European Sociological Review VOLUME 30 | NUMBER 1 | 2014 119–136

DOI:10.1093/esr/jct026, available online at www.esr.oxfordjournals.org
Online publication 29 October 2013

Measuring and Modelling Level of Education in European Societies

Heike Schröder* and Harry B. G. Ganzeboom

Abstract: We present two methods to improve the comparative measurement of level of education. The first method derives optimal scale scores for all country-specific education categories distinguished in the European Social Survey Round 1–4 [ESS R1–R4]. This results in a novel continuous comparative education measure that we label ISLED: the International Standard Level of Education. The second method further improves measurement quality by modelling level of education as a true-score latent variable that is reflected in two observed indicators. In particular, we combine ISLED and a common-denominator harmonization based on the International Standard Classification of Education (ISCED), respectively, with an independently collected duration measure. Embedded in an extended intergenerational status attainment model, this allows us to compare the measurement quality of ISLED with that of two often used comparative education measures: duration and ESS's five-category harmonized qualification indicator. ISLED outperforms both by some margin, but still attenuates measurement by 5 per cent. Full disattenuation can, however, be achieved by means of latent variable modelling, as this brings about correction of all (random) measurement error.

Introduction

In this article, we examine ways to improve the measuring and modelling of level[1] of education in international comparisons. Being a core variable in many empirical problems and the pivotal dimension of stratification in modern societies, research should treat the measurement of education level with ultimate care. Yet, an examination of existing cross-national surveys reveals an astonishing lack of comparability. Surveys use a large variety of classifications with differing numbers of education categories. Cross-national comparability is obtained (in pre- or post-harmonization) by calling on a usually overly crude common denominator approach or by using a duration measure, both of which, as we will show later in the text, underestimate the role of education considerably.

We present two methods for the cross-national comparative measurement of education level that allow us to preserve and effectively use all the available information in existing education data—even if variations in measurement occur within a country—as well as to estimate models that tap effects of the true level of education with correction for any (random) measurement error. Both methods are applied to data of the European Social Survey (ESS R1–R4).

With the first method we develop by means of optimal scaling a novel continuous comparative measure of level of education, the International Standard Level of Education (ISLED). This measure quantifies the relative value of individual country-specific education categories in the ESS. The development of ISLED is grounded in two sociological theories, the status attainment model (Blau and Duncan, 1967) and positional good theory (Hirsch, 1976), and further supported methodologically by classic measurement theory (Kelley, 1973). These theories lead us to conceive of level of education as a single intervening variable in an indirect-effects model, in which social background produces social outcomes via education. Optimal scale scores are derived within this model by maximizing the indirect effect of social background on social outcomes via education and minimizing the direct effect.

With the second method, we analyse level of education as a true-score latent variable in the same indirect-effects model, but now combining ISLED, as well as an

Department of Sociology, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands. *Corresponding author. Email: h.schroder@vu.nl.

 Submitted: March 2012; revised: August 2013; accepted: August 2013.

often-used five-category international harmonization, with a duration measure, in a latent variable model. This allows us to assess and compare the measurement quality of all three comparative indicators of education level in the ESS and shows ISLED to outperform both its competitors by a wide margin (11–14 per cent). Despite its high quality, ISLED still causes an attenuation of 5 per cent. Unattenuated measurement can, however, be achieved in the latent variable model, as this brings about the correction of all (random) measurement error.

The Comparative Measurement of Education Level: State of the Art

There are several related reasons for the rather unsatisfactory state of the art in the comparative measurement of education level. The main reason is probably the complexity of the task at hand. After all, what needs to be accommodated is a staggering diversity of education systems, which do not only vary across countries but also over time. Due to the crucial role of education in modern society, countries keep reforming their education systems, forever increasing or decreasing the number of different school types and programmes, abolishing some and adding others. Unlike with occupations, educational differentiation is primarily driven by path-dependent institutional developments that make national education systems highly idiosyncratic (Allmendinger, 1989; Shavit and Müller, 1998). Comparative measurement in cross-national designs and also in a historical perspective, to some extent, always means comparing the incomparable.

While the diversity in education is undoubtedly the heart of the problem, it is only part of the story. Many attempts have been made to implement standardized comparative measures, mostly based on common denominator harmonization (discussed later in the text). Unfortunately, different projects have opted for different standards and even where the same standard classifications are used, they have been implemented in different ways. Whatever measure is chosen, it has consequences in terms of attenuation. These consequences vary between measures and differ in severity but need to be addressed if we want to obtain unbiased statistical results. In our view, the comparability problem can best be understood as a measurement problem and solutions be drawn from classic measurement theory. In his article on causal modelling, Bentler (1980) conceptualizes measurement as a common factor model, in which a latent true score is reflected in multiple observed indicators, with differing measurement quality. The quality of indicators can be estimated and is expressed in their respective measurement coefficients (factor loadings). If we want to produce better measurement quality and hence improve on the state of the art, we first need to specify what the methodological principles are behind the various comparative measures and which consequences they each have for empirical outcomes. We discuss three such principles as found in the literature: common denominator harmonization, duration, and scaling.

Common Denominator Harmonization

The probably most frequently used method of measuring education level in cross-national surveys is harmonization by largest common denominator. The idea here is that different education systems can be made comparable by looking for equivalent elements. The difficulties with this approach are easily anticipated. First, such a strategy leads to loss of information, as any common denominator by definition contains fewer categories than the source classifications to be harmonized. Second, for some categories, it is simply not possible to find a common denominator and incomparabilities can at best be solved by compromise. Third, these difficulties increase with the number of source classifications to be harmonized. After all, the largest common denominator of 10 different classifications is cruder than that of three, and the likelihood of finding unharmonizable elements increases accordingly.

A widely used common denominator is the International Standard Classification of Education [ISCED], developed and maintained by UNESCO (2006). ISCED[2] is a very detailed and comprehensive taxonomy, which is meant to provide an ‘integrated and consistent statistical framework for the collection and reporting of internationally comparable education statistics’ (OECD, 1999: p. 7). In practice, however, the way ISCED tends to be implemented in comparative surveys produces a coarse educational distribution, rather than a detailed classification. Schneider (2009, 2010) and Schneider and Kogan (2008), among others, have evaluated the quality of ISCED-97 (OECD, 1999) and the way it is applied in the ESS and list a large number of problems. One general problem they find is that the ISCED-97 main categories contain insufficient differentiation. In particular, there is no distinction between vocational and academic programmes in secondary and tertiary education, which is, for example, relevant for the German and many other European education systems (Schneider and Kogan, 2008). Moreover, for many countries, coding into ISCED-97 is not consistent across different rounds of ESS (Schneider, 2009: pp. 101–133).

When assessing approaches to common denominator harmonization, it is useful to distinguish between pre- and post-harmonization. Pre-harmonization means that a standard classification such as ISCED is directly applied in the question format in the survey. This strategy is undesirable, as any reference to the underlying original local categories is lost forever and this information loss is beyond repair. In the ESS, this is, for example, the case for the United Kingdom in the first rounds. In post-harmonization, surveys ask for locally relevant categories, which are subsequently recoded (aggregated) into the common denominator. In principle, post-harmonization is equally damaging in terms of information loss, but because the local source categories remain accessible, the lost information can be restored, making it possible to detect coding errors or to exploit the extra detail of the local categories for analytical purposes. For these reasons, asking country-specific categories in a questionnaire and post-harmonizing them deserve preference. This has been the practice for most, but not all, countries in the ESS.

The mere availability of country-specific categories, however, in no way guarantees that this information is actually used. In fact, users of surveys like the ESS, seemingly intimidated by the great variety of distinctions and labels they are confronted with, tend to ignore them and confine themselves to the familiar common denominator variables. Schneider (2009), who illustrates how much can be gained in terms of explanatory power if country-specific categories are coded into CASMIN or EISCED, two alternative common denominator harmonizations, is a notable exception.

Duration

A simple alternative method to compare education across countries is to ask respondents about the duration of their educational career, effectively assuming that duration increases with the level achieved. The questions posed refer to either the school-leaving age or the number of years spent in education. Using duration as a comparative indicator of education level is straightforward, but has some drawbacks. Hout and DiPrete (2006) argue that duration works reasonably well for horizontally undifferentiated ('comprehensive') education systems, such as in the United States, but is much less suited to capture the distinctions of the tracked education systems found elsewhere. In the European context, Schneider (2009: p. 29) and Müller (2008) question the validity of the duration measure. Given the identical length of very different types of educational programmes within and across countries, they argue that confining measurement to duration amounts to concealing qualitative differences between them. Schneider (2009: pp. 452–454) also shows (and we will confirm this later) that the measurement quality of duration measures is lower than that of the detailed categorical measures.

