OVERVIEW, RESULTS AND DISCUSSION

1.1 Introduction

Education is a central variable in the social science research and plays an important part in numerous theories, models and analyses. To begin with, the role of education in society is the focus of a number of theories on mechanisms of social stratification, such as human capital theory (Becker, 1964), status attainment theory (Blau & Duncan, 1967), signalling theory (Spence, 1973), filter theory (Arrow, 1973), screening theory (Stiglitz, 1975), cultural reproduction theory (Bourdieu & Passeron, 1977), institutional theory (Meyer, 1977) and credentialism (Brown, 1995). In these theories education figures as an input and as an output variable. As an input variable it produces a wide range of objective and subjective effects. Most importantly, education affects a number of socioeconomic outcomes, such as employment, occupation, income, prestige and partner (e.g. Blau & Duncan, 1967; DiMaggio & Mohr, 1985; Allmendinger, 1989; DiPrete & Grusky, 1990; Ultee & Luijkkx, 1990; Mare, 1991; Müller & Shavit, 1998; de Graaff, 1998; Bills, 2003; Ganzeboom & Treiman, 2004; Shavit & Blossfeld, 1993). Apart from these direct stratification effects education impinges on many other aspects of people’s lives too, such as health (e.g. Ross & Wu, 1996; Rutter, Tizard, & Whitmore, 1981), crime (e.g. Lochner, 2004), family stability (e.g. Duncan & Duncan, 1969), mortality (e.g. Lleras-Muney, 2005), cultural participation (e.g. Bourdieu & Passeron, 1977), knowledge (e.g. Hyman, Wright & Reed, 1975), values (e.g. Hyman & Wright, 1979; Inglehart, 1971) and attitudes (e.g. Brint, 1984; Davis, 1982). As an output variable education is not simply the resultant of individual characteristics, such as effort, intelligence and interest but reflects parental education and occupation levels. Numerous studies have confirmed the effect of parental background on educational attainment (e.g. de Graaf, 1993; Mare, 1981; Breen & Jonsson, 2005; Erikson & Goldthorpe, 2008). When not thematic itself education is, moreover, frequently used as control or background variable.

Given the pivotal role of education in many research questions, its measurement quality is of critical importance and has direct consequences for the outcomes of statistical analyses. Accurate regression coefficients can only be obtained if sufficient levels of validity and reliability are assured. The measurement of the education variable therefore requires the same level of care as that of other
variables, such as for example social attitudes. This is all the more true when
education is used in comparative research, i.e. research that either compares
countries with different educational system, or between periods within countries
in which the educational systems have changed. Surprisingly, however, not much
care seems to be devoted to the measurement of education in comparative
designs. Still, some studies have recognized the importance of the measurement
of the education and provide guidelines on how best to proceed. Several
international classifications for the measurement of education have been
proposed, such as ISCED ¹, the International Standard Classification of Education
(UNESCO, 2006) and the CASMIN scheme, developed in the Comparative
Analysis of Social Mobility in Industrial Nations project (Brauns, Scherer &
Steinmann, 2003). Other studies have assessed the measurement quality of
these comparative education variables (e.g. Kerckhoff & Dylan, 1999; Kerckhoff,
Ezel & Brown, 2002; Kerckhoff, 1999; Schneider & Kogan, 2008; Schneider,
2009, 210; Braun & Müller, 1997) and have highlighted the importance of high
measurement standards and the consequences of a lack thereof.

The topic of this dissertation is the measurement quality of the education
variable in comparative survey research. Rather than merely assessing it, the
studies assembled here demonstrate that the measurement quality of existing
comparative education variables can actually be improved. The analyses provide
an indication of how much can be gained in terms of regression coefficients and
explained variance if conventionally used methods of comparative measurement,
such as common denominator harmonization ² or the use of a duration measure,
are complemented or combined. The results illustrate how important it is to
be aware of the pitfalls of conventional measurement practices and that an
improvement on the current state of the art in comparative measurement is not
just a luxury problem.

The quality of any measure in a survey is the result of two distinct processes:
data collection and data analysis. As data collection is the primary process,
which by definition precedes any analysis, the ultimate measurement quality of
a variable is bounded by the quality of the original data. The problem, moreover,
remains that analysts who simply want to use a given data set usually have no
influence on the data collection whatsoever and have to make do with what

¹ 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 revision has been launched,
ISCED-2011.

