MEASURING AND MODELING LEVEL OF EDUCATION IN EUROPEAN SOCIETIES

Paper first presented at RC 28, Haifa, May 9-11, 2010

Heike Schröder (H.Schroder@fsw.vu.nl)
Harry B.G. Ganzeboom (HBG.Ganzeboom@fsw.vu.nl)

Department of Social Research Methodology
VU University Amsterdam

Work in progress. Last revised: January 31, 2010 [HS]

Further presentations:

- **ESRA, Warsaw, June 2009**
- **SILC, October 2009**
- **UVA, November 18, 2009**
- **ISOL, December 2, 2009**
- **ECSR, Paris, December 2009**
- **IAB, Nürnberg, January 29 2010**

MEASURING AND MODELING EDUCATION LEVELS IN EUROPEAN SOCIETIES

In this paper we develop the International Standard Level of Education (ISLED) for educational qualifications in 34 European countries as measured in Round 1-4 of the European Social Survey [ESS]. Using a Multiple Indicator Multiple Cause (MIMIC) perspective, we estimate how country-specific education categories are scaled relative to each other with regard to influences of inputs (in particular parental occupations and education levels) and effects on outputs (in particular the acquisition of occupation and partner's education). The ISLED measure is defined the optimal scaling of educational qualifications that mediates inputs into outputs. Employing the presence in ESS of another harmonized measure of level of education, its duration, we assess the performance of the newly created ISLED relative to the duration measure and the often common-denominator approach to harmonizing level of education, the International Standard Classification of Education [ISCED], in a multiple indicator model. ISLED performs best, but still falls short of perfect measurement by about 6%. For ISCED and duration the loss is about 10%, respectively 18%. We find that any combination of measures produces better results than each of the measures on its own. ISLED scores may be used to scale country-specific education measures in situations when only a single indicator is available, but our final recommendation is to improve measurement by multiple indicator modeling.

1. Introduction

Educational attainment is one of the most widely used variables in the social sciences in general and in stratification research in particular. It is a core stratification variable that regularly occurs as input and output variable. As an output variable educational attainment is not only the product of individual characteristics such as intelligence, effort and motivation but also of social background characteristics, such as parental education and occupation levels. A plethora of studies has been devoted to the inequality of people's educational opportunities (Shavit & Blossfeld, 1993; Müller & Shavit, 1998; Kerckhoff, 2001; Mare, 1981). As an input variable, education produces a large number of effects. It has an impact on socio-economic outcomes relevant for social stratification, such as work, occupation, income and social prestige (DiPrete & Grusky, 1990; Ultee & Luijkx, 1990), as well as on the structure of friendship and marriage networks (Kalmijn, 1994; Van de Bunt, 1999). Apart from these direct stratification effects education affects many other aspects of people's lives, such as health (Ross & Wu, 1996; Westert et al., 2005), criminality (Lochner, 2004), family stability (Duncan & Duncan, 1969; Poortman & Kalmijn, 2002), mortality (Lleras-Muney, 2005; Doornbos & Kromhout, 1990), cultural participation (Bourdieu & Passeron, 1977; Ganzeboom, Treiman & Ultee, 1991), knowledge (Hyman, Wright, & Reed, 1975), values

(Hyman & Wright, 1979; Inglehart, 1971) and attitudes (Brint, 1984; Van de Werfhorst & De Graaf, 2004).

As a control variable education is of no lesser importance. It is used to isolate pure life course event effects on later outcomes. For instance, we assess the impact of occupation on income or that of migration on occupation preferably via models in which education is held constant, because it confounds these relationships. Another example is that when we examine the influence of unemployment on income, we better take into account that the higher educated have lower risks of becoming unemployed in the first place and when they do become unemployed, tend to keep higher incomes as well (Mooi-Reci, 2008).

The question we focus on in this paper is how educational attainment can best be measured in international comparisons. Being a crucial variable in many empirical problems, education levels should be measured with due care. Yet, an examination of existing cross-national surveys reveals a rather astonishing lack of comparability. Surveys use a large variety of classifications of education systems and the respective number of separate categories. Moreover, they differ significantly with respect to the precise wording of the questions and answer categories presented to respondents. There are several reasons for this rather dissatisfactory state of affairs. The main reason is probably the complexity of the task at hand. After all, what needs to be accommodated into the measurement is a staggering diversity of existing education systems, which do not only vary across countries but also over time, leading to different cohorts of respondents having been exposed to different education systems even within one country. What is more, 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, abolishing some and adding others.

While the diversity in education is undoubtedly at the root of the problem, this is not the end of story. Attempts have been made to implement measurement standards, based on the common denominator principle (discussed below). Unfortunately different surveys have opted for different standards and even where the same standard classifications are used, they have been implemented in different ways. Whatever choice is made, it has consequences in terms of lost categories. These consequences vary between measures and differ in severity but need to be addressed if we want to obtain unbiased coefficients in our statistical analyses. We argue that the common denominator approach is suboptimal for the internationally comparative measurement of education because it cannot avoid loss of information. We therefore propose two alternatives for accomplishing comparability. Our first alternative is the optimal scaling of the detailed country-specific education variables in a MIMIC model. We demonstrate that this method produces better results in statistical analyses than the conventional measures. The second alternative is the use of multiple indicators. This method turns out to lead to an even greater improvement of results.

In the social sciences measurement issues are most seriously addressed in the measurement of attitudes. Great efforts are made to measure concepts as precisely as possible in order to obtain statistical estimates of an acceptable level of accuracy. Here it is common practice to use multiple indicators to cover various aspects of the concept in question. These indicators are tested and combined in scales that are subsequently used for analysis. Researchers use internal consistency measures, such as Cronbach's alpha and construct index variables with strongest possible discriminatory power. The aim of all these efforts is to reduce measurement error.

Unlike attitudes, social background variables are hardly ever measured with multiple indicators, and their measurement quality is rarely addressed either. Researchers usually content themselves with single indicators of variables such as sex, age, occupation or education, likely assuming that these are fixed features that have an unequivocal content and are measured without (sizable) error. If measurement quality is addressed, discussions invariably address validity issues, not reliability.

This does not mean that the quality of social background variables has remained undebated. On the contrary, there are many debates concerning the measurement of education, occupation, income, ethnicity etc., focusing on the "single best indicator". In this paper we demonstrate that, like for attitude variables, the use of multiple indicators is the most effective way of improving measurement quality for social background variables as well.

The most profound bias of measurement error on regression coefficients arises when error occurs in the independent variable (Bohrnstedt & Carter, 1971), with random error attenuating bivariate relationships and correlated error inflating them (Munck, 1991). If we want to obtain unbiased regression coefficients it is essential that we deal with this. Hence we argue that social background variables, like attitude variables, ought to be measured with multiple indicators. This would enable us not only to diagnose measurement error, but also to correct for it in a structural equation model, which is an advantage that can hardly be undervalued. The use of multiple indicators also makes it possible to answer the following questions: How large is the measurement error for the various individual indicators? How can we correct for it? And if we improve our measurement, how great is the relative difference in terms of statistical coefficients?

The aim of the paper is three-fold. First, we assess the quality of the existing measures of educational attainment in the European Social Survey (ESS) by comparing relevant coefficients they produce in a status attainment model. Secondly, we derive a superior measure for each ESS country by means of optimal scaling. Consisting of score values for each education level, these measures can be used as a research tool and help data users improve the quality of their analyses of ESS data. This research tool will be called the International Standard Level of Education, (ISLED). Like its counterpart for occupational status, the International Socio-Economic Index (ISEI) (Ganzeboom, de Graaf & Treiman, 1992), ISLED is available on-line (ISLED 2009)[1] and can be readily applied for statistical analyses with the promise of more accurate results. Thirdly, we demonstrate that even the best single measure is outperformed if we combine two measures in a multiple indicator model. Using multiple indicators is superior not only to the common denominator approach but also to optimal scaling. We therefore conclude that the most effective way of improving the measurement of education level in the ESS is the use of multiple indicators.

2. The comparative measurement of educational attainment: the state of the art

What then are the advantages and disadvantages of different approaches to the comparative measurement of educational attainment? If we want to improve on the existing state of the art, we first need to specify what the underlying principles are and which consequences these each have for empirical outcomes. We present three different approaches: harmonization, duration and scaling.

Harmonization: largest common denominator

[1] Up to this date ISLED is available on-line for the Benelux countries and Germany.

The most frequently used method of measuring educational attainment in cross-national survey research is harmonization by largest common denominator. The idea here is that the different national education systems are to be made comparable by looking for equivalent elements in their national classifications. The problems with this approach are easy to anticipate. To begin with, such a strategy leads to a loss of information as any common denominator by definition contains fewer categories than the source classifications to be harmonized. Such aggregation error would only not occur if the collapsed categories were fully homogeneous with respect to criterion variables. Secondly, for some categories it is simply not possible to find a common denominator and incomparabilities can at best be solved by compromise. Thirdly, these problems grow with the number of source classifications that need to be harmonized. After all, the largest common denominator of 10 different classifications is smaller than that of three and the chance of finding unharmonizable elements increases accordingly. Any newly arriving dataset will increase measurement error and thus decrease statistical power, instead of reducing measurement error.

A widely used common denominator approach to the measurement of education is the application of the International Standard Classification of Education (ISCED) developed by UNESCO. This strategy has been applied in several cross-national research projects, such as PISA (Program for International Student assessment), IALS (International Adult Literacy Survey) and also the ESS. As ISCED is based upon the existing education systems of one particular year, 1997 (or 1976 for its predecessor) and confined to a limited number of countries, it is considerably less useful than its counterpart for occupations, the International Standard Classification of Occupations (ISCO). In practice, the way ISCED is implemented in these surveys produces a coarse educational distribution, rather than a detailed classification. A further important downside of ISCED as applied in the ESS is, that it yields an invalid representation of some important distinctions. For the German and Dutch contexts, for example, it fails to distinguish between vocational and academic education both at the secondary and the tertiary stage.