Using a duration measure, however, also has some important advantages. First, it exploits a feature that is intrinsically present in the organization of any education system, namely, that it takes time to pass to higher levels: you cannot start your career at more advanced levels. Second, taken over the entire distribution, duration is strongly correlated with any other indicator of education level and can therefore be interpreted as an independent measurement method. Third, duration has a meaningful analytical interpretation, as it is directly related to human capital accounts of education: irrespective of what is being learned, duration captures how long students forgo current earnings to invest in future earning capacities. Fourth, and very importantly, duration has a metric that is directly comparable across systems with no further transformation needed. Duration questions are particularly simple to ask in comparative surveys, a point that is dramatically confirmed by the treatment of education variables in the ESS: while many changes have occurred in the country-specific measures and common-denominator harmonizations (partly because with hindsight these turned out to be error-ridden), the ESS duration measure has stayed the same in all rounds and countries. For this reason, we consider the availability of the duration measure in the ESS as very valuable and exploit it as a calibration variable to derive a comparative metric for ISLED.

Scaling

A third strategy to make education categories comparable is via common scaling. We can distinguish between ad-hoc and empirical scaling methods. One widespread ad-hoc scaling method is to base the scaling on the number of years it takes to achieve a given level according to the institutionalized education system. This is conceptually similar to, but in practice rather different from, using an independent duration measure. Appropriate institutional duration measures are provided in the *Education at a Glance* publications of OECD (2011) as well as in the manual on the implementation of ISCED-97 (OECD, 1999). In the International Stratification and Mobility File (Ganzeboom and Treiman, 2012), for example, a variable is provided that expresses local categories in 'pseudo-years' of education. A related approach is proposed by Hoffmeyer-Zlotnik and Warner (2007), who organize education categories in four countries into 10 different levels that can be regarded as an ordinal hierarchy. Like scaling by 'pseudo-years', this approach is ad-hoc and non-empirical.

An example of an empirically based scaling method is the so-called *effect-proportional scaling*, where scale scores are generated by maximizing the correlation between given education categories and an output criterion variable (e.g., occupation or income). In their early comparison of the US and British status attainment regimes, Treiman and Terrell (1975), for example, derive comparative education scores, using an output variable (respondent's occupation) as a criterion. They motivate this choice with the argument that it is a primary function of the education system to prepare individuals for employment and that the correlation between education and occupation is particularly strong. By contrast, Smith and Garnier (1987) generate an education scale using an input (father's occupation) as criterion variable. Like respondent's occupation, father's occupation too is strongly associated with education level. Consistent with the term effect-proportional scaling used with output criterion variables, we suggest to label scaling with input criterion variables as *cause-proportional*. No matter how it is done, empirical scaling has the potential of using all available information.

Scaling has not, however, remained without criticism either. Braun and Müller (1997), for example, contend that in effect-proportional scaling, we have to assume that the explanatory power of the respective country-specific measurements is comparable. Also, the criterion variable would have to be measured in a strictly comparable metric, which transfers the problem of deriving a comparable education metric to the criterion variable. We question the validity of this argument because even if the criterion variables are poorly measured, this does not necessarily affect the ordering of the education levels nor the relative distances between them. Braun and Müller (1997) confuse pattern and strength of association.

Measuring and Modelling Level of Education in an Indirect-Effects Model

Our approach to the comparative measurement of level of education consists of two separate methods. With the first method, we measure the value of the education categories contained in the country-specific variables by means of optimal scaling. With the second method, we model level of education in a double-indicator latent variable model.

Method 1: Measuring Level of Education Via an Optimal Scaling Procedure: ISLED

The basic model for our optimal scaling procedure is shown in Figure 1. Here, discrete (national) education

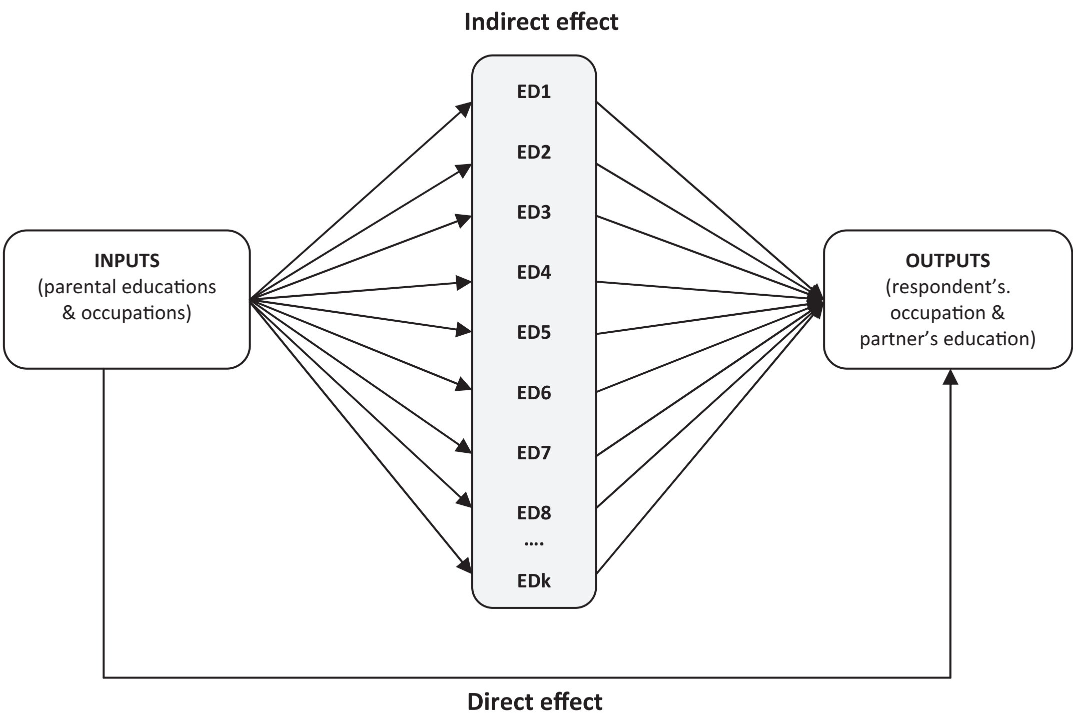

Figure 1 Measuring education levels: an optimal scaling procedure. *Note*: ED1, ED2, etc., are the respective categories within a national classification of education.

categories are interpreted as intervening between inputs and outputs in the stratification process.

The input variables tap parental resources that condition offspring's attained education level. Education level for its part affects multiple output variables. Education is thus understood as the mechanism converting social resources into social outcomes, whereby the value of educational qualifications is on the one hand revealed by the rewards they yield in the labour and marriage markets and on the other hand by the appeal they have for the social status groups competing for them. We merge these two different but related aspects and optimally scale education categories such that the indirect effects of inputs on outputs via education are maximized and the direct effects are minimized. In other words, education level is operationally defined as the scaling of education that best accounts for the conversion of social resources into social outcomes. In our scaling procedure, we follow-up on previous scaling approaches, but improve on them in two ways. First, we integrate cause- and effect-proportional approaches in one unified model. Rather than choosing between either input (Smith and Garnier, 1987) or output variables (Treiman and Terrell, 1975), we combine the two. Our approach is therefore labelled *cause-and-effect-proportional scaling*. Second, instead of confining ourselves to single criterion variables, we use several inputs and several outputs. This makes that the scaling is not crucially dependent on specific patterns of educational attainment, occupational achievement, or homogamy. Finally, we solve the comparative metric problem independently by using duration as a calibration measure.

Method 2: Modelling Level of Education in a Latent Variable Model

With the second method, we again analyse the role of education as intervening variable in the status attainment process, but now model education as a latent variable with two indicators. Provided that the indicators are collected independently, the latent variable model makes it possible to identify the unique true-score information as well as to estimate and correct the measurement error in each indicator. Latent variable modelling can lead to further improvement of the measurement quality of the education variable, over and above optimal scaling and can also be applied independently.

Theoretical Backgrounds

Our procedures find theoretical support in two substantive theories on the role of education in society: the status attainment model (Blau and Duncan, 1967) and positional good theory (Hirsch, 1976). Methodological support can be derived from classic measurement theory.

The Status Attainment Model

A first theoretical anchor for our scaling procedure is provided by the intergenerational status attainment model (Blau and Duncan, 1967). The model shows education to be the pivotal mechanism in intergenerational status transfer. This not only provides a clear conceptualization of the role of education in society, but also a theoretical rationale for our choice of criterion variables. On the output side of the model, we adopt both respondent's occupational status and partner's education level as criteria. While occupational status is the only output variable in the classic model, status attainment research has solidly shown that around the world, the education level of an individual is also strongly associated with that of his/her partner (Smits, Ultee and Lammers, 1998). For this reason, we have included partner's education level in our model as well. On the input side, we work with the combined effects of father's and mother's education levels and occupational statuses. We use information on both parents because there is ample evidence that both fathers' and mothers' educations and occupations strongly affect a person's educational attainment (Korupp, 2002).