² Common denominator harmonization means that the number of categories found in country-
specific variables is reduced to the categories all countries have in common.
they happen to find in the data. As existing data cannot be recollected, what is needed instead are ways to improve measurement quality that circumvents problems originating in the data collection process. This dissertation attempts to do exactly that: to fill this gap and provide researchers with analytical tools to improve the quality of the education variable after data collection, post-hoc as it were. Despite its secondary nature, data analysis does in fact offer various ways to compensate for or refurbish weaknesses found in the measures as they were originally collected or processed. The potential of such secondary analytical means to improve the education variable, as will be explained below, also has important repercussions for the collection of data. In particular, two different methods are proposed here, both of which rely upon the maximal exploitation of all information contained in existing data.

The first method, optimal scaling, exploits all the details contained in the original country-specific measures, which tend to contain much more detail than their harmonized counterparts. The scaling process results in a novel continuous education measure, labelled the International Standard Level of Education [ISLED] (Schröder & Ganzeboom, 2014). ISLED is a continuous measure that can be used instead of conventional comparative education measures, with the promise of yielding more accurate structural coefficients in statistical analyses. While in principle any categorical education variable can be optimally scaled, the quality of the derivative will critically depend upon the number of categories distinguished in the source variable. The more categories the respective source variable contains, the better its scaled derivative will be.

The second method, latent variable modelling, relies upon the exploitation of all information contained in two different indicators. The improvement in measurement quality is achieved by of model of error correction. Such error correction is not possible if only one indicator is used and is rather dependent on the availability of a second independent indicator. If such a second indicator is available (even if it is an inferior measure), latent variable modelling produces a measurement quality equal or superior to that of any single indicator, including ISLED. Compared to optimal scaling, the applicability of latent variable modelling is somewhat more limited, because it is not universal practice in surveys to collect the two required independent measures. If two measures are available, however, the potential of the method to improve measurement quality can hardly be overestimated. While optimal scaling and latent variable modelling are two independent methods, in order to obtain the best possible results, they may also be combined. The aim of both methods is to improve the measurement quality of the education variable in order to obtain accurate regression coefficients.
1.2 Conventional approaches to the measurement of education level in surveys

Comparative survey designers generally have the choice between two different types of education questions. The first type is formed by so-called qualification questions. Here most often questions are asked about the highest education level a person has achieved. Such questions are typically (e.g. in ESS and ISSP) phrased in a country-specific format, with commonly used national education classifications being presented to respondents as answer categories on a showcard. Such showcards typically vary widely in number of categories used and sensitivity to historical changes in education systems. The second type consists of so-called duration questions. Here the idea is that the length of an individual’s educational career, while abstracting away from the actual level achieved, is functionally equivalent to it and may therefore be used as a proxy. This strategy is also frequently used in comparative surveys, sometime next to qualification questions (e.g. in ESS and ISSP). The questions most often asked in surveys typically are single shot questions that either pertain to the total number of years spent in education or the school-leaving age.

The question type used has important consequences for the data analysis. Qualification questions result in country-specific categorical variables, which, for obvious reasons, are not immediately comparable. If they are to be used in comparative analysis, the categories first need to be made comparable. This is usually done by looking for those elements all country-specific classifications have in common, a process known as common denominator harmonization. This process inevitably leads to a loss of information because some distinctions are relevant in some but not in other countries. As a result, harmonized variables are by definition less informative than their country-specific source variables.

Surveys differ in the way harmonisation is implemented, whereby two main methods may be distinguished: pre- and post-harmonization. Pre-harmonization means that the common-denominator variable is directly implemented in the survey questionnaire. In other words, harmonized answer categories are presented to respondents (at best with country-specific examples) and the underlying country-specific variables are no longer collected. As pre-harmonization means that country-specific distinctions are irretrievably lost, post-harmonization deserves preference. Post-harmonization means that the answer categories presented to respondents are country-specific and harmonized at a later stage. The country-specific source variables may or may not be preserved in the data files, but in general the information remains retrievable.
In contrast to qualification questions, duration questions produce continuous variables that are directly comparable, without requiring much further transformation. While this makes duration questions user-friendly, they have their own drawbacks. One major problem concerns their validity, which often has been questioned. Not everybody is convinced that duration is actually a suitable way to measure level of education. Another problem is that duration questions require some arithmetic on the part of the respondent, which inevitably leads to mistakes and as a consequence to enhanced levels of random measurement error. Moreover, an cursory review of question formats used in even one and the same survey reveals a sometimes astounding variation in the exact question formulations as well as in the accompanying specifications of what respondents should and should not include in the count. For surveys such as the International Social Survey Project [ISSP], this casts severe doubts on the comparability of its duration measure. Be this as it may, duration measures are thus far the only continuous education indicator available. Despite their demonstrably low measurement quality, duration measures are therefore functional and useful, and are frequently used by researchers. However, this dissertation will argue that their best use is as a second measure of level of education.