Kerckhoff & Dylan (1999) have compared several implementations of ISCED-76 in surveys and conclude that the way standard categories are derived from country-specific classifications can cause large differences in the type of results produced by comparative researchers. Schneider (2008a) has evaluated the quality of ISCED-97 and the way it is applied in the ESS and lists a large number of problems. One general problem she finds is that the ISCED-97 categories, while being more detailed than those of its predecessor, still contain insufficient differentiation for some levels.

When assessing approaches to harmonization, it is useful to distinguish between pre- and post-harmonization. Pre-harmonization means that a harmonized standard classification such as ISCED is directly applied in the question format of a survey. For the ESS this approach has been used for the measurement of the education levels of parents and spouse. The preformatted questions contain the seven main ISCED categories. This is the least desirable approach, as any reference to the underlying original local categories is lost for good. With this method the loss of information is beyond repair.

In principle, post-harmonization can be equally damaging in terms of lost information. Here surveys ask for locally relevant categories, which are subsequently recoded (aggregated) into for example ISCED. Given that the source categories remain accessible, the information lost can, however, be restored, making it possible to detect coding errors or exploit the extra detail of the local categories for analytical purposes. For these reasons post-harmonization is preferable. The mere availability of detailed local categories, however, in no way guarantees that this information is actually exploited. On the contrary, ESS data users tend to ignore them and confine themselves to the familiar ISCED. In practice the local variables tend not to be used by analysts, who are often lost in view of the great variety of distinctions and labels they are confronted with.

Duration

Another widely used method to compare levels of education cross-nationally (e.g. in the ISSP, the IALS and the ESS) is to take the duration of educational careers as an indicator, effectively assuming that the duration of an educational career increases by-and-large with the education level to be achieved. The questions posed refer to either the school-leaving age or the number of years spent in education. Both presuppose a strictly hierarchically ordered education system. Duncan and Hodge (1963) were the first to implement such a duration measure, scoring people’s educational attainments according to the number of years of schooling they successfully completed.

Using duration as the basis for the measurement of educational attainment avoids some pitfalls of common denominator harmonization. What is exploited instead is a feature that is intrinsically present in the organization of any education system, namely that it takes time to pass to higher levels. In effect duration is a measure that appears to be directly comparable across systems with no further transformation needed. While the advantages of this approach must be acknowledged, it has drawbacks too. As Hout and DiPrete (2006) note, this method works reasonably well for the rather undifferentiated education system in the United States but is much less suited to capture the distinctions of education systems elsewhere. Schneider (2007) too questions the validity of the duration measure and draws attention to the fact that it is not evident what 'years of education' actually means. Given the identical length of very different types of educational programmes within and across countries, she argues that confining measurement to duration amounts to concealing qualitative differences between them.

Scaling

A third strategy to make educational categories comparable is via a common scaling. Such a scaling tends to be constructed on more or less arbitrary grounds. One common practice is to measure level of education by the years it takes to achieve a given level according to the institutionalized education system. Such measures are provided in the *Education at a Glance* publications of OECD and also appear in the ISCED manual. Clearly, such scaling by 'institutional years' is vulnerable to the same criticism as duration: a certain amount of years in the system may lead to entirely different levels of outcomes, if the system is tracked. In the International Stratification and Mobility File (Ganzeboom & Treiman 2008) a variable is provided that expresses local categories in 'pseudo' years of education. The scale score expresses the level of categories relative to the highest track, i.e. the number of years is reduced when the obtained qualification is in a lower track. A related approach is proposed by Hoffmeyer-Zlotnik & Warner (2007), who organize education categories in four countries into 10 different levels that can be regarded as an ordinal hierarchy. A striking commonality of all these approaches is their ad-hoc nature.

In an empirically driven approach, the so-called *effect-proportional scaling*, scale scores for educational categories are generated in such a way that the correlation between the derived education scale and a given criterion variable (e.g. occupation or income) is maximized. For example, in a comparison of the US and British education systems Treiman & Terrell (1975) derive comparable education scores, by using an outcome, the occupational status of the respondent, as a criterion variable. They motivate this choice with the argument that it is a primary function of the education system to prepare individuals for professional life and that the correlation between education and occupation is particularly high. Smith & Garnier (1986) by contrast, generate an educational scale by using an input variable, namely father's occupation, as criterion variable in a loglinear model. Like the occupation of the respondent, the occupation of the father too, is strongly associated with education level. Given that this approach uses an input as criterion variable, we will label it *cause-proportional scaling*.

Similar to the other measurements of educational attainment, effect- or cause-proportional scaling has not remained without criticism. Braun & Müller (1997) for example contend that in effect-proportional scaling we have to assume that the respective country-specific measurements exploit the potential explanatory power of educational attainment to the same degree. According to these authors, the problem of comparability is transferred from the education variable to the criterion variable. Consequently, the respective criterion variable would have to be measured in a strictly comparable way, which is rarely the case. We deny the validity of this argument because even if the criterion variables are poorly measured, this will not affect the respective ordering of the education levels.

3. A MIMIC approach to optimal scaling

In our analysis we will follow up on previous scaling approaches, improving upon them in two ways. We integrate cause and effect-proportional approaches in one unified model. While Treiman & Terrell (1975) used a single output variable as a basis for their scaling and Smith & Garnier (1986) a single input variable, we use both types of variables. Our approach is therefore best described as *cause-and-effect-proportional scaling*. Instead of confining ourselves to single input and output variables, we, moreover, combine several variables of each type in a MIMIC (Multiple Indicator Multiple Cause) model.

The basic model for such a construction of an optimal scale score is shown in Figure 1. Here, discrete educational categories are interpreted as intervening between a number of input and a number of output variables. The input variables basically measure parental resources or constraints that condition their offspring's educational attainment. The education levels for their part lead to a number of outcomes. We restrict ourselves to two of them, which are strongly associated with education, however measured: occupational status of the respondent and educational attainment of the spouse.

The MIMIC model leads to an optimal scaling of educational categories if we combine it with classic measurement theory as applied to a causal chain model.

$$X \rightarrow Z \rightarrow Y$$

In a causal chain of three variables we presume that X affects Y only via Z. Whether or not this presumption corresponds to empirical data depends not only on the quality of the theoretical model. In particular, the quality of measurement of the intervening variable Z is important. If X or Y are measured imperfectly we find attenuated coefficients b(YZ) and b(ZX), but it remains true that the direct effect of X on Y remains zero (b(YX|Z)=0). If Z is not perfectly measured, however, this outcome changes and we will find some direct effect of X on Y not via Z (bYX|Z > 0). In order to be able to adequately estimate a causal chain it is therefore necessary to correct measurement error in Z (Kelley, 1973).

We can apply such reasoning to the problem of optimally scaling education categories. A suboptimal scaling is one that weakens the mediating role of education in the status attainment process and yields larger direct effects of inputs on outputs. An optimal scaling is one that maximizes the central role of the acquired education level.

Theoretical support for this procedure can, furthermore, be derived from two substantive theories, status attainment theory and positional good theory, which allow us to ground our measurement procedure in sociological theory and uncover the empirical basis of our scaling procedure.

Status attainment theory

Status attainment theory (Blau & Duncan, 1967) is concerned with people’s positions in society and investigates which factors determine occupational status and to what degree. On the one hand, occupational status is affected by achievement factors, such as educational attainment, on the other hand by ascription factors, such as parental occupations and educational statuses. In the theory, education plays a pivotal role, as the main determinant of occupational status and the mechanism transferring occupational status from one generation to the next. Blau & Duncan's celebrated status attainment model describes the relevant variables and causal relationships between them and provides us with a structural embedding for our measurement procedure. We conceive of education as an integral part of the stratification process and take both sides of the status attainment process into account. On the acquisition side the value of an educational qualification reflects how different status groups value different levels of education. Social groups compete with each other and the outcome of that process informs us of the perceived values of educational credentials. Higher status groups are more successful in ensuring higher-level certificates for their offspring than lower status groups.

Regarding the outcomes of the status attainment process, the achievements of certificate holders in the labour and marriage markets reflect how various educational certificates are rewarded in society. The most obvious indicators of such achievements are the occupation and/or the income acquired with it. These, however, are not the only effects of education. Educational attainment also has strong effects on other positions in society. Most notable is hereby the strong association between the education levels of partners, parents and offspring or that among friends. The obtainment of a certain certificate not only determines what sort of job a person will find, but also who they are going to mix with, who they get married to and have children with and also what education level these children are likely to acquire.

The status attainment model provides an important rationale for the choice of a MIMIC model to scale occupations. The underlying theoretical assumption in scaling education levels is that educational qualifications, where or whenever obtained, are typically ordered in a one-dimensional common hierarchy. Various qualifications have a certain distance to each other, which, in the ultimate case is zero (their levels collapse). How can we form an image of such an educational hierarchy and how can we determine the relative positions (scale values) of educational qualifications? If we conceive of education as part of the stratification process, it makes sense to scale qualifications in a way that takes both sides of that process into account.

Schematically, education level is regarded to be a variable that mediates between social backgrounds and outcomes in the further life course. The respective backgrounds are multiple and so are the outcomes. It is therefore justified to speak of a multiple indicators, multiple causes [MIMIC] model. This model provides us with a nice way of finding an unequivocal optimal scale of education levels: we define this scale as **the relative positions of education level on a single dimension, such that the direct effects of social backgrounds on social outcomes are minimized.** We thus conceive of education level as a latent variable that is reflected in the various education measures to varying degrees.