Positional Good Theory

A second theoretical anchor for our scaling procedure is derived from positional good theory (Hirsch, 1976; Ultee, 1980). While material goods can in principle be produced in unlimited quantities, positional goods are of fixed supply. Educational qualifications may be classified as positional goods and should be interpreted as relative positions. Positional good theory argues that education systems at all times and places, regardless of their institutional differences, have in common that they are hierarchically organized and allocate people to positions in a single rank-order. This hierarchical rank-order of individuals corresponds to the theoretical notions of job queue (Thurow, 1975) or, in the case of assortative mating (Kalmijn, 1994), candidate queue. The position of individuals in this hierarchy is determined by the relative value of qualifications. In case of increased demand, however, educational qualifications may become subject to congestion or crowding ('credential inflation'). It is therefore not the absolute value of a person's education that counts, but its relative value compared with that of competitors in the queue. This logic provides the rationale that in all countries, education levels are hierarchically ordered on a one-dimensional scale, which informs the derivation of ISLED.

Classic Measurement Theory

Methodologically, our optimization approach can be justified as follows. In an indirect-effects model, the total effect of inputs on outputs is the sum of the direct and the indirect effects. How much of the total effect is direct and how much is indirect crucially depends on the quality of the mediating variable (Kelley, 1973). If the mediator is poorly measured, the direct effect is over-estimated, while the indirect effect is underestimated. By this reasoning, minimizing measurement error equals minimizing the direct effect. We apply this argument to the optimal scaling of education categories. Starting from the assumption that the size of the real direct effect is an empirical matter, we conclude that if we want to establish its true size, we need to filter out the part of the direct effect that is caused by measurement error. As the size of this effect is inversely proportional to the amount of measurement error in the education variable, a scaling that yields larger direct effects of inputs on outputs and weakens the mediating role of education in the status attainment process is suboptimal. A scaling, by contrast, that maximizes the intervening role of the education level and minimizes the direct effect contains the least amount of measurement error and therefore yields the best measure.

Data Sources and Constructed Variables

The ESS is a high-quality survey, which has been held biennially in 34 European countries starting in 2002. We use the data of the first four rounds, referred to as ESS R1–R4, with some 198,000 available cases. On the basis of our selection criteria, excluding respondents <25 and >74 years of age as well as students and respondents without valid education data, we obtain an effective sample of 150,567 cases.

The ESS research design calls for two independent measurements of level of education: a country-specific classification and a duration measure. As the country-specific classification is post-harmonized into a five-category version of ISCED-97, the two measurements yield three different indicators in the data: the country-specific measures, their ISCED-based harmonization, and the duration measure. The presence of two independent measurements of education level in the ESS allows us to apply latent variable modelling, which is required for a direct comparison of the three indicators as well as for the correction of measurement error. The ESS data are, moreover, particularly well suited for our purposes due to their richness in criterion variables.

The ESS Education Variables

Until ESS Round 5, it has been one of the policies of the ESS to leave countries the option to use country-specific education typologies, which serve as source variables for post-harmonization. Countries were not instructed how to design their education showcards, but could use their own formats, the only requirement being that it could be recoded into the (seven) main ISCED-97 levels. For most countries, the country-specific measures have been included in the main ESS data file. We will refer to them as EDLVXX, as for R1–R4, the names of these variables in the ESS are EDLVAT . . . EDLVUA (XX is replaced by the ISO country abbreviations AT (Austria) to UA (Ukraine)). Across all countries, we found 1,154 individual categories.

An inspection of these variables reveals a number of problems. First, countries have interpreted the recommended strategy in different ways. For Austria, Finland, Iceland, Slovenia, Turkey, and the United Kingdom, no detailed country-specific measure is available, or at least not for all rounds. The Irish, Italian, and Ukrainian country-specific measures are available but turn out to be identical to their ISCED equivalent in at least one round. Second, for the remaining countries, the number of categories distinguished in the country-specific measure varies from 5 for the United Kingdom to 19 for Luxembourg. This illustrates that the detail available and hence the information that can be lost in the harmonization process varies considerably between countries.

A third problem in processing the country-specific variables is that many countries have changed their variables between rounds. In fact, there are only three countries (Germany[3], The Netherlands, and Sweden) that have not made any changes over the four rounds. Fortunately, the changes that have been made are easy to track, as ESS flags them by adding a character to the variable name: for instance, the variable names EDLVCH, EDLVaCH, EDLVbCH, and EDLVcCH indicate that Switzerland changed its measurement system with each new round. Changes can be characterized as either *splits* or *mergers*: splits occur when a category is divided into two or more branches and mergers when two or more categories that were distinct in one round are collapsed in a subsequent round. As can be seen in our online appendix (ISLED, 2012), we have processed this information by organizing it in a hierarchical digit system: if a category (say 4) is split in a new round, we refer to its branches as 4.1, 4.2, etc. If subcategories present in one round are merged in the next round, the reverse occurs and digits disappear. Not all changes, however, are simple splits or mergers. In particular, in the Estonian, French, and Swiss cases, 'layered' splits occur, meaning that divisions created in one round are further split in the next round. This required the use of second and occasionally even third digits. A particular convenience of this digit system is that it allows us to estimate the level of the cruder category by averaging it with the individual values of its more detailed branch categories.

The largest common-denominator strategy used by ESS is derived from and documented by ISCED-97. Until 2011, the ISCED measure in the ESS used to be called EDULVL and contained seven categories (0 Less than primary, 1 Primary, 2 Lower Secondary, 3 Upper Secondary, 4 Post-secondary, 5 First stage of Tertiary, 6 Second stage of Tertiary). As ISCED-levels 0 and 1 as well as 5 and 6 could not always be properly identified due to a lack of differentiation in the country-specific source variables, these distinctions could not be maintained consistently across countries. For this reason, in a 2011 revision[4] of the data, EDULVL was replaced by a five-category harmonization, EDULVLa. In EDULVLa, levels 0 and 1 as well as levels 5 and 6 of the former EDULVL[5] are merged.

For the respondent, the ESS data contain an independent second education measure: duration (EDUYRS). The question asked is about the number of years (in 'full-time equivalents') the respondent has spent in education. For our analyses, we have truncated EDUYRS at 24 years to exclude improbably long durations.

The Criterion Variables

For the input side of the model, we have chosen for parental occupations and education levels, yielding four variables. The ESS data of R1–R3 include only one indicator of father's and mother's education, the harmonized ISCED, stored as variables EDULVLFa and EDULVLMa. While the harmonization process was the same as for the respondent, here the country-specific source variables were not archived. In R4, many countries have complemented the harmonized variables with country-specific measures, but to preserve comparability, we have not taken this change into account. For parental occupations, two indicators are available in ESS: a crude precoded measure (OCCF14 and OCCM14, with two revisions) and a detailed code, measured in ISCO-88 (ISCOCOF and ISCOCOM).[6] To process this occupational information, we have converted all of it into the International Socio-Economic Index of occupational status (ISEI) (Ganzeboom, de Graaf and Treiman, 1992; Ganzeboom and Treiman, 1996). We averaged the two ISEI indicators for fathers and mothers, respectively, before using them as criterion variables.

For the output side of the model, we have chosen respondent's occupation and partner's education. The respondent's occupation too is measured in ISCO-88 (ISCOCO), which we have converted into ISEI scores. The education level of the partner is again measured in a harmonized ISCED format (EDULVLaP).

Algorithm and Models

Optimal Scaling

The algorithm we use to find the optimal scaling of education categories, our first method, is a variation of the algorithm used for the development of the ISEI index for occupational status, which was developed for this purpose by De Leeuw (in Ganzeboom, de Graaf and Treiman, 1992). Here occupational status was defined and calculated as the optimal scaling of occupations: ISEI is the scaling of occupations that mediates best the influence of education on income. In our application to derive ISLED, we first reduce complexity by assembling the unweighted averages of the standardized input variables and then of the standardized output variables in two composite indices. To lose as little information as possible, we have used an available-cases strategy, meaning that the criterion indices average whatever is available as inputs or outputs. The optimal scale score is a weighted average of the Z-standardized composite inputs and composite outputs; the optimal solution is found by updating the relative weights of the input and output composites. This is done by systematic search in a few iterative steps in an Ordinary Least Squares regression. The search stops when the remaining direct effect of inputs on outputs is at a minimum. In the ESS data, this happens to be the case for 0.61 (inputs) and 0.39 (outputs). These weights are constrained to be the same for all countries and all rounds.[7]

The resulting optimal scores are initially Z-standardized within countries, which makes levels of education comparable within but not across countries. The within-country standardized metric may satisfy many needs (in particular when doing analyses on a country-by-country basis, or pooling an analysis of multiple countries), but will not allow the analyst to compare means and dispersions between countries, or to control for educational composition in a cross-national analysis. To accomplish these goals, a common cross-national metric needs to be established. We define this metric by calibrating the optimized scale on an external measure, the duration measure EDUYRS, which is available in all ESS rounds and for all countries. Despite its somewhat poorer measurement quality (see later), the duration measure has the advantage of producing a remarkably stable image of between-country variations in the underlying educational distributions and of having directly comparable means and dispersions across contexts.[8]

The procedure to develop the comparable metric consists of two steps:

Calibration: Define an intermediate metric Z^* by equalizing country-specific means and dispersions between the estimated optimal scale and the duration measure:

$$Z^* = (Z(OPTI) + M_c(Z(EDUYRS))^* SD_c(Z(EDUYRS)),$$

in which M_c and SD_c represent the country-specific means and standard deviations of the duration distribution, respectively, and OPTI is the within-country optimal scale score.

