or some other form of scaling.

It is then up to the analyst to make the most of the information provided by recoding into simpler categories that are adequate for the research question at hand, that can be used in statistical analyses. Looking at the distribution of all codes in the harmonized dataset, analysts will be better able to aggregate education variables in a substantially meaningful way for analysis than opting for just broad ISCED levels from the outset and accepting substantial information loss. Further research should look into how reliably data can be harmonized using GISCED and the proposed extension variables, and look more deeply into the substantive research potential and cross-national comparability of the resulting data compared to other harmonization schemes.

Acknowledgements I would like to thank my colleague Verena Ortmanns, the anonymous reviewers as well as the editors of this special issue for their valuable comments that substantially helped improve the paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with ethical standards

Funding This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG), as part of the development of the National Research Data Infrastructure (NFDI) Consortium for Social, Behavioural, Educational, and Economic Sciences (KonsortSWD), grant number NFDI 19/1, project number 442494171.

Conflict of interest The authors declare that they have no conflict of interest.

Data availability Data used in the empirical analyses is available online and references are provided.

Code availability Code used for data analysis is made available as supplementary material.

References

Allen J, Massing N, Schneider SL, van der Velden R (2017) Proposal for the revision of the Background Questionnaire on education and training for the 2nd cycle of PIAAC. Tech. rep., ROA, Maastricht, https://doi.org/10.1017/CBO9781107415324.004 arXiv: 1011.1669v3

Allmendinger, J.: Educational systems and labor market outcomes. Eur. Sociol. Rev. **5**(3), 231–250 (1989)

Andersen, R., van de Werfhorst, H.G.: Education and occupational status in 14 countries: The role of educational institutions and labour market coordination. Br. J. Sociol. **61**(2), 336–355 (2010)

Barone, C., Ruggera, L.: Educational equalization stalled? Trends in inequality of educational opportunity between 1930 and 1980 across 26 European nations. Eur. Soc. **20**(1), 1–25 (2018)

Bol T, van de Werfhorst HG (2011) Measuring Educational Institutional Diversity : External Differentiation , Vocational Orientation and Standardization Thijs Bol. Amsterdam Centre for Inequality Studies Working Paper Series 1(November)

Bol, T., Van de Werfhorst, H.G.: Measuring educational institutional diversity: tracking, vocational orientation and standardisation. In: Hadjar, A., Gross, C. (eds.) Education systems and inequalities: international comparisons, pp. 73–94. Policy Press, Bristol (2016)

Bol, T., Witschge, J., Van de Werfhorst, H.G., Dronkers, J.: Curricular tracking and central examinations: counterbalancing the impact of social background on student achievement in 36 Countries. Social Forces **92**, 1545–1572 (2014)

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S.: Data resource profile: the survey of health, ageing and retirement in Europe (SHARE). Int. J. Epidemiol. **42**(4), 992–1001 (2013)

Bourdieu, P.: Homo Academicus. Stanford University Press, Stanford (CA) (1988)

Boyle EH, King M, Sobek M (2019) IPUMS-Demographic and Health Surveys: Version 7 [dataset]. https://doi.org/10.18128/D080.V7

Braun, M., Müller, W.: Measurement of education in comparative research. Comp. Social Res. **16**, 163–201 (1997)

Brauns H, Steinmann S (1999) Educational reform in France, West-Germany, and the United Kingdom. Updating the CASMIN educational classification. MZES Working Paper 23, MZES, Mannheim

Brauns, H., Scherer, S., Steinmann, S.: The CASMIN educational classification in international comparative research. In: Hoffmeyer-Zlotnik, J.H., Wolf, C. (eds.) Advances in cross-national comparison: a European working book for demographic and socio-economic variables, pp. 221–244. Kluwer Academic/Plenum, New York (2003)

Breen, R. (ed.): Social mobility in Europe. Oxford University Press, Oxford (2004)

Breen, R., Luijkx, R., Müller, W., Pollak, R.: Nonpersistent inequality in educational attainment: Evidence from eight European countries. Am. J. Sociol. **114**(5), 1475–1521 (2009)