Although our procedure is related, optimal scaling of qualifications from a MIMIC perspective is not the same as optimal scaling based on only inputs or only outputs. From a theoretical point of view, our approach has the advantage that it takes both sides of the status attainment model into account. The approaches of Treiman & Terrell (1975), who focus on one effect, and that of Smith & Garnier (1987), who concentrate on one cause, are both valid. By combining both types of variables, we make use of more information. Pragmatically, a MIMIC scale stands midway between the two alternatives. This implies that a MIMIC scale may be suboptimal to both criteria separately. The resemblance between the two will be stronger the more in- and output scalings resemble each other. If they are identical, both ways of scaling yield identical results. In using multiple variables on both sides, the impact of potential measurement errors of the individual criterion variables is reduced, countering yet again Braun & Müller's (1997) afore-mentioned criticism against effect-proportional scaling that the problem of comparability is transferred from the education to the criterion variable. To sum up, we use all available relevant information, keeping measurement error at bay.

Positional good theory

What all education systems at all times and places, regardless of all their institutional differences have in common is that the hierarchical organization of education systems allocates people to positions in a rank order. This hierarchy of individuals corresponds to the theoretical notions of job queue (Thurow, 1975) or, in the case of assortative mating, candidate queue. It is this very idea that is at the heart of positional good theory (Hirsch, 1976).

According to this theory, goods differ from each other in terms of supply. While material goods are in principle available in unlimited quantities, positional goods are of fixed quantity and through more extensive demand subject to congestion or crowding. As the number of available jobs and educations depends on their place in the hierarchy and decreases the higher up in the hierarchy they are situated, Hirsch lists both among the positional goods. Education in this view determines an individual's rank in a hierarchy of job seekers, before they enter the labour market. Therefore it is not the absolute value of an individual's education that is relevant, but its relative value compared to that of competitors in the job queue. People compete over scarce jobs and educational credentials are crucial in employers' decisions on who gets access to what kind of job and hence to what privileges. This implies that it is not the labour market that is the major source of income inequality, but the educational institutions producing skill (or credential) differentials in the first place. It is educational credentials that determine an individual's rank in the job queue and thereby their subsequent position in the labour market. Viewed like this education functions as a means of social closure.

If we investigate the way education facilitates social closure within a status attainment model, education is relevant in two ways. On the one hand a given education level produces outcomes in the labour and marriage markets. On the other hand, education is itself a positional good since the higher-level educational programmes leading to better paid and valued jobs are themselves in limited supply as well.

The positional good perspective, which focuses on the single hierarchy of people produced by the education system, provides another theoretical anchor for the educational scale we want to derive. As the complete process of social stratification is at issue, the positional good perspective is applied in a social closure interpretation, paying tribute to the role education plays in intergenerational status reproduction and assortative mating.
It should be clear from the above that the principles underlying existing measurement practices of educational attainment are rather static and crude. Whether the measures are based on the content or the duration of educational programmes, they all pertain to formal features of education systems and lack theoretical underpinning. Education is treated as an isolated variable of individual scope, rather than as a working mechanism in social processes, such as tapped by the intergenerational status attainment process. The only implicit link to theory seems to be that education systems are essentially hierarchical institutions. No reference is made though to the function these institutions fulfil in the status attainment process. By generating cross-nationally comparable educational scales based on positional good theory and embedded in the status attainment model, the methodological and theoretical concerns with conventional measurement procedures raised above are addressed simultaneously. The focus in the measurement thereby shifts as it were from the intrinsic

structural hierarchy of the education system itself to the hierarchy of individuals that system produces.

With this approach we take up Grusky and Van Rompaey's (1992) suggestion that, when attempting to improve measurements, sociologists should apply vertical scales that capture middle-range concepts of some sociological significance. Relying solely on an empiricist critique, by contrast, would amount to the impossible task of adjudicating between the various "corrections" that might be put forward.

4. Source data and constructed variables

The ESS is a high-quality survey that has been held biennially in about 30 European countries, starting in 2002. It includes two independent measurements of educational attainment, a detailed country-specific classification and duration. Because the detailed classification is post-harmonized into ISCED, the two measurements yield three different measures. For this reason the ESS provides a useful illustration and testing ground for the advantages and disadvantages of pre- and post-harmonization and duration and allows us to demonstrate the benefits of multi-indicator models.

Country-specific education categories

It has been one of the policies of the ESS to leave countries the freedom to employ a detailed country-specific education typology, as a means to complement standard measurement and a possible remedy for problems encountered through harmonization. These measures have been included in the main ESS data file. Closer inspection of these variables reveals some problems though. A majority of countries has followed this strategy, but not all. For Austria, Finland, Iceland, Slovenia and Turkey no detailed measures are available. The Irish, Italian and Ukrainian country-specific measures turn out to be identical to their ISCED equivalent in at least one round. For the remaining countries the number of categories distinguished in the country-specific measure varies from 5 for the UK (which is less than the ISCED) to 19 for Luxembourg. This illustrates that the detail available and hence the information that can be lost varies considerably between countries.

Common denominator measure

Like many other surveys the ESS employs a largest common denominator strategy using seven categories, ranging from (0) No education to (6) Second Degree of Tertiary training. This information is recorded in the standard ESS variable EDULVL.

Schneider (2008a) bemoans that in the ESS only the seven major categories have been implemented, reducing the much more complex classification that ISCED is actually meant to be to its bare bones. Moreover, researchers make mistakes with the implementation of ISCED, resulting in aggregation error. Schneider adds some more specific problems to the general one of aggregation error. She notes for example that some of the available criteria for level 2, lower secondary education, do not work well and that for other levels criteria are missing to distinguish between qualifications.

An illustration: the case of Luxembourg

We have argued that common denominator harmonization by its very nature entails loss of information. But how serious a problem is this in practice? In order to illustrate how grave the consequences of harmonization can be we examine the case of Luxembourg in more detail. Within the ESS Luxembourg is the country with the most detailed country-specific education classification. For this reason it lends itself particularly well as a test case.

Harmonization means loss of information because the number of categories is reduced and relevant distinctions are lost. For the Luxembourg case this implies that the original country-specific classification that contains 19 categories is reduced to seven ISCED categories. In other words, as many as 12 categories are lost. Table 1 illustrates the conversion of the country-specific Belgian, Dutch and Luxembourg classifications into ISCED. For Luxembourg table 1c shows that ISCED categories retain between one and five of the original distinctions. This table also shows the frequencies of the ISCED categories and that the distribution of the respondents over the various categories is very uneven. The smallest category, 6, contains 22 cases, while the largest, category 3, contains as many as 818. It is evident that the ISCED classification includes a lot less detail than its source typology. In our analysis section the consequences of this loss of information in terms of regression coefficients and explained variance will become apparent.

Duration

For the respondent, the ESS data contain a second education measure, duration. The question asked is the number of years (in full-time equivalents) the respondent has spent in education. This information is provided in the standard ESS variable EDUYRS. Schneider (2008b) has compared the quality of the duration measure with the ISCED and concludes that it performs much worse than ISCED.

Notwithstanding these problems, the availability of the duration measure in the ESS is very valuable. Whether it is a good or a bad measure, duration is a second, independent measurement of educational attainment and therefore a means of tracing measurement error via a multiple-indicator model. Unfortunately, it is only available for the respondent.
For the education levels of mothers, fathers and spouses we have to make do with ISCED measures.

Criterion variables

The choice of criterion variables to be used in the construction of our optimal scale depends on the theoretical model, the availability and utility of possible variables in the ESS data and the strength of association of a given variable with education. The choice is further limited because only variables that have direct (not indirect) causal relations with education are considered suitable.

For the input side of the model we have chosen for parental occupations and education levels. ESS data (Rounds 1-2-3) include only one indicator of parental education, the pre-harmonized ISCED: EDULVLF and EDULVLM for father and mother, respectively. The nature of the parental occupation indicators varies between countries. A poor crude occupation question is available for all countries: OCCF14a and OCCM14a, for father and mother respectively. In some countries, however, parental occupations have also been measured in the much more detailed ISCO format. All of the occupational information has been converted into ISEI scores.

For the output side of the model we have chosen the respondent’s occupation and the education level of the spouse. The respondent’s occupation is measured in ISCO/ISEI, educational attainment of the spouse is again measured in a pre-harmonized ISCED format (EDULVLP). Note that we do not use income, a feasible alternative for occupation, because the effect of education on income runs via occupation.

5. Optimal scaling

The algorithm we use to find the optimal scaling of education levels is a variation of the algorithm used for the development of the ISEI index (Ganzeboom, De Graaf & Treiman, 1992), where occupational status was defined and calculated as the scaling of occupations that mediates best the influence of education on income.

The scaling of education levels differs from that of occupations in a number of ways. First, when scaling occupations we are dealing with a large amount of occupational groups and therefore with relatively thin data. The number of education levels is comparatively very limited, which simplifies the solution of the optimisation problem. Secondly, the scaling of occupations was based on a single input and a single output variable. For education we use several input and several output variables. This complicates matters. Thirdly, the issue of scaling may also be viewed from an institutional position: what are we to do if the optimal scale value of a continuation course turns out to be lower than its preceding level? It remains to be seen if this actually happens but if it does, it will be difficult to promote a resulting ‘optimal’ scaling as the best evaluation of certificates.