Transformation: After restandardizing Z^* into an overall Z, we project back into a 0..100 metric using the anti-logistic transformation:

$$ISLED = 100^*(\exp(Z)/(1 + \exp(Z)).$$

As a result of the calibration step, country-specific means and dispersions of the ISLED distributions are proportional to those of the duration measure. As a result of the transformation step, the final scores range between 0 and 100.[9] The anti-logistic transformation (Hauser and Warren, 1997) is preferred over a linear transformation because by reducing differences in scores at either extreme of the scale, we avoid out-of-range projections. Note that duration is only used to define the overall metric of the score distribution, but does not determine the relative distances between the score values of education categories within countries—these are solely determined by the association with the criterion variables. The ISLED scores thus obtained are now comparable within and between countries and can be interpreted as giving an indication of the relative value of educational qualifications (in Europe in as far as being represented in the ESS). We label them ISLED, the International Standard Level of Education, because the scores are comparable between countries and designate the value of each and every education level represented by the individual country-specific categories that we have scaled on a one-dimensional international educational hierarchy.

Table 1 presents the means and standard deviations of ISLED per country, together with those of the duration measure on which they are based.

Mean levels of education are fairly similar between most European countries, with Portugal and Turkey as striking exceptions. Iceland has the highest educated population (58.7), closely followed by Norway (57.4) and Italy (43.4). Greece (40.8) and Spain (43.4) trail behind the main pack of countries. Portugal (25.2) and Turkey (22.4) are outliers by a substantial distance. The dispersions vary much more between countries. The greatest contrast can be found between the Southern European countries with wide distances between the lower and higher educated and Eastern European countries with much more compressed educational distributions.

Latent Variable Modelling

In the modelling part of our analysis, where we apply our second method, we examine the measurement

Table 1 Average level of education per country, ESS R1–R4 (age 25–74)

ISO	Country	*N*	EDUYRS		ISLED	
			Mean	SD	Mean	SD
AT	Austria	5,177	12.5	3.1	51.8	16.9
BE	Belgium	5,388	12.6	3.7	52.1	20.3
BG	Bulgaria	2,940	11.3	3.5	46.6	18.5
CH	Switzerland	6,279	11.7	3.7	48.0	18.9
CY	Cyprus	1,787	12.0	3.8	48.6	20.7
CZ	Czech Republic	5,140	12.5	2.5	51.7	13.7
DE	Germany	8,849	13.5	3.3	56.0	17.2
DK	Denmark	4,767	13.5	4.1	56.8	20.8
EE	Estonia	3,809	12.7	3.2	52.7	17.2
ES	Spain	5,903	11.4	5.1	43.4	25.3
FI	Finland	6,083	12.9	4.0	54.2	20.9
FR	France	5,799	12.4	4.0	50.9	20.8
GB	United Kingdom	6,605	13.2	3.5	55.4	18.9
GR	Greece	5,561	10.5	4.4	40.8	22.6
HR	Croatia	1,119	11.7	3.7	48.3	19.8
HU	Hungary	4,810	12.0	3.6	49.0	19.0
IE	Ireland	6,146	13.1	3.6	54.8	18.9
IL	Israel	3,463	13.3	3.6	55.5	19.5
IS	Iceland	420	13.8	4.3	58.7	21.4
IT	Italy	2,163	10.9	4.5	43.4	22.9
LT	Lithuania	1,516	12.8	3.3	53.1	17.8
LU	Luxembourg	2,283	11.8	4.2	48.1	21.4
LV	Latvia	2,868	12.4	3.4	51.3	18.3
NL	The Netherlands	6,511	13.1	4.0	55.2	20.7
NO	Norway	5,498	13.7	3.7	57.4	19.1
PL	Poland	5,188	12.0	3.4	49.3	18.3
PT	Portugal	6,115	7.8	4.8	25.2	22.4
RO	Romania	3,252	11.4	3.7	45.9	20.0
RU	Russia	3,704	12.6	3.0	52.2	16.6
SE	Sweden	5,701	12.8	3.5	53.3	19.1
SI	Slovenia	4,207	11.7	3.6	48.3	19.3
SK	Slovakia	3,809	12.5	3.1	52.1	16.4
TR	Turkey	3,181	6.2	4.2	22.4	18.9
UA	Ukraine	4,526	12.1	3.3	50.0	17.9
Total/average		150,567	12.1	4.0	49.7	21.0

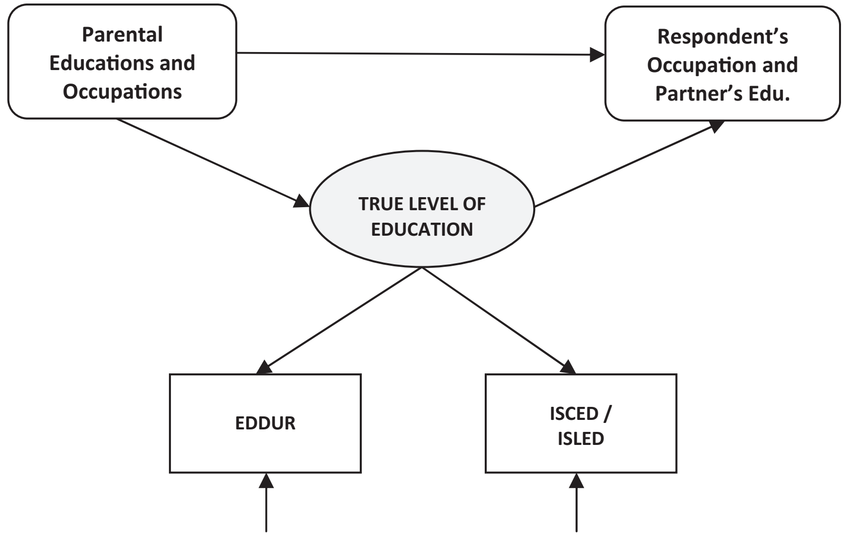

Figure 2 Modelling education levels: a latent variable indirect-effects model. *Note*: EDDUR = duration of education; ISCED = International Standard Classification of Education; ISLED = International Standard Level of Education. The duration measure is the first indicator in all models, while the second indicator is ISCED or ISLED, respectively.

quality of education indicators using an indirect-effects latent variable model (Figure 2). The model consists of two parts, a measurement part and a structural part. The measurement model illustrates how the latent education variable (represented by an oval) is reflected in two indicators (represented by rectangles). In one model we combine EDUYRS with ISLED, in another with EDULVLa (the ISCED-97-based harmonization in ESS R1–R4). The measurement model with two indicators is not identified in itself, but becomes identified when we embed it in a structural model, by including input and output variables, the criterion variables introduced earlier. We estimate the parameters using Full Information Maximum Likelihood in LISREL 8.8 (Jöreskog and Sörbom, 1996). The Full Information Maximum Likelihood approach computes a casewise likelihood function using only those variables that are observed for a given case (Enders and Bandalos, 2001). The estimates of the parameters are weighted with the N of the pertinent correlations: if the estimate is based on a larger N, it gets a small standard error, whereas an effect that models correlations with a smaller N gets a large standard error.

We can assess the relative measurement quality of indicators by comparing their measurement coefficients. These are inversely related to the (attenuated) size of the structural coefficients in single-indicator models and can be interpreted as (dis)attenuation coefficients. For reasons of exposition, we precede the double-indicator models by showing the three separate single-indicator models. We can assess measurement quality by comparing the explained variance *in* education level and the explained variance *by* education level, with higher amounts of explained variances signifying better indicator quality. A related way of assessing indicator quality is the comparison of effect sizes. The smaller the direct effects of inputs on outputs and, by the same token, the larger the indirect effects via education, the better the indicator. The double-indicator latent variable model allows us to diagnose and correct random measurement error. As even a small amount of random error may have large consequences in terms of the attenuation of structural coefficients (Allison and Hauser, 1991), error correction should be worth our while. We will show that this is the case here as well. The latent variable model takes our principle of fully using all available information one step further. Rather than restricting ourselves to a single indicator, we make use of the extra information contained in a second indicator. Provided that they are based on independent measurement, even suboptimal measures will contribute some information that is not tapped by the other indicator.