To sum up, both harmonized and duration measures of education have their own weaknesses and, as is demonstrated in this dissertation, may yield variables with rather high levels of measurement error. With some degree of care it is possible to improve both variable types in data collection. This may be done by synchronizing question formats across countries, by avoiding aggregation error or by introducing better harmonisations; all of these been attempted by the European Social Survey (ESS). Such improvements, however, are not always feasible and, moreover tend to cause new problems. Variables may for example end up not being comparable across rounds any more or new improved variables may only be available in later rounds of data collection.

1.3 Two methods to improve the measurement quality of the education variable in the analysis

Two post-hoc methods are proposed here to improve the measurement quality of the education in comparative research. Both methods use the measures that happen to have been collected in a given survey as the starting point and improve measurement quality by complementing or combining conventional methods respectively.
The first method, optimal scaling, serves to improve the measurement quality of qualification variables. The principle of this method is to optimally scale the educational categories contained in a given educational classification, using the intergenerational status attainment model. In this model education level is the mediating variable in an indirect effects model, which contains parental education and occupation levels as input and respondent’s occupation and partner’s education as output variables. These variables serve as criterion variables, on which the educational categories are scaled. The scaling is considered optimal when the direct effect of inputs on outputs is minimal and the indirect effect running via education is maximal. Figure 1.1 displays the model used for the scaling procedure.

This scaling procedure can in principle be applied to any categorical education variable, whether country-specific or harmonized. Its greatest potential, however, lies in the scaling of detailed and unharmonized (even: unharmonizable) country-specific variables. As country-specific variables tend to contain a great deal more information than their harmonisations, their scaling may produce a variable with increased explanatory power. In this dissertation scale scores are derived for the
country-specific variables of the ESS (ESS R1-5), yielding a novel comparative variable labelled the International Standard Level of Education (ISLED). When the scaling methodology is applied to harmonized qualification variables, which is also done in this dissertation, the advantage of the resulting variables lies less in its increased explanatory power, but rather in the increased user-friendliness. If the source variable contains a lot of detail, as for example is the cases for the new harmonization introduced in the ESS in round 5 (27 categories), the explanatory power of the scaled variable is the same as that of the underlying categorical variable. The scaled variable is, however, continuous and much easier to use than a categorical variable with such a large number of categories. As the analyses presented here unequivocally demonstrate, moreover, the measurement quality of the scaled variable is much higher than that of the duration measure, which would be the alternative continuous indicator.

In sum, ISLED scale scores then have two important advantages. First, they render categorical qualification variables continuous and if these are country-specific, make them comparable. Second, by exploiting the extra detail contained in the country-specific variables, which gets lost through harmonization, the resulting variables have higher measurement quality. Given that any qualification variable can be scaled, the applicability of the method is in principle very broad. As, however, the improvement in measurement quality is greatest when it is applied to country-specific variables, it is here that the method is at its best.

The second method, latent variable modelling, improves measurement quality by compensating for the weakness of one indicator using the extra information contained in a second, independent indicator. The reason why the combination of two indicators in one measurement model improves measurement quality is mainly that this allows for the correction of the measurement error contained in either of the indicators. Figure 1.2 displays a latent variable measurement model and shows how the education level of a respondent (EDU) is modelled as a latent variable with two indicators: the duration measure EDDUR and either ISLED or a harmonized qualification measure (e.g. ISCED). Due to the correction of random measurement error (indicated by single errors pointing to the measured variables), latent variable modelling by definition maximizes measurement quality.

If, as is the case in Figure 1.2, the model is expanded with a second latent variable (PEDU) for the education level of another person whose education level has been measured with the same indicators (here respondent’s partner), it becomes
possible to also correct for systematic measurement error: this error becomes tractable by repeating the measurement. This error is modelled here with the connected arrows pointing to the respective indicators of the same type, for example EDDUR (duration of education respondent) and PEDDUR (duration of education partner).