We apply a simplified version of the algorithm devised by Ganzeboom, De Graaf and Treiman (1992) to derive the ISEI scores. The first simplification is that we do not control for age as a confounder (age differences are pertinent to occupational status, but much less so to educational status). Complexity is also reduced by first assembling the four input variables in a single composite variable, and then the two outputs as well. We use unweighted averaging of available standardized indicators as the ingredients of the two composites. The derived scale score is basically a weighted and standardized average of these composite inputs and composite outputs. Optimization is reached for a certain combination of relative weights for the two composites that we can simply find by a systematic search. The search stops when the remaining direct effect is at a minimum. In the ESS data[2] this happens to be the case for 0.55

[2] In the present version of the paper, the empirical results only refer to the three Benelux countries in ESS Round 1-2-3.

(inputs) and 0.45 (outputs), values that are fairly representative for what we have found for other countries. The resulting weighted average is then Z-standardized. The column in Table 1 labelled OPTI shows the obtained scores. It is immediately apparent that the ISCED aggregation hides substantive variations within its categories, and also that the ISCED categories do not always constitute a proper rank order.

The constructed optimal scores are Z-standardized within countries: they are comparable within countries, but not between countries. As a consequence, a similar or identical education programme may obtain a different level score in two countries, depending upon the specific association with the criterion variables in the two countries. The within-country standardized metric may satisfy many needs (in particular when doing analyses on a country-by-country basis), but will not allow the analyst to control level of education as it may be different between countries, nor compare means and dispersion between countries. For this, a common metric needs to be established.

Transferring the optimal scores into a common metric may be achieved in different ways, but will always involve the choice of calibration points at which two countries are regarded to be the same. To set a common scale, at least two calibration points need to be used, as well as a transformation function. We offer the user two different versions of the International Standard Level of Education. In one version of the ISLED scale (A), we have transformed the optimal scores into a metric that resembles (but is not identical with) a duration measure. We choose two calibration points, establish their value in duration and project all the other scores using a simple linear equation. In the second version (B), we derive our calibration points from a Europe-wide distribution of education levels as it is presented in our effective sample of 25-74 year old men and women in over 30 European countries. The calibration points reflect the relative standing (percentiles) in this distribution. We then project the optimal scores into a 0...100 metric using a logistic transformation in order to avoid out-of-range projections.

We have chosen our calibration points close to two ends of the scale. One is the completion of primary education, which in the Benelux countries corresponds to 6 years of education and a 10% percentile rank. The other is the end of the highest academic level of secondary education, which corresponds to 12 years of education and a 71% percentile rank. Table 1 provides the original optimal scores and their corresponding ISLED scores for each country, provisionally labelled ISLED-A (duration metric) and ISLED-B (percentile metric).

In practice, it will make very little difference which version of ISLED is being used. Over de Benelux countries, they correlate extremely strongly (0.98). The difference becomes more relevant when one has to phrase results, and in this respect we find ISLED-B (percentile metric) a bit more convenient to talk about than ISLED-A (duration metric). A score in ISLED-B would refer to the expected number of people lower than that rank in a standard European population. Percentile statements (‘I am the best’, ‘my university is top 10 percent’, ‘the poorest 20 %’) are often used and easily understood by wider audiences. The ISLED-A metric also refers to an easily understood metric (how long it takes), but this may create confusion as well: the reference is relative to the duration *of the calibration points* and we fear that the comparison with actual duration might cloud the clarity of interpretation.

6. Estimating the quality of the ESS education measures

So far we have concentrated on generating a single best indicator of educational attainment. How can we measure the quality of the constructed indicator? We can compare the relative quality of indicators by examining how these indicators behave in predicting outcome variables and when predicted by input variables themselves. We will now take the principle of maximally using all available information a step further. Rather than restricting ourselves to comparing single measures, we make use of the extra information contained in the remaining indicators in a multiple indicator measurement model.

We use multiple indicators for a number of reasons. Indicators relate to a concept in a part-whole relationship. Consequently, different indicators measure (slightly) different aspects of a concept. They have a common and a unique (but systematic) part. If only one indicator is used, this may cause bias. Indicators are also unreliable measurements, i.e. some part of their variance is random. The use of a multiple indicator model is a way of detecting and correcting for measurement error. As even small random error may have large consequences in terms of accuracy of regression coefficients, correcting for it should be worth our while.

For our analysis we regard educational attainment as a latent variable, which has three reflective indicators. Since these three measures are based on two independent measurements, they reflect the value of education with different amounts of error. This error can be estimated and corrected in a structural equation model. The three indicators each contain differing amounts of error, but none of them is perfect.

Conceiving of educational attainment as a latent variable with a continuous true score, we combine all available education measures, including our optimized ISLED, in one model. This allows us to reconstruct the effect of the actual, i.e. the true level of education, to assess indicator validity (Alwin, 2007) and to obtain estimates of the valid variance of the individual measures in the form of validity coefficients. Apart from that we also obtain estimates of the invalid variance of the measures, which consist of method variance and unique variance (Saris & Andrews, 1991). While it is not possible in our model to separate the two from each other, the model corrects for the composite error contained in the individual measures, at the same time countering attenuation effects. Even if we have suboptimal measures, we can limit the loss of information because a multiple indicator model makes it possible to balance the weaknesses of the individual indicators.

A multiple indicator model also allows us to compare the quality of the individual measures. To evaluate the gain achieved by using multiple indicators, we compare the explained variance in educational attainment and the explained variance by educational attainment. Our analyses show that any combination of indicators leads to higher ultimate measurement quality than any single indicator on its own, including ISLED.

The model

Figure 2 is an illustration of the indirect effects simultaneous equation model. It shows how the inputs are transferred into outputs and how level of education is affected by the inputs and for its part affects the outputs. Our model consists of two parts, a measurement model and a structural model. As is customary in simultaneous equation models latent variables are represented by ovals and measured variables by rectangles. Figure 2 shows the full model, consisting of a structural and a measurement model. The measurement model illustrates how the latent education variable is measured with three indicators, ISCED, duration and our ISLED scale. It is not identified by itself but becomes identified if we embed it into a structural model, by including input and output variables.

7. Results

We have analysed the data of those countries and rounds that contain country-specific education measures and double indicator measurement of parental occupation. The first part consists of a detailed analysis of the Benelux data (rounds 1, 2 and 3) presented in table 2. Due to the large amount of detail contained in the country-specific measure, Luxembourg (table 2c) presents the most extreme case. Here a maximum amount of information is lost through aggregation, which is why this case is particularly well suited to demonstrate the points we are trying to make. This example serves to illustrate the consequences of various choices of indicators both for the structural models in terms of regression coefficients and explained variance as for the measurement model in terms of differences in the factor loadings, which indicate the validity of the individual measures. By comparing indicator validities, we can decide which measure performs best. Subsequently we demonstrate the effects of possible strategies to improve our results.

The second part of this section pertains to the general findings across countries and rounds. These results contain a comparison of the measurement models as well as percentages of explained variance in and by the education variables per country and round. We will conclude our results with a list of score values, the ISLED, which, like the ISEI, are ready to be applied in the statistical analysis of the ESS data. Although multiple indicator models are deemed to produce even better results, we consider these scores a useful pragmatic solution to improve the quality of analyses.

The Benelux countries

Table 2 presents five relevant SEM models for the Benelux countries, in the form of three standardized regression equations. Educational attainment and occupation of the respondent and educational attainment of the spouse are the respective dependent variables. The first three models differ in the variable used as the single indicator of the latent education variable. Model 1 uses ISCED, model 2 duration and model 3 the ISLED scale. When comparing the models, we can consider the size of the regression coefficients or the explained variance of the separate structural equations. Unsurprisingly, the ISLED turns out to be the best single indicator. Its superiority is reflected in higher direct effects of educational attainment both on respondent’s occupation and spouse’s education. The indirect effects of parental education on spouse’s education and parental occupations on respondent’s occupation are accordingly smaller. The ISLED scale also yields the highest levels of explained variance for all three structural models in all the three countries. The differences between the ISCED and duration measures are somewhat less marked, but reveal that ISCED is a slightly better indicator of educational attainment than duration. That the differences in quality between the ISCED and duration measures are so slight is rather striking. Schneider (1988b) who compared the same two indicators in a simple regression model of ISEI on education found that using duration leads to a loss of information of almost 20 % compared to ISCED. In our models the difference between the two is only an average of 7 %. For the Luxembourg data duration even turns out to perform slightly better than ISCED. This illustrates that the choice of criterion variables is not trivial. It remains to be seen whether the results obtained with different (sets of) criterion variables may actually contradict each other but what appears to be true is that the loss of information through the use of a specific indicator varies.

Up till now we have followed other researchers in comparing the quality of the various indicators. We now go a step further and examine what happens when we proceed to a multiple indicator measurement model. Instead of one single measure we use various combinations of the available measures. In model 4 ISCED is combined with duration, which leads to a dramatic increase in explained variance in all three regression equations. The explained variance is on average more than 5 % higher than in the best single indicator model. In accordance with this improvement we see a stronger effect of education as the determinant of respondent’s occupation and partner’s education and a greater predictability of education from any of the parental background variables.

In Model 5 we repeat the exercise, this time with duration and ISLED as indicators. The results change but the picture varies with regard to both explained variance and regression coefficients. Some are higher, some lower. What does improve consistently across the three countries are the explained variance as well as the effect size of education with the respondent’s occupation as a dependent variable. They are higher in all three countries. In Model 6 finally, we combine all three indicators. The estimated coefficients do not lead to more improvement compared to those in the two-indicator models.

An important piece of information concerns the measurement relations between the latent education variable and the respective indicators, presented in the bottom of the table. In model 5 this ranges from 0.764 to 0.831 for duration, and from 0.931 to 0.958 for the ISLED. If duration were only a weakened measurement of level of education and the ISLED score a perfect one, the measurement relation for the ISLED should equal 1.0. A deviation from an average of 0.94 to 1.00 may not be earth shattering, but it has a considerable impact on the estimated coefficients (cf. model 5 to model 3). This result not only shows that duration is a reasonable indicator of educational attainment, but also that this measure contains information relevant to the status attainment process that is unique for this indicator.