Results

We present our findings in two sections. In the first section, we discuss one country, Germany, in detail, while in the second section, we briefly discuss the results across countries.

Table 2 Summary of the German country-specific education categories: ESS R1–R4 ($N = 8{,}849$)

1 Cat	2 ISCED	3 *N* R1	4 *N* R2	5 *N* R3	6 *N* R4	7 Country-specific education category	8 OPTI R1	9 OPTI R2	10 OPTI R3	11 OPTI R4	12 OPTI	13 ISLED
0	0	42	61	48	44	Grundschule nicht beendet[a]	−1.622	−1.640	−1.471	−1.551	−1.568	**26.9**
1	2	104	108	90	76	Hauptschule	−1.328	−1.364	−1.353	−1.376	−1.344	**37.4**
1.1	3	18	23	34	19	Hauptschule + *Beruflich-betriebliche Anlernzeit mit Abschlusszeugnis, aber keine Lehre*	−1.637	−1.508	−1.447	−1.307	−1.460	**28.7**
1.2	3	20	28	13	22	Hauptschule + *Teilfacharbeiterabschluss*	−1.161	−1.645	−1.223	−1.619	−1.449	**28.9**
1.3	3	346	324	364	281	Hauptschule + *Abgeschlossene gewerbliche oder landwirtschaftliche Lehre*	−1.096	−1.046	−0.997	−1.233	−1.079	**35.4**
1.4	3	129	138	126	115	Hauptschule + *Abgeschlossene kaufmännische Lehre*	−0.623	−0.405	−0.522	−0.747	−0.565	**45.5**
1.5	2	5	6	2	3	Hauptschule + *Berufliches Praktikum, Volontariat*	−0.759	−0.821	−1.865	−0.474	−0.795	**40.9**
1.6	5	34	29	29	16	Hauptschule + *Fachschulabschluss*	−0.494	−0.598	−0.092	−0.481	−0.420	**48.5**
1.7	3	43	30	23	30	Hauptschule + *Berufsfachschulabschluss*	−0.768	−0.678	−0.158	−0.809	−0.642	**44.0**
1.8	5	55	42	41	33	Hauptschule + *Meisterabschluss*	−0.700	−0.500	−1.001	−0.724	−0.726	**42.3**
2	3	43	21	33	60	Realschule	−0.314	−0.784	−0.383	−0.194	−0.338	**54.5**
2.1	3	4	7	11	7	Realschule + *Beruflich-betriebliche Anlernzeit mit Abschlusszeugnis, aber keine Lehre*	0.736	0.224	−0.519	−0.432	−0.125	**54.5**
2.2	3	3	4	6	3	Realschule + *Teilfacharbeiterabschluss*	−0.240	−0.611	−0.500	−1.378	−0.642	**44.0**
2.3	3	247	230	268	233	Realschule + *Abgeschlossene gewerbliche oder landwirtschaftliche Lehre*	−0.351	−0.330	−0.533	−0.573	−0.448	**47.9**
2.4	3	228	206	250	232	Realschule + *Abgeschlossene kaufmännische Lehre*	0.258	0.029	0.072	−0.110	0.057	**58.1**
2.5	2	8	2	9	1	Realschule + *Berufliches Praktikum, Volontariat*	0.491	−0.128	−0.344	2.201	0.115	**59.3**
2.6	5	103	112	91	82	Realschule + *Fachschulabschluss*	0.243	0.142	0.287	0.242	0.218	**61.3**
2.7	3	66	55	74	67	Realschule + *Berufsfachschulabschluss*	−0.209	0.110	0.112	0.333	0.079	**58.6**
2.8	5	79	69	47	67	Realschule + *Meisterabschluss*	−0.088	0.055	−0.274	−0.150	−0.102	**54.9**
3	3	5	3		8	Fachhochschulreife	1.881	0.675		−0.435	0.396	**62.3**
3.1	3		1	2		Fachhochschulreife + *Beruflich-betriebliche Anlernzeit mit Abschlusszeugnis, aber keine Lehre*		−1.199	−0.026		−0.011	**56.8**
3.3	3	8	6	13	8	Fachhochschulreife + *Abgeschlossene gewerbliche oder landwirtschaftliche Lehre*	0.621	−0.344	−0.026	−0.446	−0.011	**56.8**
3.4	4	20	20	14	22	Fachhochschulreife + *Abgeschlossene kaufmännische Lehre*	0.026	−0.131	0.703	0.253	0.171	**60.4**
3.5	5		3		3	Fachhochschulreife + *Fachschulabschluss*		0.235		0.619	−0.011	**56.8**
3.6	4	19	24	13	21	Fachhochschulreife + *Berufsfachschulabschluss*	0.447	0.687	0.487	0.343	0.492	**66.4**
3.7	4	15	10	14	16	Fachhochschulreife + *Berufliches Praktikum, Volontariatpracticum*	0.370	0.385	0.565	0.509	0.453	**65.7**
3.8	5	24	22	5	15	Fachhochschulreife + *Meisterabschluss*	0.354	−0.116	0.665	−0.013	0.140	**59.8**
4	3	12	12	20	25	Abitur	1.222	0.605	0.606	0.765	0.751	**71.1**
4.1	4	1	1	3		Abitur + *Beruflich-betriebliche Anlernzeit mit Abschlusszeugnis, aber keine Lehre*	1.806	−0.429	1.678		0.630	**68.9**
4.3	4	15	11	21	14	Abitur + *Abgeschlossene gewerbliche oder landwirtschaftliche Lehre*	0.863	0.511	0.472	0.573	0.630	**68.9**
4.4	4	29	40	42	47	Abitur + *Abgeschlossene kaufmännische Lehre*	0.733	0.921	0.833	0.748	0.800	**71.8**
4.5	4	1	1	4	1	Abitur + *Berufliches Praktikum, Volontariatpracticum*	−0.967		1.003	0.192	0.630	**68.9**
4.6	5	24	28	20	23	Abitur + *Fachschulabschluss*	0.810	0.609	1.324	0.762	0.830	**72.3**
4.7	4	16	8	10	10	Abitur + *Berufsfachschulabschluss*	0.797	0.320	1.680	0.694	0.874	**73.0**
4.8	5	19	16	9	6	Abitur + *Meisterabschluss*	0.773	0.548	0.342	0.577	0.589	**68.2**
5	5	127	148	163	197	Fachhochschule	0.729	0.929	0.933	0.793	0.836	**72.4**
6	5	77	74	65	100	Hochschule/Universität: Zwischenprüfung, Vordiplom; Bachelor	1.691	1.626	1.609	1.149	1.469	**81.5**
7	5	243	229	211	238	Abgeschlossenes Studium an Hochschule, Universität, Akademie, Polytechnikum (Diplom, Magister, Master, Staatsexamen)	1.729	1.914	1.782	1.602	1.732	**84.5**
8	5	31	27	33	41	Promotion; Habilitation	2.611	2.805	2.961	2.713	2.737	**92.5**

Note: Bold values indicates final results of the calculations.

[a]*The category 'Grundschule beendet', without finishing any further education (ISCED 1) does not occur in the ESS data.*

OPTI-R1 –R2 –R3 –R4: optimal scale scores per round; OPTI: average optimal score.

An Example: Germany

We have chosen Germany as our example because its country-specific variable in the ESS is by far the most complex and in our rendition ultimately the most detailed one too. For Germany, therefore, a maximum amount of information is lost through harmonization, or vice versa can be retained through scaling. This makes the German case particularly well suited to demonstrate the points we are trying to make. By comparison, the country-specific education variables for the other ESS countries are straightforward and self-explanatory.

The German country-specific question has been asked in exactly the same format in each round. It consists of two separate questions: one on the highest school qualification and one on the highest vocational qualification. To exploit all the information contained in the two questions, we constructed a combined variable, which is listed in Column 7 of Table 2[10] (Column 1 assigns a number to each category and Columns 3–6 present the number of selected respondents per round). Category 0 (*Grundschule nicht beendet*), the lowest level, as well as the four tertiary education categories 5–8: (*Fachhochschule, Bachelor, Master, Promotion*) remain undifferentiated. Together with the four secondary school levels (categories 1–4: *Hauptschule, Realschule, Fachhochschulreife, Abitur*), this yields nine main levels in German education. The four secondary levels have each been combined with eight types of non-university vocational training, resulting in potential 32 subcategories. Two of these combinations are not filled and others have very small counts. With as many as 39 effective categories, we have, however, obtained a very detailed variable indeed, which preserves the available information as fully as possible.