Breen, R., Karlson, K.B., Holm, A.: Interpreting and understanding logits, probits, and other nonlinear probability models. Annual Rev. Sociol. **44**(1), 39–54 (2018)

Brons A, Mooyaart JE (2018) The Generations & Gender Programme: Constructing harmonized, continuous socio-economic variables for the GGS Wave 1. Tech. rep., Den Haag, https://pure.knaw.

nl/ws/files/9348005/2018_Brons_Mooyaart_Continuous_Harmonized_SES_Variables_GGP_working_paper_publ_vers_v142.pdf
Delaruelle, K., Van Houtte, M., Bracke, P.: Educational inequalities in general health: Does the curricular tracking system matter? Acta Sociologica **63**, 63–81 (2020)
Dieckhoff, M.: Skills and occupational attainment: a comparative study of Germany, Denmark and the UK. Work, Employ.Soc. **22**(1), 89–108 (2008)
Doiron, D., Raina, P., L'Heureux, F., Fortier, I.: Facilitating collaborative research: implementing a platform supporting data harmonization and pooling. Norsk Epidemiologi **21**(2), 352 (2012)
Dubrow, J.K., Tomescu-Dubrow, I.: The rise of cross-national survey data harmonization in the social sciences: emergence of an interdisciplinary methodological field. Qual. Quant. **50**(4), 1449–1467 (2016)
Duncan, O.D., Hodge, R.W.: Education and occupational mobility a regression analysis. Am. J. Sociol. **68**(6), 629–644 (1963)
Ehling, M.: Harmonising data in official statistics. In: Hoffmeyer-Zlotnik, J.H., Wolf, C. (eds.) Advances in cross-national comparison: a European working book for demographic and socio-economic variables, pp. 17–31. Kluwer Academic/Plenum, New York (2003)
Erikson, R., Goldthorpe, J.H.: The constant flux: a study of class mobility in industrial societies. Clarendon Press, Oxford (1992)
Erikson, R., Goldthorpe, J.H., Portocarero, L.: Intergenerational class mobility in 3 Western European societies - England, France and Sweden. Br. J. Sociol. **30**(4), 415–441 (1979)
ESS ERIC (2018a) European Social Survey Round 5 Data (2010). Data file edition 3.4. NSD - Norwegian Centre for Research Data–Data Archive and distributor of ESS data for ESS ERIC, Bergen, https://doi.org/10/ghfg9k
ESS ERIC (2018b) European Social Survey Round 6 Data (2012). Data file edition 2.4. NSD - Norwegian Centre for Research Data – Data Archive and distributor of ESS data for ESS ERIC, Bergen, https://doi.org/10/ghfg9j
ESS ERIC (2018c) European Social Survey Round 7 Data (2014). Data file edition 2.2. NSD - Norwegian Centre for Research Data – Data Archive and distributor of ESS data for ESS ERIC, Bergen, https://doi.org/10/ghfg9m
ESS ERIC (2020a) European Social Survey Round 8 Data (2016). Data file edition 2.2. NSD - Norwegian Centre for Research Data – Data Archive and distributor of ESS data for ESS ERIC, Bergen, https://doi.org/10/ggnsh6
ESS ERIC (2020b) European Social Survey Round 9 Data (2018). Data file edition 3.0. NSD - Norwegian Centre for Research Data – Data Archive and distributor of ESS data for ESS ERIC, Bergen, https://doi.org/10/ghfg9n
European Social Survey (2018a) Appendix A1–Education, ESS5–2010 ed. 4.2. In: ESS5 Data Documentation Report, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC, Bergen
European Social Survey (2018b) Appendix A1–Education, ESS6–2012 ed. 2.0. In: ESS6 Data Documentation Report, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC, Bergen
European Social Survey (2018c) Appendix A1–Education, ESS7–2014 ed. 3.2. In: ESS7 Data Documentation Report Ed. 3.0, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC, Bergen, pp 1–80
European Social Survey (2020a) Appendix A1–Education, ESS8–2016 ed. 2.3. In: ESS8 Data Documentation Report, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC, Bergen
European Social Survey (2020b) Appendix A1 Education, ESS9–2018 ed. 3.0. In: ESS5 Data Documentation Report, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC
European Social Survey (2020c) ESS-9 2018 Documentation Report. Edition 3.0. European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC, Bergen
European Values Study (2020) European Values Study 2017: Integrated Dataset (EVS 2017). ZA7500 Data File Version 3.0.0. GESIS Data Archive, Cologne, https://doi.org/10.4232/1.13511
Eurostat (2007) Task Force on Core Social Variables Final Report. Office for Official Publ. of the European Communities, Luxembourg, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-006/EN/KS-RA-07-006-EN.PDF
Eurostat (2020) European Union Labour Force Survey (EU LFS). https://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey

Eurostat (2021) National ISCED Mappings. https://circabc.europa.eu/w/browse/c2dc65ad-5163-4935-b0c2-e5ea1f44929b
Flabbi, L., Paternostro, S., Tiongson, E.R.: Returns to education in the economic transition: a systematic assessment using comparable data. Econ. Educ. Rev. **27**(6), 724–740 (2008)
Forster, A.G., Bol, T., van de Werfhorst, H.G.: Vocational education and employment over the life cycle. Sociol. Sci. **3**, 473–494 (2016)
Ganzeboom HB, Treiman DJ (2019) International Stratification and Mobility File: Conversion Tools. http://www.harryganzeboom.nl/ismf/index.htm
Ganzeboom, H.B., de Graaf, P.M., Treiman, D.J.: A standard international socio-economic index of occupational status. Social Sci. Res. **21**(1), 1–56 (1992)
Ganzeboom HBG (2019) International stratification and mobility file: Overview of harmonized variables. http://www.harryganzeboom.nl/ismf/harmonizing.htm
Granda P, Blasczyk E (2010) Data Harmonization. In: Cross-Cultural Survey Guidelines, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor
Granda, P., Wolf, C., Hadorn, R.: Harmonizing survey data. In: Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Mohler, P.P., Pennel, B.E., Smith, T.W. (eds.) Survey methods in multinational, multiregional, and multicultural contexts. Wiley, Hoboken (2010)
Hancock A (2013) Best Practice Guidelines for Developing International Statistical Classifications. Expert Group Meeting on International Statistical Classifications. New York, 13-15 May 2013. May, United Nations Department of Economic and Social Affairs, Statistics Division, New York, http://unstats.un.org/unsd/class/intercop/expertgroup/2013/AC267-5.PDF
Hanushek, E.A., Schwerdt, G., Woessmann, L., Zhang, L.: General education, vocational education, and labor-market outcomes over the lifecycle. J. Human Res. **52**(1), 48–87 (2017)
Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.E., Smith, T.W.: Comparative survey methodology. In: Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.E., Smith, T.W. (eds.) Survey methods in multinational, multiregional, and multicultural contexts. Wiley, Hoboken (2010)
Hauser, R.M.: Comparable metrics: Some examples. Chin. J. Sociol. **2**(1), 3–33 (2016)
International Labour Organisation (2007) Resolution Concerning Updating the International Standard Classification of Occupations
Kerckhoff, A.C., Dylan, M.: Problems with international measures of education. J. Socio-Econ. **28**(6), 759–775 (1999)
Kerckhoff, A.C., Ezell, E.D., Brown, J.S.: Toward an improved measure of educational attainment in social stratification research. Social Sci. Res. **31**(1), 99–123 (2002)
Kim, J.O.: Multivariate Analysis of Ordinal Variables. Am. J. Sociol. **81**(2), 261–298 (1975). https://doi.org/10.1086/226074
König W, Lüttinger P, Müller W (1988) A comparative analysis of the development and structure of educational systems. CASMIN Working Paper 12, University of Mannheim, Mannheim
Labovitz, S.: The assignment of numbers to rank order categories. Am. Sociol. Rev. **35**(3), 515–524 (1970). https://doi.org/10.2307/2092993
Lucas, S.R.: Effectively maintained inequality: education transitions, track mobility, and social background effects. Am. J. Sociol. **106**(6), 1642–1690 (2001)
Luxembourg Income Study (2019) Generic Codebook, LIS Database, 2019 Template. Tech. rep., LIS office, Luxembourg, https://www.lisdatacenter.org/wp-content/uploads/files/data-lis_codebook.pdf
Luxembourg Income Study (n.d.) The standardisation routines for highest level of completed education. Tech. rep., http://www.lisdatacenter.org/wp-content/uploads/standardisation-of-education-levels.pdf
Marginson, S.: The worldwide trend to high participation higher education: dynamics of social stratification in inclusive systems. Higher Educ. **72**(4), 413–434 (2016)
Mincer, J.: Schooling, Experience, and Earnings, vol. 2. National Bureau of Economic Research distributed by Columbia University Press, New York (1974)
Minnesota Population Center (2019) Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Tech. rep., Minnesota Population Center, Minneapolis, MN, https://doi.org/10.18128/D020.V7.2
Möhring K, Schmidt A (2012) Multilevel tools - an ado-package for Stata. In: German Stata Users' Group Meetings 2012, Stata Users Group, Berlin, https://ideas.repec.org/p/boc/dsug12/06.html
OECD: Technical Report of the Survey of Adult Skills (PIAAC), 2nd edn. OECD, Paris (2016)
OECD (2017) Education at a Glance 2017 - OECD Indicators. OECD Publishing, Paris, http://www.oecd.org/edu/education-at-a-glance-19991487.htm
OECD, Eurostat, : Joint Eurostat-OECD guidelines on the measurement of educational attainment in household surveys. Tech. Rep, September, OECD, Eurostat (2014)