The measurement indicators in model 4 show that ISCED is a better measure of educational attainment than duration but only by a slight margin. Model 6 confirms this result and reveals that the ISLED performs best, followed by ISCED and duration.
These measurement relations provide direct insight into the loss of information that incurs if we confine ourselves to one indicator of educational attainment:

- By only using duration we lose approximately 20% of the information.
- By only using ISCED, we lose approximately 13% of the information.
- By only using the ISLED, we lose approximately 6% of the information.

These results warrant the conclusion that the best way of improving our results is not through optimal scaling, but by using two separate indicators.

Findings across countries

TO BE COMPLETED FOR THE RC 28 CONFERENCE IN HAIFA IN MAY 2010

8. Summary and conclusions

It was our aim to improve on the existing measures and we have proposed two alternative strategies of achieving this goal. These two strategies share a common maxim: they both attempt a maximal exploitation of all available information. The first strategy, optimal scaling, makes use of the extra detail contained in the country-specific education measures as collected in the ESS. These measures, rendered comparable by way of optimal scaling, perform better than both ISCED and years of education. For the Benelux countries the average gain in explained variance is more than 30% compared to years of education. Compared to ISCED, ISLED leads to an increase in explained variance of 4%.

While the improvements achieved through ISLED are considerable, our second strategy, the use of a multiple-indicator model outperforms it. Compared to ISLED, a combination of two indicators yields an additional average gain in explained variance of 5.5%. Compared to ISCED the average total gain in explained variance is 8.5%, compared to duration 12%. We further conclude that adding a third indicator does not lead to further improvement.

If we compare the sizes of regression coefficients, results point in the same direction. For the Luxembourg case, for example, the effect of education on occupation increases from 0.50 for model 1 to 0.72 for model 5, which is an astounding difference. The effect of education on spouse's education increases comparably, from 0.39 for the worst to 0.73 for the best measure, again an enormous increase. That such tremendous increases in the size of regression coefficients may be due solely to improved measurement should convince even the most sceptical among us that dealing with measurement error is necessary and well worth our while.

What does all this mean then? The first lesson to be drawn is that ignoring measurement error amounts to no less than negligence. While standard procedures to estimate and avoid measurement error are common practice in the measurement of attitudes, we plead for an equally meticulous procedure in the measurement of social background variables. This leads to some general conclusions concerning both data collection and analysis. Anybody setting out to gather new data would be advised to collect multiple parallel indicators of social background variables. The good news is that there is no need to invest much time in trying to produce a perfect indicator. Instead, the mere presence of two or more indicators is sufficient, provided that they are based on independent measurements. In other words, what we need to do is ask more rather than better questions.

As far as data analysis is concerned, we plead for the use of multiple indicator models. Even the use of a mediocre parallel measure, such as duration, leads to a greater improvement of the measurement quality of the education variable than an improvement reached through the optimization of individual measures. Multiple indicator models 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. As our analyses clearly demonstrate just how much we have to gain in terms of accuracy of regression coefficients and explained variance, we hope to have increased awareness of the problems caused by measurement error and possible remedies for it.

While a multiple indicator model deserves preference, we realize that it is unrealistic to expect that from now on all researchers will apply the structural equation modelling techniques necessary for the correction of measurement error. Given that we have also achieved considerable improvement of measurement through optimal scaling, albeit to a lesser degree, we provide country registers of optimal scores for each education level, which can be readily applied in OLS regression analysis of the ESS data. The applications of these scores will, we believe, parallel to the familiar ISEI scores for occupation, noticeably improve empirical results, in terms of both accuracy of regression coefficients and explained variance.

References

Alwin, D.F. (2007) *Margins of Error: A Study of Reliability in Survey Measurement*. New York: Wiley.

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

Bohrnstedt, G.W. &Carter, T.M. (1971) 'Robustness in Regression Analysis', *Sociological Methodology* 3:118-146.

Bourdieu, P. & Passeron, J.-C. (1977) *Reproduction in Education, Society and Culture*. London: Sage.

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

Brint, S. (1984) ''New-Class' and Cumulative Trend Explanations of the Liberal Political Attitudes of Professionals', *American Journal of Sociology*, 90(1): 30-71.

Bunt, G.G. van de, Duijn, M.A.J. van & Snijders, T.A.B. (1999) 'Friendship Networks Through Time: An Actor-Oriented Dynamic Statistical Network Model', *Computational & Mathematical Organization Theory*, 5(2): 167–192.

Churchill, G. A. (1979) 'A Paradigm for Developing Better Measures of Marketing Constructs', *Journal of Marketing Research*: 64-73.

DiPrete, T.A. & Grusky, D.B. (1990) 'Structure and Trend in the Process of Stratification for American Men and Women', *American Journal of Sociology*, 96(1): 107-143.

Doornbos, G. & Kromhout, D. (1990) **'**Educational Level and Mortality in a 32-Year Follow-up Study of 18-Year-Old Men in the Netherlands', *International Journal of Epidemiology*, 19(2): 374-379.

Duncan, B. & Duncan, O.D. (1969) 'Family Stability and Occupational Success', *Social Problems*, 16(3): 273-285.

Duncan, O.D. & Hodge, R.W. (1963). Education and Occupational Mobility: A Regression Analysis. *The American Journal of Sociology*, 68(6), 629-649.

Ganzeboom, H.B.G., de Graaf, P.M. & Treiman, D.J. (1992) 'A Standard International Socio-Economic Index of Occupational Status', *Social Science Research*, 21: 1-56.

Ganzeboom, H.B.G., Treiman, D.J. & Ultee, W.C. (1991) 'Comparative Intergenerational Stratification Research: Three generations and Beyond', *Annual Review of Sociology*, 17: 277-302.

Grusky, D.B. & Van Rompaey, S.E. (1992) ’The Vertical Scaling of Occupations: Some Cautionary Comments and Reflections’, *American Journal of Sociology*, 97(6): 1712-1728.

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

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

Hout, M. & DiPrete, T.A. (2006) ‘What we Have Learned: RC28's Contributions to Knowledge about Social Stratification’, *Research in Social Stratification and Mobility*, 24(1): 1-20.

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

Hyman, H.H., Wright, C.R. & Reed, J.S. (1975) *The Enduring Effects of Education*. Chicago and London: The University of Chicago Press.

Inglehart, R. (1971) ‘The Silent Revolution in Europe: Intergenerational Change in Post-Industrial Societies’, *The American Political Science Review*, 65(4): 991-1017.

ISCED (1997) Manual: http://www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED_A.pdf.

ISLED (2009) Syntax: http://home.fsw.vu.nl/hbg.ganzeboom/isled.htm.

Kalmijn, M. (1994) ‘Assortative Mating by Cultural and Economic Occupational Status’, *American Journal of Sociology*, 100(2): 422.

Kelley, J. (1973) ‘Causal Chain Models for the Socioeconomic Career’, American Sociological Review, 38: 481-493.

Kerckhoff, A.C. (2001) ‘Education and Social Stratification Processes in Comparative Perspective’. *Sociology of Education* (Extra Issue): 3-18.

Kerckhoff, A.C. and Dylan, M. (1999) ‘Problems with International Measures of Education’, *The Journal of Socio-Economics*, 28: 759-775.

Lleras-Muney, A. (2005) ‘The Relationship between Education and Adult Mortality in the United States’, *Review of Economic Studies*, 72(1): 189-221.

Lochner, L. (2004) ‘Education, Work, and Crime: A Human Capital Approach’, *International Economic Review*, 45(3): 811-843.

Mare, R.D. (1981) ’Change and Stability in Educational Stratification’, *American Sociological Review*, 46(February): 72-87.

Mooi-Reci, I. (2008) *Unemployed and Scarred for Life?: Longitudinal Analyses of How Unemployment and Policy Changes Affect Re-employment Careers and Wages in the Netherlands, 1980-2000*. Vrije Universiteit: ISBN: 978-90-9023654-4.

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

Munck, I.M.E. (1991) 'Path Analysis of Cross-National Data Taking Measurement Errors into Account'. In: *Measurement Errors in Surveys*, edited by Biemer, P. et al., New York: Wiley.

Poortman, A.-R. & Kalmijn, M. (2002) 'Women's Labour Market Position and Divorce in the Netherlands: Evaluating Economic Interpretations of the Work Effect', *European Journal of Population*, 18(2): 175-202.

Ross, C.E.R. & Wu, C.-L. (1996) 'Education, Age, and the Cumulative Advantage in Health', *Journal of Health and Social Behavior*', 37(1): 104-120.

Rutter, M., Tizard, J. & Whitmore, K.E. (1981) *Education, Health and Behaviour. Huntington.* New York: Longman.

Saris, W.E. & Andrews, F.M. (1991) 'Evaluation of Measurement Instruments Using a Structural Modeling Approach'. In: *Measurement Errors in Surveys*, Biemer et al. (Eds). New York: Wiley.

Schneider, S.L. (2007) 'Measuring Educational Attainment in Cross-National Surveys: The Case of the European Social Survey', paper presented at the EDUC workshop of the EQUALSOC network, Dijon, 22-24 Nov.

Schneider, S.L. (2008a) 'The International Standard Classification of Education (ISCED-97)', Mannheimer Zentrum für Europäische Sozialforschung.

Schneider, S.L. (2008b) 'Nominal Comparability is not Enough: Evaluating Cross-national Measures of Educational Attainment Using ISEI Scores', *Sociology Working Papers*.

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

Smith, H.L. & 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. & Terrell, K. (1975) 'The Process of Status Attainment in the United States and Great Britain', *American Journal of Sociology*, 81: 563-583.