The first set of results for Germany can be found in Columns 8–11 of Table 2, headed OPTI-R1–R4, and provide the optimal scale scores per category and round. Per category the scale scores are then averaged across rounds, producing OPTI in Column 12. In Column 13, OPTI is transformed into ISLED, yielding values ranging from 26.9 to 92.5. The results show that the nine main education levels discerned in the German country-specific variable are strictly hierarchically scaled by the criterion variables and correspond to the implicit ordering of the presented answer categories. The values of the sublevels, by contrast, vary considerably per round and do not always follow the nominal hierarchy either. It can, however, be seen that sublevels of the same type (e.g., 1.5, 2.5, 3.5, and 4.5) are hierarchically ordered among themselves. This implies that although the vocational qualifications attained are actually identical, differences in the preceding general education dominate the ultimate scale scores.

Table 3 presents the results of five pertinent simultaneous equation models for Germany, each consisting of three standardized regression equations, with education level, occupation, and education level of the partner being the three dependent variables. Models 1–3 are single-indicator models, which alternately use one of the three different indicators, EDUYRS, EDULVLa, and ISLED. Models 4 and 5 are double-indicator models, whereby Model 4 combines EDUYRS with EDULVLa and Model 5 EDUYRS with ISLED.

When comparing the models, we can consider the measurement coefficients, the explained variance for the separate equations, or the size of the regression coefficients. On all accounts, ISLED turns out to produce the best results. The measurement coefficients for the three indicators in Models 4 and 5 are 0.812[11] (EDUYRS), 0.803 (EDULVLa), and 0.946 (ISLED). These coefficients provide direct insight into the loss of information we suffer per indicator, which can be expressed in percentage points: EDUYRS causes 19 per cent, EDULVLa 20 per cent, and ISLED 5 per cent attenuation of any covariance-based association in the German data.

The better quality of ISLED is already visible in the single-indicator Models 1–3. Here it is reflected in higher levels of explained variance. Compared with EDUYRS, ISLED explains more of the variance in all dependent variables: 5 per cent more in respondent's education, 6 per cent more in partner's education, and 11 per cent more in respondent's occupation. ISLED's quality can also be observed in the effect sizes. Compared with Models 1 and 2, in Model 3 (ISLED), the indirect effect that is mediated by education level is largest: ISLED produces the lowest direct effects of parental educations and occupations on respondent's education, as well as by respondent's education on occupation and education of the partner. Accordingly, the indirect effects of parental education on respondent's and partner's education and parental occupation on respondent's occupation are larger. The differences between EDULVLa and EDUYRS are somewhat less marked, but for Germany, EDUYRS performs better than EDULVLa. EDULVLa's poor performance may be explained by the fact that the new revised harmonization EDULVLa contains very little detail because it lumps most secondary as well as tertiary educations together, disguising distinctions that are particularly relevant in the German case.

In Models 1–3, we have followed other researchers (Kerckhoff and Dylan, 1999; Schneider, 2009) in assessing measurement quality by using single indicators. We now go a step further and examine what happens when we proceed to double-indicator models. In Model 4, EDULVLa is combined with EDUYRS, which has a mixed but only small effect on the explained variance in the dependent variables, which is either marginally higher (respondent's education

Table 3 Model parameters for GERMANY, ESS R1–4 ($N = 8{,}849$)

	Single-indicator models			Double-indicator models	
	1 EDUYRS	2 EDULVLa	3 ISLED	4 EDUYRS EDULVLa	5 EDUYRS ISLED
Structural models					
Independent variables			Dependent variables		
			Respondent's education		
Father's education	0.166	0.152	0.205	0.188	0.209
Mother's education	0.081	0.054	0.091	0.072	0.088
Father's occupation	0.163	0.151	0.176	0.201	0.198
Mother's occupation	0.140	0.119	0.143	0.161	0.195
R^2	0.213	0.161	0.267	0.273	0.302
			Spouse's education		
Father's/Mother's education[a]	0.138	0.164	0.104	0.107	0.082
Respondent's education	0.138	0.324	0.478	0.469	0.527
R^2	0.291	0.260	0.348	0.341	0.377
			Respondent's occupation		
Father's/Mother's occupation[a]	0.097	0.117	0.043	0.039	0.012[b]
Respondent's education	0.500	0.481	0.641	0.645	0.703
R^2	0.354	0.348	0.463	0.464	0.509
Measurement models					
Indicator			Measurement coefficients		
EDUYRS	1			0.862	0.812
EDULVLa		1		0.803	
ISLED			1		0.946
Fit statistics					
RMSEA	0.029	0.033	0.026	0.030	0.026

Completely standardized parameters.
[a]*Effects constrained to be equal.*
[b]*Not significant.*
EDUYRS: duration; EDULVLa: ISCED-harmonization; ISLED: optimal scaling of country-specific measures.

and occupation) or slightly lower (partner's education) than in Model 3. By contrast, Model 5, where we combine EDUYRS with ISLED instead of EDULVLa, improves the results more visibly. Compared with Model 4, the explained variance is higher in all three dependent variables, by an average of 3 per cent. These results show that double-indicator models produce better results than any single indicator and that the combination of EDUYRS and ISLED yields the best results. Effect sizes increase accordingly, with effects of parental educations and occupations on respondent's education, as well as effects of the latter on occupation and partner's education increasing, while the direct effect of parental occupations on respondent's occupation decreases and in fact becomes insignificant. For Germany, it is even true that if education level is measured and modelled appropriately, no such direct effect remains.

Model 5 also shows that EDUYRS, despite being a relatively poor indicator, still contributes some information, even when it is combined with ISLED. If the duration question were only a weak measurement of education level and ISLED a perfect one, the measurement coefficient for ISLED would equal 1.0. A deviation by 5 per cent may not seem much, but it still has a noticeable impact on the estimated coefficients (cf. Model 5 with Model 3). This result shows that duration contains some information relevant to the status attainment process that is unique for this indicator.

Results Across Countries

After having discussed the German case in some detail, we now briefly consider the remaining countries jointly. Table 4 provides the results for the simultaneous equation

Table 4 Model parameters for all countries, ESS R1–4 ($N = 150{,}567$)

	Single-indicator models			Double-indicator models	
	1 EDUYRS	2 EDULVLa	3 ISLED	4 EDUYRS EDULVLa	5 EDUYRS ISLED
Structural models					
Independent variables			Dependent variables		
			Respondent's education		
Father's education	0.191	0.198	0.204	0.218	0.215
Mother's education	0.165	0.155	0.165	0.173	0.173
Father's occupation	0.137	0.128	0.149	0.154	0.161
Mother's occupation	0.093	0.092	0.112	0.109	0.119
R^2	0.234	0.224	0.269	0.291	0.302
			Spouse's education		
Father's/Mother's education[a]	0.130	0.129	0.109	0.091	0.088
Respondent's education	0.423	0.437	0.477	0.523	0.525
R^2	0.325	0.337	0.358	0.388	0.387
			Respondent's occupation		
Father's/Mother's occupation[a]	0.095	0.093	0.059	0.048	0.036
Respondent's education	0.501	0.520	0.594	0.614	0.642
R^2	0.348	0.367	0.420	0.433	0.455
Measurement models					
Indicator			Measurement coefficients		
EDUYRS	1			0.878	0.859
EDULVLa		1		0.892	
ISLED			1		0.949
Fit statistics					
RMSEA	0.021	0.020	0.015	0.016	0.018

Completely standardized parameters.
[a]*Effects constrained to be equal.*
EDUYRS: duration, EDULVLa: ISCED-harmonization; ISLED: optimal scaling of country-specific measures.

models for all countries combined. Like the German table, it reveals how the various indicators differ in measurement quality and how they affect explained variance and effect sizes accordingly. The results are clear and unequivocal: ISLED outperforms EDULVLa and EDUYRS by a considerable margin. For the pooled sample the measurement coefficients are: EDUYRS (measurement: 0.859; attenuation: 14.1 per cent); EDULVLa (measurement: 0.892; attenuation: 10.8 per cent); ISLED (measurement: 0.949; attenuation: 5.1 per cent).

With 5 per cent loss, ISLED produces the best measurement quality of the three single indicators. Note that in contrast with Germany, for all countries combined, EDULVLa performs better than EDUYRS. Also note that both EDUYRS and EDULVLa perform better for the combined than for the German data.

Again we observe that among the single indicators, ISLED (Model 3) produces the largest explained variances in all three dependent variables, which comes with the familiar effect pattern: ISLED produces the largest indirect and smallest direct effects. Just like in the German case, here too, ISLED is outperformed by double-indicator models (Models 4 and 5). In contrast to Germany, for all countries combined, parental occupations continue to have a significant (albeit small) direct effect on respondent's occupation. It must be stressed again, however, that it is double-indicator latent variable modelling that tops off measurement quality.