Oesch, D.: Redrawing the class map: stratification and institutions in Britain, Germany, Sweden, and Switzerland. Palgrave Macmillan, Basingstoke (2006)
Rose D, Harrison E (eds) (2010) Social Class in Europe: An Introduction to the European Socio-Economic Classification. Routledge
Saar E, Roosalu T, Helemäe J, Roosmaa EL, Martma L, Kazjulja M (2017) Education and labour market outcomes for vocational education and training graduates in different types of VET systems in Europe. Tech. rep., cedefop, https://www.cedefop.europa.eu/files/working_paper_4.1_education_and_labour_market_outcomes_for_vet_graduates_in_different_types_of_vet_systems_in_europe.pdf
Schneider SL (2008) Suggestions for the cross-national measurement of educational attainment: Refining the ISCED-97 and improving data collection and coding procedures. In: Schneider SL (ed) The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries, MZES, Mannheim, pp 311–330,https://doi.org/10.13140/RG.2.1.4614.9526
Schneider, S.L.: Nominal comparability is not enough: (In-)equivalence of construct validity of cross-national measures of educational attainment in the European Social Survey. Res. Social Stratification Mobil **28**(3), 343–357 (2010)
Schneider, S.L.: The international standard classification of education 2011, comparative social research. In: Birkelund, G.E. (ed.) Class and stratification analysis. Emerald Group Publishing Limited, Canada (2011)
Schneider, S.L.: The Conceptualisation, Measurement, and Coding of Education in German and Cross-National Surveys. Tech. rep, GESIS - Leibniz Institute for the Social Sciences, Mannheim (2016).. (https://doi.org/10.15465/gesis-sg_en_020)
Schneider SL, Kogan I (2008) The International Standard Classification of Education 1997: Challenges in the application to national data and the implementation in cross-national surveys. In: Schneider SL (ed) The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries, MZES, Mannheim, pp 13–46, https://doi.org/10.13140/RG.2.1.2517.8004
Schneider SL, 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. GESIS - Leibniz Institute for the Social Sciences, https://www.surveycodings.org/sites/default/files/content_files/SERISS%20Deliverables%20D8.8%20Database%20of%20educational%20attainment_20190830_final_edit20190920.pdf
Schneider, S.L., Joye, D., Wolf, C.: When translation is not enough: background variables in comparative surveys. In: Joye, D., Smith, T.W., Fu, Y.C. (eds.) Wolf C, pp. 288–307. The SAGE Handbook of Survey Methodology, Los Angeles (2016)
Schröder H (2014) Levels and Loadings: Two Methods to Improve the Measurement of Education in Comparative Research. PhD thesis, Vrije Universiteit Amsterdam, Amsterdam, https://research.vu.nl/en/publications/levels-and-loadings-two-methods-to-improve-the-measurement-of-edu
Schröder, H., Ganzeboom, H.B.: Measuring and modelling level of education in European societies. Eur. Sociol. Rev. **30**(1), 119–136 (2014)
Shavit, Y., Müller, W. (eds.): From school to work: a comparative study of educational qualifications and occupational destinations. Clarendon Press, Oxford (1998)
Shavit, Y., Arum, R., Gamoran, A. (eds.): Stratification in higher education: a comparative study. Stanford University Press, Stanford (CA) (2007)
Slomczynski, K.M., Tomescu-Dubrow, I.: Basic principles of survey data recycling. In: Johnson, T.P., Pennell, B.E., Stoop, I.A., Dorer, B. (eds.) Advances in comparative survey methods, pp. 937–962. Wiley, Hoboken (2018). https://doi.org/10.1002/9781118884997
Smith, H.L., Garnier, M.A.: Scaling via models for the analysis of association: social background and educational careers in France. Sociol. Methodol. **17**, 205–245 (1987)
Snijders, T.A.B., Bosker, R.J.: Modeled variance in two-level models. Sociol. Methods Res. **22**(3), 342–363 (1994)
Survey Data Harmonization Team: Slomczynski KM, Jenkins JC, Tomescu-Dubrow I, Kołczyńska M, Wysmułek I, Oleksiyenko O, Powałko P, Zieliński M: Master file documentation. SDR Master Box Version **1**, (2017). https://doi.org/10.7910/DVN/VWGF5Q
Treiman, D.J.: Problems of concept and measurement in the comparative study of occupational mobility. Social Sci. Res. **4**(3), 183–230 (1975)
Triventi, M.: The role of higher education stratification in the reproduction of social inequality in the labor market. Res. Social Stratification Mobil. **32**(1), 45–63 (2013a)
Triventi, M.: Stratification in higher education and its relationship with social inequality: a comparative study of 11 European countries. Eur. Sociol. Rev. **29**(3), 489–502 (2013b)