Ultee, W. & Luijkx, R. (1990) ‘Educational Heterogamy and Father–to–son Occupational Mobility in 23 Industrial Nations: General Societal Openness or Compensatory Strategies of Reproduction?’, *European Sociological Review*, 6: 125-149.

Werfhorst, H.G. van de & Graaf, N.D. de (2004) ‘The Sources of Political Orientations in Post-industrial Society: Social Class and Education Revisited’, *British Journal of Sociology* 55(2): 211-235.

Westert, G.P., Schellevis, F.G., Bakker, D.H. de, Groenewegen, P.P., Bensing, J.M. & Zee, J. van der (2005) ‘Monitoring Health Inequalities through General Practice: The Second Dutch National Survey of General Practice’, *The European Journal of Public Health*, 15(1): 59-65.

Figure 1: A MIMIC model of education

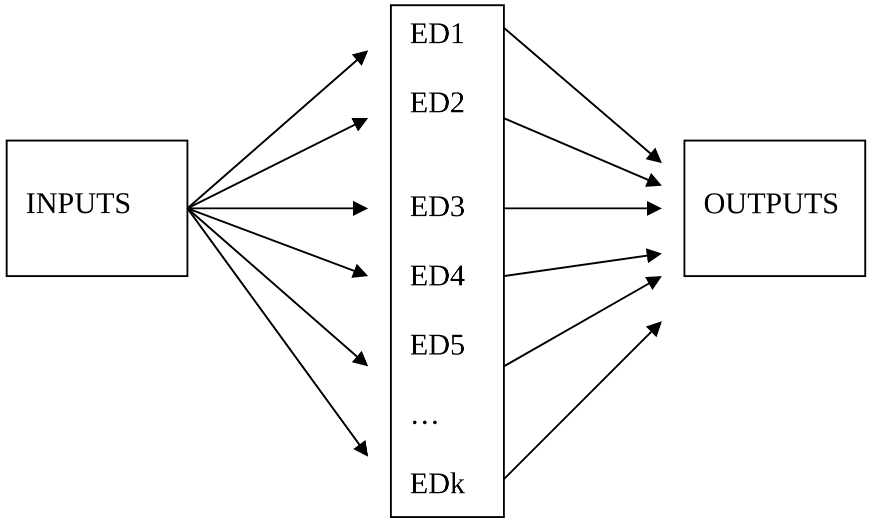

Figure 2: Indirect effects model

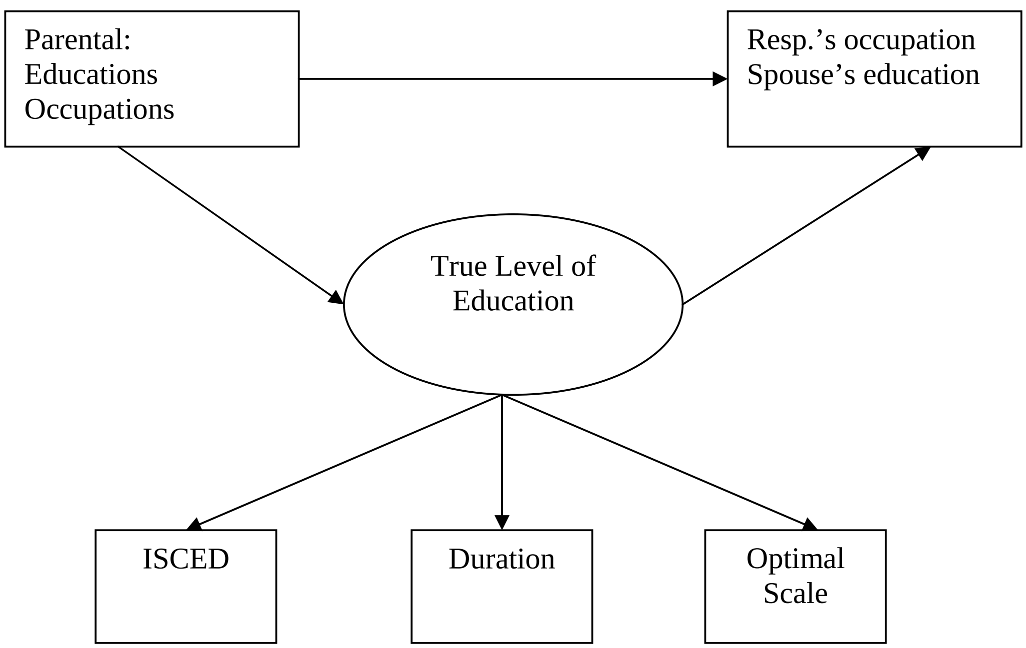

Table 1a: Summary of country-specific education categories and ISCED levels BELGIUM, ESS Round 1-2-3

Cat	Country specific education category	N	Cat	ISCED	N	Opti	ISLED	
							A	B
0	Not completed primary education	51	0	Not completed primary education	51	-1.38	5.9	9.5
1	Primary, basic, and special primary education	456	1	Primary or first stage of basic	456	-1.35	6.0	10.0
2	Lower secondary vocational education	463	2	Lower secondary or second stage of basic	761	-1.01	7.3	18.0
3	Lower secondary general education	298				-0.61	8.9	32.7
4	Higher secondary vocational education	482	3	Upper secondary	1520	-0.68	8.6	29.7
5	Higher secondary technical, or 7th year vocational education	575				-0.16	10.6	54.4
6	Higher secondary general education	463				0.20	12.0	71.0
7	Higher education, short type (HOKT)	702	5	First stage of tertiary	702	0.94	14.9	91.5
8	Higher education, long type (HOLT)	192	6	Second stage of tertiary	612	1.33	16.4	95.9
9	University education	376				1.82	18.3	98.4
10	Doctoral and postdoctoral education	44				1.63	17.5	97.7

Table 1b: Summary of country-specific education categories and ISCED levels NETHERLANDS, ESS Round 1-2-3								
Cat	**Country specific education category**	**N**	**Cat**	**ISCED**	**N**	**Opti**	**ISLED**	
							A	**B**
1	Not completed primary education	48	0	Not completed primary education	48	-1.82	5.4	7.4
2	Primary school or first stage of basic education	376	1	Primary or first stage of basic	376	-1.58	6.0	10.0
3	Lower secondary school, technical training (lbo)	882	2	Lower secondary or second stage of basic	1532	-1.14	7.2	16.8
4	Lower secondary school, theoretical training (mulo,mavo)	650				-0.43	9.0	34.7
5	Short upper secondary professional education (kmbo, vhbo)	113	3	Upper secondary	1436	-0.39	9.1	36.0
6	Upper secondary professional education (mbo)	930				-0.13	9.8	44.5
8	Higher secondary school (mms, havo)	208				0.27	10.9	58.0
9	Pre-scientific secondary school (hbs, vwo)	185				0.69	12.0	71.0
7	Post secondary, non-tertiary education (mbo plus)	350	4	Post secondary, non-tertiary	350	0.38	11.2	61.6
10	Tertiary professional education (hbo)	925	5	First stage of tertiary	1369	1.06	13.0	80.2
11	Tertiary scientific education, university	385				1.73	14.7	91.0
12	Tertiary post-scientific education (teachers, doctors)	59				2.02	15.5	93.7
13	Second stage of tertiary education, Ph.D. education	23	6	Upper secondary Second stage of tertiary	23	2.48	16.7	96.6

Table 1c: Summary of country-specific education categories and ISCED levels LUXEMBOURG, ESS Round 1-2								
Cat	**Country specific education category**	**N**	**Cat**	**ISCED**	**N**	**Opti**	**ISLED**	
							A	**B**
0	No qualification	43	0	Not completed primary education	43	-1.35	5.6	8.3
1	Primary school	331	1	Primary or first stage of basic	751	-1.22	6.0	10.0
2	Upper primary school	161				-0.77	7.4	18.7
3	Complementary school	165				-0.86	7.1	16.6
4	Lower technical secondary school	94	2	Lower secondary or second stage of basic	251	-0.33	8.8	31.9
5	Craftsman diploma	40				-0.50	8.3	26.3
11	General lower secondary school	117				-0.09	9.5	40.9
6	Skilled craftsman	34	3	Upper secondary	818	-0.69	7.7	20.8
7	First professional diploma	42				-0.32	8.8	32.3
8	Second professional diploma	429				-0.33	8.8	31.9
9	First technical high school diploma	46				0.51	11.4	64.7
10	Second technical high school	65				0.30	10.8	56.6
12	Secondary diploma	202				0.69	12.0	71.0
13	Master craftsman diploma	89	4	Post-secondary. non-tertiary	89	0.04	10.0	46.1
14	High school + 2 years university	94	5	First stage of tertiary	448	1.13	13.4	83.3
15	High school + 3 years university	122				1.25	13.8	85.8
16	High school + 4 years university	117				1.92	15.9	94.7
17	High school + 5 years university without obt. dipl.	115				2.13	16.5	96.2
18	Doctorate. PhD	22	6	Second stage of tertiary	22	2.42	17.4	97.6

Table 1d: Summary of country-specific education categories
GERMANY, ESS Round 1-2-3-4