These findings have an important ramification for the interpretation of the results we achieve using ISLED. If latent variable modelling produces benchmark unattenuated measurement, ISLED on its own does not quite match this result, but comes much closer to it than the other two indicators. This illustrates that although the criterion variables used in the validation model are the same as those used for the derivation procedure, ISLED by no means overestimates education effects, but rather still attenuates them.

Table 5 presents the measurement coefficients for the three indicators per country, which can again be directly translated into attenuation factors.

The results are surprisingly consistent across countries. ISLED outperforms both EDUYRS and EDULVLa in all countries, except for Greece, where EDULVLa just about surpasses ISLED by two points in the third decimal. How well ISLED does for a given country is of course dependent on the country-specific source variables. The differences in quality that we find are attributable to how well the respective country-specific variables represent a given national education system. In this sense, the quality of ISLED is bounded by these source variables. In countries such as France, Estonia, Germany, Luxembourg, and Switzerland, which have the most detailed source variables, the potential gain in measurement quality by using ISLED is most pronounced.

Table 5 Measurement coefficients for the three different indicators by country, ESS R1–4 ($N = 150{,}567$)

ISO	Indicator EDUYRS	EDULVLa	ISLED
AT	0.880	0.921	0.949
BE	0.763	0.941	0.969
BG	0.960	0.956	0.978
CH	0.785	0.790	0.928
CY	0.923	0.961	0.974
CZ	0.849	0.881	0.972
DE	0.812	0.804	0.946
DK	0.780	0.847	0.907
EE	0.903	0.857	0.921
ES	0.880	0.926	0.945
FI	0.884	0.876	0.901
FR	0.836	0.892	0.961
GB	0.785	0.859	0.908
GR	0.930	0.969	0.967
HR	0.863	0.917	0.967
HU	0.880	0.913	0.978
IE	0.835	0.887	0.937
IL	0.892	0.882	0.972
IS	0.808	0.848	0.952
IT	0.953	0.958	0.963
LT	0.880	0.847	0.962
LU	0.878	0.926	0.960
LV	0.832	0.874	0.942
NL	0.784	0.898	0.934
NO	0.872	0.870	0.914
PL	0.937	0.934	0.980
PT	0.958	0.969	0.974
RO	0.918	0.935	0.956
RU	0.909	0.847	0.983
SE	0.898	0.899	0.938
SI	0.868	0.907	0.960
SK	0.774	0.908	0.977
TR	0.936	0.951	0.961
UA	0.767	0.843	0.972
Mean	0.865	0.897	0.953
SD	0.059	0.046	0.023

Conclusions and Discussion

It was our goal to improve on the state of the art of the comparative measurement of education level. We have proposed two complementary but independent methods to achieve this goal. With the first method, we measured the value of each category of the ESS country-specific education variables by means of optimal scaling, resulting in a novel high-quality single indicator: ISLED. With the second method, we modelled education level as a latent variable with double indicators, resulting in unattenuated measurement coefficients. These two methods share a common maxim: the full exploitation of all available information. The first method, optimal scaling, makes use of all extra detail contained in the country-specific education measures. Across all ESS countries, the derived variable ISLED has been found to outperform both EDULVLa (ISCED harmonization) and EDUYRS (duration) by some distance. The second method, latent variable modelling, makes use of the unique information contained in each indicator. This method has been found to have the edge over the standard single ESS indicators available in R1–R4, as well as over ISLED, albeit to a lesser degree. We conclude that together the two methods lead to a significant improvement of the state of the art in the measurement of level of education.

While the results presented here are clear and promising, we acknowledge a number of limitations. Some limitations pertain to the measurement, others to the modelling part of our analyses. Concerning the measurement part, a first limitation is that the analyses have so far been confined to European countries in the ESS. We would like to stress that the intent is to produce ISLED scores for all countries where pertinent data are available. Analyses for a wider range of countries are possible with, for example, the 2009 Social Inequality module data from the International Social Survey Project (ISSP).

A second limitation is that we compared ISLED with EDUYRS and EDULVLa, both of which are known to be of poor quality, which may portray ISLED in too favourable a light. In Round 5, ESS has introduced two new common denominator harmonizations, EISCED and EDULVLb, with which ISLED still needs to be compared. As they are not (EDULVLb) or only partially (EISCED) available for R1–R4, we did not include them in the analyses reported here. We plan, however, to make up for this in a forthcoming study (Schröder and Ganzeboom, 2012b), where we analyse the ESS R5 data.[12]

A third limitation is that we present validation models for ISLED, which use the same data that were used for its derivation. An important way of further testing ISLED would be to apply it to fresh data. First results by Schröder and Ganzeboom (2012c) with the ISSP 2009 Social Inequality data are promising and suggest that ISLED can be successfully applied to fresh data, with compatible education categories. We will continue to test ISLED and invite other researchers to do so as well.

A fourth limitation is that ISLED has not yet been tested with different criterion variables, that is, variables that were not involved in its derivation. Although it was derived within a status attainment model, we expect that ISLED, just like ISEI for occupations, is not limited to social stratification research, but can in principle be applied in any research context. The reason for this is that in essence, ISLED, just like any other education indicator, measures educational resources that are important in determining outcomes, be it in stratification, attitudes and values, cultural participation, or health (to name but a few). We believe that it is always one and the same (latent) education level that is tapped by no matter what education measure. Education effects are merely captured by different measures to different degrees. We are therefore confident that ISLED can be applied in non-stratification contexts as well. Whether it will produce superior results in these other contexts too remains to be seen.

The most important limitation concerning the modelling part of our analyses is that we could only correct random measurement error. Measures may, however, also contain systematic measurement error. To be able to estimate and correct that, we need to repeat the measurement error, which requires the availability of double measures not only for the respondent but also for another person, for example, the partner. Unfortunately, the ESS data do not contain this kind of information. Schröder and Ganzeboom (2012a) present some first results on the impact of systematic measurement error, using Dutch ISSP data, which contain a duration as well as a country-specific education measure for both respondent and partner.

A conclusion to be drawn from the measurement part of our analyses is that the detail contained in education measures is of crucial importance. This leads us to some general recommendations concerning both data collection and analysis. With regard to data collection, we conclude that education data should be collected with as much country-specific detail as possible—very much along the way ESS has been heading. If country-specific variables are harmonized, the country-specific source variables should remain available in the data so that they can be used. In the data analysis, this country-specific detail can be matched with ISLED (as documented by ISLED 2012) and promises to reduce attenuation by measurement error in the analysis. We therefore recommend the use of ISLED as a single indicator in any quantitative comparative study that involves education level as an independent, a dependent, or a control variable.

A conclusion to be drawn from the modelling part of our analyses is that ignoring measurement error amounts to no less than negligence. While latent variable procedures for estimating and correcting measurement error are common practice in the research on attitudes, we plead for an equally meticulous procedure for the measurement of social background variables. This leads us to further recommendations concerning data collection and analysis. Anybody setting out to collect new data would be advised to collect double indicators of social background variables. The good news is that with this method, the mere presence of two parallel indicators is sufficient. Provided that they are based on independent measurements, their individual measurement quality becomes less of an issue. Concerning data analysis, we plead for latent variable modelling, wherever feasible. Even the addition of a weaker parallel measure, such as duration, as a second indicator leads to higher measurement quality of the education variable than the perfection of any individual indicator. Latent variable models have the advantage that they do not require any lengthy procedure to try and fix poor-quality measures. Not only do such models make it possible to estimate the amount of measurement error, they also allow for its correction and thereby produce unattenuated coefficients.

Because our analyses clearly demonstrate how much we can gain in terms of explained variance and regression coefficients, we hope to have increased awareness of the problems caused by measurement error in social background variables and possible remedies for it. While latent variable models deserve preference, we realize that it is unrealistic to expect that all researchers will apply the simultaneous equation modelling techniques required for the correction of measurement error. Given that we have also achieved some considerable improvement of measurement quality through optimal scaling, in our online appendix (ISLED, 2012), we provide country registers with ISLED scores for each education category contained in the

ESS R1–R4 data, which as continuous indicators are ready to be applied in statistical analyses. The application of these ISLED scores will, we believe, like the familiar ISEI scores for occupations, noticeably improve empirical results.