UNESCO Institute for Statistics: International Standard Classification of Education: ISCED 1997 (Re-Edition). UNESCO, Montreal (2006)
UNESCO Institute for Statistics: International Standard Classification of Education - ISCED 2011. UNESCO Institute for Statistics, Montreal (2012)
UNESCO Institute for Statistics (2021) ISCED Mappings. http://uis.unesco.org/en/isced-mappings
van de Werfhorst, H.G.: Vocational and academic education and political engagement: the importance of the educational institutional structure. Comp. Educ. Rev. **61**(1), 111–140 (2017)
van de Werfhorst, H.G., Mijs, J.J.B.: Achievement inequality and the institutional structure of educational systems: a comparative perspective. Annual Rev. Sociol. **36**, 407–428 (2010)
Wilkinson, M.D., Dumontier, M.: Comment: the FAIR guiding principles for scientific data management and stewardship. Scientif. Data **3**, 1–9 (2016)
Winship, C., Mare, R.D.: Regression models with ordinal variables. Am. Sociol. Rev. **49**(4), 512 (1984)
Wolf, C., Schneider, S.L., Behr, D., Joye, D.: Harmonizing survey questions between cultures and over time. In: The, S.A.G.E. (ed.) Handbook of survey methodology, pp. 502–524. Sage, Los Angeles (2016)
Wysmulek, I., Oleksiyenko, O., Powalko, P., Zielinski, M.W., Slomczynski, K.M.: Towards standardization: target variable report template in the harmonization project. Ask **24**(1), 85 (2015)