<table>
<tr><th>Cat</th><th>Country specific education category</th><th>R 1</th><th>R 2</th><th>R 3</th><th>R 4</th><th>Opti</th><th>Optix</th><th colspan="2">ISLED</th></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>A</td><td>B</td></tr>
<tr><td>0</td><td>Grundschule nicht beendet</td><td>64</td><td>89</td><td>71</td><td>71</td><td>-1.82</td><td>-1.82</td><td>5.0</td><td>7</td></tr>
<tr><td>1</td><td>Hauptschule</td><td>232</td><td>266</td><td rowspan="3">942</td><td>211</td><td>-1.48</td><td rowspan="3">-1.02</td><td>6.0</td><td>10</td></tr>
<tr><td>2</td><td>Hauptschule & basic voc training</td><td>578</td><td>550</td><td>470</td><td>-0.92</td><td>7.6</td><td>18</td></tr>
<tr><td>3</td><td>Hauptschule & advanced voc training</td><td>151</td><td>115</td><td>91</td><td>-0.61</td><td>8.5</td><td>25</td></tr>
<tr><td>4</td><td>Realschule</td><td>154</td><td>134</td><td rowspan="3">940</td><td>142</td><td>-0.42</td><td rowspan="3">-0.11</td><td>9.1</td><td>30</td></tr>
<tr><td>5</td><td>Realschule & basic voc training</td><td>548</td><td>504</td><td>504</td><td>-0.21</td><td>9.7</td><td>36</td></tr>
<tr><td>6</td><td>Realschule & advanced voc training</td><td>265</td><td>257</td><td>232</td><td>0.09</td><td>10.6</td><td>45</td></tr>
<tr><td>7</td><td>Fachhochschulreife</td><td>25</td><td>24</td><td rowspan="3">86</td><td>26</td><td>0.20</td><td rowspan="3">0.20</td><td>10.9</td><td>48</td></tr>
<tr><td>8</td><td>Fachhochschulreife & basic voc training</td><td>33</td><td>34</td><td>39</td><td>0.07</td><td>10.5</td><td>44</td></tr>
<tr><td>9</td><td>Fachhochschulreife & advanced voc training</td><td>64</td><td>63</td><td>57</td><td>0.32</td><td>11.25</td><td>52</td></tr>
<tr><td>10</td><td>Abitur</td><td>138</td><td>169</td><td rowspan="3">301</td><td>136</td><td>0.92</td><td rowspan="3">0.65</td><td>13</td><td>70</td></tr>
<tr><td>11</td><td>Abitur & basic voc training</td><td>62</td><td>64</td><td>75</td><td>0.75</td><td>12.5</td><td>65</td></tr>
<tr><td>12</td><td>Abitur & advanced voc training</td><td>61</td><td>59</td><td>44</td><td>0.79</td><td>12.6</td><td>66</td></tr>
<tr><td>13</td><td>Fachhochschule</td><td>135</td><td>156</td><td>178</td><td>213</td><td>0.83</td><td>0.83</td><td>12.7</td><td>67</td></tr>
<tr><td>14</td><td>BA</td><td>92</td><td>94</td><td rowspan="3">269</td><td>130</td><td>1.24</td><td rowspan="3">1.78</td><td>13.9</td><td>78</td></tr>
<tr><td>15</td><td>MA</td><td>267</td><td>249</td><td>261</td><td>1.80</td><td>15.6</td><td>88</td></tr>
<tr><td>16</td><td>PhD</td><td>33</td><td>30</td><td>46</td><td>2.94</td><td>18.9</td><td>97</td></tr>
</table>

Table 2a: Model parameters for BELGIUM ESS Round 1-2-3						
Belgium	**1**	**2**	**3**	**4**	**5**	**6**
	ISCED	**Duration**	**ISLED**	**1 & 2**	**2 & 3**	**1, 2 & 3**
A. Structural models						
EDUCATION R.						
FEDUC	0.276	0.261	0.264	0.308	0.286	0.284
MEDUC	0.225	0.188	0.239	0.242	0.243	0.241
FOCC	0.125	0.098	0.144	0.146	0.153	0.144
MOCC	0.014	0.018#	0.023#	0.053	0.030#	0.036
R2	**0.320**	**0.235**	**0.325**	**0.399**	**0.364**	**0.384**
SPOUSE'S EDU						
FEDUC	0.143	0.206	0.144	0.076	0.110	0.116
MEDUC	0.080	0.136	0.077	0.037	0.058	0.060
EDUC	0.460	0.311	0.463	0.577	0.519	0.511
R2	**0.359**	**0.287**	**0.360**	**0.418**	**0.388**	**0.356**
OCCUPATION R.						
FOCC	0.076	0.152	0.060	0.038	0.027	0.040
MOCC	0.045	0.092	0.048	0.029#	0.030	0.031
EDUC	0.572	0.397	0.588	0.611	0.642	0.625
R2	**0.390**	**0.265**	**0.404**	**0.407**	**0.446**	**0.430**
B. Measurement models						
ISCED	1			0.896		0.947
Duration		1		0.743	0.777	0.764
ISLED			1		0.948	0.958
C. Fit statistics						
Chi-square	59.4	240.0	54.4	43.2	49.7	81.7
RMSEA	.037	0.079	0.035	0.035	0. 028	0.043
Standardized parameters. # = non-significant.						

Table 2b: Model parameters for the NETHERLANDS ESS Round 1-2-3						
NETHERLANDS	**1**	**2**	**3**	**4**	**5**	**6**
	ISCED	**Duration**	**ISLED**	**1 & 2**	**2 & 3**	**1, 2 & 3**
A. Structural models						
EDUCATION R.						
FEDUC	0.227	0.212	0.226	0.257	0.237	0.246
MEDUC	0.206	0.224	0.189	0.245	0.214	0.215
FOCC	0.087	0.104	0.120	0.108	0.124	0.130
MOCC	0.069	0.050	0.096	0.073	0.087	0.097
R2	**0.229**	**0.231**	**0.255**	**0.308**	**0.322**	**0.309**
SPOUSE'S EDU						
FEDUC	0.146	0.172	0.134	0.101	0.123	0.104
MEDUC	0.087	0.102	0.085	0.044	0.070	0.057
EDUC	0.427	0.348	0.440	0.513	0.459	0.500
R2	**0.309**	**0.261**	**0.315**	**0.352**	**0.282**	**0.346**
OCCUPATION R.						
FOCC	0.099	0.117	0.073	0.075	0.058	0.045
MOCC	0.069#	0.058	0.010#	0.011#	-0.001#	-0.012#
EDUC	0.530	0.422	0.570	0.546	0.602	0.621
R2	**0.338**	**0.249**	**0.365**	**0.339**	**0.391**	**0.400**
B. Measurement models						
ISCED	1			0.881		0.892
Duration		1		0.825	0.871	0.789
ISLED			1		0.963	0.931
C. Fit statistics						
Chi-square	37.1	144.2	22.4	58.4	55.4	188.71
RMSEA	0.025	0.054	0.017	0.028	0.027	0.042
Standardized parameters. # = non-significant.						

Table 2c: Model parameters for LUXEMBOURG ESS Round 1-2						
LUXEMBOURG	**1**	**2**	**3**	**4**	**5**	**6**
	ISCED	**Duration**	**ISLED**	**1 & 2**	**2 & 3**	**1, 2 & 3**
A. Structural models						
EDUCATION R.						
FEDUC	0.307	0.301	0.298	0.376	0.338	0.321
MEDUC	0.170	0.203	0.219	0.212	0.225	0.232
FOCC	0.069	0.115	0.095	0.116	0.112	0.114
MOCC	0.041	0.003#	0.092	0.057	0.082	0.089
R2	**0.239**	**0.266**	**0.325**	**0.392**	**0.379**	**0.372**
SPOUSE'S EDU						
FEDUC	0.145	0.124	0.088	-0.013#	0.022	0.035
MEDUC	0.141	0.116	0.079	0.053#	0.049	0.046
EDUC	0.394	0.457	0.532	0.671	0.736	0.625
R2	**0.315**	**0.352**	**0.392**	**0.477**	**0.455**	**0.450**
OCCUPATION R.						
FOCC	0.124	0.092	0.075	0.034	0.037	0.040
MOCC	0.086	0.095	0.011#	0.002#	-0.023	-0.015#
EDUC	0.504	0.532	0.650	0.705	0.635	0.720
R2	**0.349**	**0.371**	**0.469**	**0.518**	**0.550**	**0.533**
B. Measurement models						
ISCED	1			0.762		0.782
Duration		1		0.820	0.836	0.831
ISLED			1		0.919	0.931
C. Fit statistics:						
Chi-square	197.4	156.2	52.6	32.8	42.6	71.3
RMSEA	0.097	0.086	0.052	0.028	0.034	0.035
Standardized parameters. # = non-significant.						

Tabel 2 d: Model parameters for GERMANY ESS Round 1-2-3-4						
	1	**2**	**3**	**4**	**5**	**6**
	ISCED	**Duration**	**ISLED**	**1 & 2**	**2 & 3**	**1, 2 & 3**
A. Structural models						
EDUCATION R.						
FEDUC	0.120	0.114	0.139	0.141	0.144	0.145
MEDUC	0.091	0.099	0.086	0.100	0.091	0.093
FOCC	0.160	0.204	0.233	0.243	0.251	0.249
MOCC	0.091	0.147	0.186	0.163	0.180	0.177
R2	**0.134**	**0.202**	**0.251**	**0.267**	**0.285**	**0.283**
SPOUSE'S EDU						
FEDUC	0.154	0.140	0.111	0.101	0.096	0.096
MEDUC	0.134	0.115	0.102	0.091	0.090	0.090
EDUC	0.240	0.278	0.338	0.366	0.372	0.372
R2	**0.158**	**0.173**	**0.201**	**0.215**	**0.217**	**0.217**
OCCUPATION R.						
FOCC	0.170	0.137	0.080	0.057	0.046	0.046
MOCC	0.079	0.042	0.009#	-0.004#	-0.016#	-0.015#
EDUC	0.434	0.487	0.600	0.647	0.664	0.665
R2	**0.303**	**0.330**	**0.414**	**0.453**	**0.462**	**0.464**
B. Measurement models						
ISCED	1			0.743		0.729
Duration		1		0.837	0.823	0.823
ISLED			1		0.940	0.940
C. Fit statistics						
Chi-square	122.58	90.39	39.24	126.47	83.99	59.61
RMSEA	0.055	0.047	0.029	0.030	0.024	0.023
Standardized parameters. # = non-significant.						