Notes

1. Education levels here are defined as distinct categories. While the term *level* may be understood to refer to having a beginning and an end, we refer to a level here solely in terms of completion. Educational attainment would therefore be defined as the completion of an education programme of a given level and education level is regarded and used here as a functional equivalent of educational attainment.
2. ISCED was first launched by UNESCO in 1976 for a limited number of OECD countries and then revised in 1997. Our discussion refers to ISCED-97. Recently, a revised version has been launched, ISCED-2011 (UNESCO, 2012).
3. Note that the German country-specific variables are not included in the main ESS data files but must be retrieved from the country-specific files.
4. The same revision also introduced EISCED. EISCED is a variable that was developed by Schneider (2009) as an alternative simplification of the detailed ISCED-97, which separates qualifications with access to tertiary education and qualifications without at upper secondary education and distinguishes between Bachelor and Master. Unfortunately, for Rounds 1–4, EISCED is not available for all countries, which is why we use EDULVLa for our comparisons.
5. We use the old EDULVL as country-specific measure where no alternative is available.
6. The ESS data contain detailed parental occupations for the most part as uncoded strings. This information has now been coded into ISCO-88 (Ganzeboom, 2009). Crude and detailed measures do not differ substantially in measurement quality. The coded parental occupation data are available at www.harryganzeboom.nl/ess-devo.
7. The finding that inputs obtain a higher weight may be somewhat surprising. The weights must not, however, be confused with correlations. Despite the lower weight, the output composite correlates more strongly (0.682) with ISLED than the input composite (0.508).
8. The calibration criterion is the mean duration per country over all four rounds. Not all countries have taken part in all rounds, so the time point of calibration ('Europe around 2005') is slightly different between countries. We did not adjust the means for representation of countries in rounds, because changes between rounds in mean duration are very minor.
9. The extreme values 0 and 100 do not arise. The empirically found extremes are 4 and 97.
10. All the information in Table 2 can also be found in our online appendix (ISLED, 2012), where this information is available for all countries.
11. We chose the lower values from Models 4 and 5 because we regard them as the most accurate, arguing that if the second indicator is of poorer measurement quality, the measurement quality of the duration measure is overestimated.
12. Schröder and Ganzeboom (2012b) assign ISLED scores to the categories of the much more detailed ISCED-2011.

Acknowledgements

We gratefully acknowledge critical remarks and helpful comments received at presentations at the ESRA, Warsaw (June 2009); in the SILC research group of the Department of Sociology at the VU University Amsterdam (October 2009); at the Department of Sociology at the University of Amsterdam (November 2009), at the ECSR, Paris, (December 2009); at the IAB, Nürnberg (January 2010); the ISA RC28 (Haifa, May 2010); Hong Kong University of Science and Technology (December 2010); the Interuniversity Working Group on Social Inequality and the Life Course (ISOL); and at the ESS Working Group meeting on Quality Enhancement of Education, at GESIS, Mannheim (November 2011), as well as by three anonymous reviewers of European Sociological Review. Erella Grassiani, Nataliya Nikitina, Hilde Orten, Silke Schneider, and Kadri Täht provided important information for interpreting the country-specific education measures and their harmonizations in the ESS.

Funding

Support for this research was provided by a Strategic Grant of the Faculty of Social Sciences of VU University and by Grant 471-09-005 of the Netherlands Organization of Scientific Research to Ganzeboom.

References

Allison, P. D. and Hauser, R. M. (1991). Reducing bias in estimates of linear models by remeasuring a random subsample. *Social Methods and Research*, **19**, 466–492.

Allmendinger, J. (1989). Educational systems and labor market outcomes. *European Sociological Review*, **5**, 231–250.

Bentler, P. M. (1980). Multivariate analysis with latent variables: causal modeling. *Annual Review of Psychology*, **31**, 419–456.

Blau, P. and Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.

Braun, M. and Müller, W. (1997). Measurement of education in comparative research. *Comparative Social Research*, **16**, 163–201.

Enders, C. K. and Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, **8**, 430–457.

Ganzeboom, H. B. G., de Graaf, P. M. and Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, **21**, 1–56.

Ganzeboom, H. B. G. and Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, **25**, 201–239.

Ganzeboom, H. B. G. (2009). *Improving the Measurement of Social Background Variables in ESS*. Project Proposal Awarded by Netherland Organisation for Scientific Research, Amsterdam: VU University, 471-09-005.

Ganzeboom, H. B. G. and Treiman, D. J. (2012). International Stratification and Mobility File. Available from <http://www.harryganzeboom.nl/ismf> [accessed 28 March 2012].

Hauser, R. M. and Warren, J. R. (1997). Socioeconomic index of occupational status: a review, update, and critique. *Sociological Methodology*, **27**, 177–298.

Hirsch, F. (1976). *Social Limits to Growth*. London: Routledge and Kegan Paul.

Hoffmeyer-Zlotnik, J. H. P. and Warner, U. (2007). How to survey education for cross-national comparisons: the Hoffmeyer-Zlotnik/Warner-Matrix of education. *Metodološki Zvezki*, **4**, 117–148.

Hout, M. and DiPrete, T. A. (2006). What we have learned: RC28 s contributions to knowledge about social stratification. *Research in Social Stratification and Mobility*, **24**, 1–20.

ISLED (2012). Available from <http://www.harryganzeboom.nl/isled> [accessed 2 February 2012].

Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8 User's Reference Guide*. Lincolnwood: SSII.

Kalmijn, M. (1994). Assortative mating by cultural and economic occupational status. *American Journal of Sociology*, **100**, 422–452.

Kelley, J. (1973). Causal chain models for the socioeconomic career. *American Sociological Review*, **38**, 481–493.

Kerckhoff, A. C. and Dylan, M. (1999). Problems with international measures of education. *The Journal of Socio-Economics*, **28**, 759–775.

Korupp, S. (2002). *Mothers and the Process of Social Stratification*. Dissertation, Utrecht University.

Müller, W. (2008). Foreword. In Schneider, S. L. (Ed.), *The International Standard Classification of Education (ISCED-97)*. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.

OECD (1999). *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD Countries, 1999 Edition*. Paris: OECD Publications Service.

OECD (2011). *Education at a Glance*. Paris: OECD Publishing, available from <http://www.oecd.org/edu/skills-beyond-school/48631582.pdf> [accessed 17 October 2013].

Schneider, S. L. (2009). *Confusing Credentials: The Cross-Nationally Comparable Measurement of Educational Attainment*. DPhil thesis, University of Oxford, Nuffield College.

Schneider, S. L. (2010). Nominal comparability is not enough: (In-)Equivalence of construct validity of cross-national measures of educational attainment in the European Social Survey. *Research in Social Stratification and Mobility*, **28**, 343–357.

Schneider, S. L. and Kogan, I. (2008). The International Standard Classification of Education (ISCED-97): Challenges in the Application to National Data and the Implementation in Cross-national Surveys. In Schneider, S. L. (Ed.), *The International Standard Classification of Education (ISCED–97). An Evaluation of Content and Criterion Validity for 15 European Countries*. Mannheim: MZES, pp. 13–46.

Schröder, H. and Ganzeboom, H. B. G. (2012a). De Waarde van Diploma's in Nederland. De ESS-NL Kwantificaties Getoetst. In Aarts, K. and Wittenberg, M. (Eds.), *Proceedings Fourth ESS Workshop*. Amsterdam: Aksant, pp. 63–78.

Schröder, H. and Ganzeboom, H. B. G. (2012b). Lessons from ESS education measures: exploring the potential of ISCED-2011 to develop a new international standard indicator of level of education. unpublished manuscript.

Schröder, H. and Ganzeboom, H. B. G. (2012c). The comparative measurement of level of education in the ISSP—an assessment and application of the ISLED Scale. unpublished manuscript.

Shavit, Y. and Müller, W. (1998). *From school to work*. Oxford: Clarendon Press.

Smits, J., Ultee, W. and Lammers, J. (1998). Educational homogamy in 65 countries: an explanation of differences in openness using country-level explanatory variables. *American Sociological Review*, **63**, 264–285.

Smith, H. L. and Garnier, M. A. (1987). Scaling via Models for the Analysis of Association: Social Background and Educational Careers in France. *Sociological Methodology*, **17**, 205–245.

Thurow, L. C. (1975). *Generating Inequality*. London: The Macmillan Press ltd.

Treiman, D. J. and Terrell, K. (1975). The process of status attainment in the United States and Great Britain. *American Journal of Sociology*, **81**, 563–583.

Ultee, W. C. (1980). Is education a positional good? An empirical examination of alternative hypotheses on the connection between education and occupational level, Netherlands. *Journal of Sociology*, **16**, 135–153.

UNESCO (2006). ISCED 1997: International Standard Classification of Education. May 2006 Re-edition. UNESCO Institute for Statistics. Available from http://www.uis.unesco.org/Library/Documents/isced97-en.pdf> [accessed 17 October 2013].

UNESCO (2012). ISCED 2011: International Standard Classification of Education. May 2006 Re-edition. Montreal: UNESCO Institute for Statistics. Available from <http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf> [accessed 17 October 2013].