Appendix A: Benelux correlation matrices

Cntry		feducyr	meducyr	fisei	misei	edulvl	eddur	isled	isei	seducyr
BE	feducyr	1.000	0.655	0.619	0.429	0.519	0.452	0.520	0.361	0.435
	meducyr	0.655	1.000	0.467	0.546	0.488	0.414	0.492	0.325	0.399
	fisei	0.619	0.467	1.000	0.457	0.420	0.355	0.430	0.335	0.354
	misei	0.429	0.546	0.457	1.000	0.343	0.277	0.333	0.271	0.286
	edulvl	0.519	0.488	0.420	0.343	1.000	0.755	0.932	0.618	0.574
	eddur	0.452	0.414	0.355	0.277	0.755	1.000	0.733	0.476	0.460
	isled	0.520	0.492	0.430	0.333	0.932	0.733	1.000	0.630	0.576
	isei	0.361	0.325	0.335	0.271	0.618	0.476	0.630	1.000	0.430
	seducyr	0.435	0.399	0.354	0.286	0.574	0.460	0.576	0.430	1.000
LU	feducyr	1.000	0.573	0.463	0.493	0.489	0.472	0.513	0.417	0.406
	meducyr	0.573	1.000	0.360	0.463	0.438	0.418	0.467	0.335	0.378
	fisei	0.463	0.360	1.000	0.476	0.320	0.329	0.356	0.312	0.275
	misei	0.493	0.463	0.476	1.000	0.336	0.300	0.386	0.298	0.322
	edulvl	0.489	0.438	0.320	0.336	1.000	0.753	0.902	0.615	0.563
	eddur	0.472	0.418	0.329	0.300	0.753	1.000	0.775	0.591	0.564
	isled	0.513	0.467	0.356	0.386	0.902	0.775	1.000	0.681	0.614
	isei	0.417	0.335	0.312	0.298	0.615	0.591	0.681	1.000	0.504
	seducyr	0.406	0.378	0.275	0.322	0.563	0.564	0.614	0.504	1.000
NL	feducyr	1.000	0.581	0.536	0.423	0.422	0.419	0.441	0.288	0.377
	meducyr	0.581	1.000	0.400	0.504	0.407	0.414	0.417	0.260	0.346
	fisei	0.536	0.400	1.000	0.458	0.322	0.330	0.361	0.283	0.257
	misei	0.423	0.504	0.458	1.000	0.308	0.300	0.342	0.238	0.247
	edulvl	0.422	0.407	0.322	0.308	1.000	0.728	0.951	0.571	0.524
	eddur	0.419	0.414	0.330	0.300	0.728	1.000	0.733	0.478	0.462
	isled	0.441	0.417	0.361	0.342	0.951	0.733	1.000	0.600	0.534
	isei	0.288	0.260	0.283	0.238	0.571	0.478	0.600	1.000	0.365
	seducyr	0.377	0.346	0.257	0.247	0.524	0.462	0.534	0.365	1.000

Appendix B: Distribution of variables

cntry		feducyr	meducyr	fisei	misei	edulvl	eddur	isled	isei	seducyr
BE	feducyr	3641	3476	3338	1323	3634	3609	3641	3464	2709
	meducyr	3476	3724	3353	1367	3717	3691	3724	3540	2766
	fisei	3338	3353	3674	1323	3668	3643	3674	3494	2724
	misei	1323	1367	1323	1486	1485	1475	1486	1427	1093
	edulvl	3634	3717	3668	1485	4102	4061	4102	3886	3024
	eddur	3609	3691	3643	1475	4061	4070	4070	3861	3003
	isled	3641	3724	3674	1486	4102	4070	4113	3895	3030
	isei	3464	3540	3494	1427	3886	3861	3895	3895	2895
	seducyr	2709	2766	2724	1093	3024	3003	3030	2895	3030
LU	feducyr	2063	1960	1914	598	1926	2008	2063	1889	1493
	meducyr	1960	2134	1939	634	1992	2081	2134	1952	1557
	fisei	1914	1939	2148	592	1995	2090	2148	1960	1553
	misei	598	634	592	695	639	675	695	643	514
	edulvl	1926	1992	1995	639	2213	2159	2213	2019	1589
	eddur	2008	2081	2090	675	2159	2316	2316	2119	1663
	isled	2063	2134	2148	695	2213	2316	2379	2171	1706
	isei	1889	1952	1960	643	2019	2119	2171	2171	1572
	seducyr	1493	1557	1553	514	1589	1663	1706	1572	1706
NL	feducyr	4698	4563	4356	1204	4695	4654	4698	4560	3298
	meducyr	4563	4791	4398	1249	4788	4745	4791	4649	3358
	fisei	4356	4398	4660	1189	4657	4618	4660	4524	3263
	misei	1204	1249	1189	1312	1311	1298	1312	1290	868
	edulvl	4695	4788	4657	1311	5134	5085	5134	4969	3561
	eddur	4654	4745	4618	1298	5085	5088	5088	4926	3533
	isled	4698	4791	4660	1312	5134	5088	5139	4973	3564
	isei	4560	4649	4524	1290	4969	4926	4973	4973	3462
	seducyr	3298	3358	3263	868	3561	3533	3564	3462	3564

Appendix C: ISLED codes for the Benelux

(also available at: http://home.fsw.vu.nl/hbg.ganzeboom/isled.htm)

```
Do if (cntry eq "BE" and essround eq 1).
recode edlvbe (1=9.5)  into isled.
recode edlvbe (2=10.0)  into isled.
recode edlvbe (3=18.0)  into isled.
recode edlvbe (4=32.7)  into isled.
recode edlvbe (5=29.7)  into isled.
recode edlvbe (6=54.4)  into isled.
recode edlvbe (7=71.0)  into isled.
recode edlvbe (8=91.5)  into isled.
recode edlvbe (9=95.9)  into isled.
recode edlvbe (10=98.4)  into isled.
recode edlvbe (11=97.7)  into isled.
End if.

Do if (cntry eq "BE" and essround ge 2).
recode edlvbe (0=9.5)  into isled.
recode edlvbe (1=10.0)  into isled.
recode edlvbe (2=18.0)  into isled.
recode edlvbe (3=32.7)  into isled.
recode edlvbe (4=29.7)  into isled.
recode edlvbe (5=54.4)  into isled.
recode edlvbe (6=71.0)  into isled.
recode edlvbe (7=91.5)  into isled.
recode edlvbe (8=95.9)  into isled.
recode edlvbe (9=98.4)  into isled.
recode edlvbe (10=97.7)  into isled.
End if.

recode edlvnl (1=7.4)  into isled.
recode edlvnl (2=10.0)  into isled.
recode edlvnl (3=16.8)  into isled.
recode edlvnl (4=34.7)  into isled.
recode edlvnl (5=36.0)  into isled.
recode edlvnl (6=44.5)  into isled.
recode edlvnl (7=61.6)  into isled.
recode edlvnl (8=58.0)  into isled.
recode edlvnl (9=71.0)  into isled.
recode edlvnl (10=80.2)  into isled.
recode edlvnl (11=91.0)  into isled.
recode edlvnl (12=93.7)  into isled.
recode edlvnl (13=96.6)  into isled.

recode edlvlu (0=8.3)  into isled.
recode edlvlu (1=10.0)  into isled.
recode edlvlu (2=18.7)  into isled.
recode edlvlu (3=16.6)  into isled.
recode edlvlu (4=31.9)  into isled.
recode edlvlu (5=26.3)  into isled.
recode edlvlu (6=20.8)  into isled.
recode edlvlu (7=32.3)  into isled.
recode edlvlu (8=31.9)  into isled.
recode edlvlu (9=64.7)  into isled.
recode edlvlu (10=56.6)  into isled.
recode edlvlu (11=40.9)  into isled.
recode edlvlu (12=71.0)  into isled.
recode edlvlu (13=46.1)  into isled.
recode edlvlu (14=83.3)  into isled.
recode edlvlu (15=85.8)  into isled.
recode edlvlu (16=94.7)  into isled.
recode edlvlu (17=96.2)  into isled.
recode edlvlu (18=97.6)  into isled.
```

```
** GERMANY **.

add value labels edlvde
 (00) Grundschule nicht beendet
 (05) Kein Abschluss
 (10) Hauptschule
 (11) Hauptschule & basic voc training
 (12) Hauptschule & advanced voc training
 (20) Realschule
 (21) Realschule & basic voc training
 (22) Realschule & advanced voc training
 (30) Fachhochschulreife
 (31) Fachhochschulreife & basic voc training
 (32) Fachhochschulreife & advanced voc training
 (40) abitur
 (41) abitur & basic voc training
 (42) abitur & advanced voc training
 (50) fachhochschule
 (60) BA
 (70) MA
 (80) PhD.

do if (cntry eq "DE" and essround ne 3).
recode edlvde
 (00= 7)
 (10=10)
 (11=18)
 (12=25)
 (20=30)
 (21=36)
 (22=45)
 (30=48)
 (31=44)
 (32=52)
 (40=70)
 (41=65)
 (42=66)
 (50=67)
 (60=78)
 (70=88)
 (80=97)
 into isled.
end if.
do if (cntry eq "DE" and essround eq 3).
recode edlvde
 (00= 7)
 (10=16)
 (20=39)
 (30=48)
 (40=62)
 (50=67)
 (70=87)
 into isled.
end if.
```