It is safe to say that any measure contains some amount of measurement error. Latent variable modelling is the only way to fully correct it. Accordingly, latent variable modelling bears more potential than the improvement of any single indicator, including ISLED. Compared with ISLED, however, latent variable modelling is much more limited in its practical applicability. First, it requires that indeed two independent education measures are collected in one survey, which is not general practice. Second, it requires the use of simultaneous equation modelling, which is not always desirable and not accessible to any analyst. Apart from these practical limitations, however, latent variable modelling is the best method available to improve the measurement quality of the education variable and therefore deserves to be applied as widely as possible.
1.4 Underlying theoretical assumptions

Both optimal scaling and latent variable modelling are grounded in substantive as well as methodological theory. To begin with, optimal scaling is embedded in the status attainment model (Blau & Duncan, 1967) and positional good theory (Hirsch, 1976). In order to empirically determine the position of educational qualifications and to derive an ISLED-score for them, we need criterion variables. Suitable criterion variables are variables that are directly and strongly associated with education level. Such variables can be found in the status attainment model (Blau & Duncan, 1967). In this model, the level of education is the mediating variable in an intergenerational status transfer process, in which social inputs, such as parental education and occupation levels, determine social outcomes, such as occupation and education of the partner, via education level. Apart from providing us with suitable criterion variables for the scaling procedure, the status attainment model also generated the theoretically informed definition of education level as the mechanism by which social backgrounds are converted into social outputs. For these reasons the status attainment model is fundamental to this dissertation. It serves both as the model for the optimal scaling procedure as well as for its validation.

Positional good theory (Hirsch, 1976; Thurow, 1975) provides the rationale for the one-dimensional hierarchy that is assumed to be underlying ISLED. The idea here is that in all countries educational qualification can be hierarchically ordered on a single continuum. The position of a given qualification is to some extent determined by the institutionalized structure of the national education system. In other words, there is a nominal or systemic component, which depends on things like programme length, the chronological ordering of successive programmes and formal entry requirements. Instead of relying on such institutional factors to establish the position of a given educational programme in the hierarchy, this position may also be inferred empirically by means of optimal scaling. As positional good theory provides a substantive interpretation for the derived scale scores as an indication of the value of a given educational qualification in society, it is a second theoretical anchor for ISLED.

Methodologically, the optimal scaling procedure is grounded in classic measurement theory (Kelley, 1973), which provides an empirical test for the quality of the education variable. Education is regarded as the mediating variable in an indirect effects (or causal chain) model, where social resources generate social outcomes via education, which functions as a transfer mechanism. The
indirect effect model, as displayed in Figure 1.3, explains how the input variables, parental educations and occupations, determine a person’s occupation and the education level of their partner both directly and indirectly via education. According to classic measurement theory, the quality of the mediating variable will affect the relative sizes of the direct and indirect effects in this model. The better the measurement quality of the education variable, the smaller the direct effects of inputs on outputs will be, and consequently the larger the indirect effects via education. By the same token, the relative size of the direct and indirect effects may be used to infer the quality of measurement of the education variable. The better the measurement quality of the mediating variable is, the smaller the remaining direct effect will be. Consequently, if education is measured with a single indicator, the quality of that indicator is decisive for the results. When comparing single indicator models, a larger indirect effect denote better measurement. The better the indicator, the smaller the direct effect and the larger the indirect effects, whereby the best results will be achieved if two indicators are combined in a latent variable model, such is in Figure 1.3.
1.5 Chapter overview

The general structure of this dissertation is as follows: two chapters are devoted to the derivation of ISLED and two to its validation. While ESS-data were used for the derivation of ISLED, ISSP-data were used for the validation. The derivation and validation chapters are linked. The country-specific ISLED, which is derived in chapter 2, is validated for the Netherlands in chapter 4. The universal ISLED, which is derived in chapter 3, is validated for all countries it is available for in the ISSP in chapter 5. Table 1.1 provides an overview of all chapters and summarizes what was done, which data and method were used as well as the main results.

Chapter 2

In chapter 2 two methods are introduced and developed to improve the measurement quality of the education variable. The first method, optimal scaling, concentrates on how to better measure the level of education.

The second method, latent variable modelling, focuses on how to better model level of education. For both methods data were used from rounds 1-4 from the European Social Survey (ESS R1-R4). ISLED is derived by way of optimally scaling all country-specific education categories the ESS in an extended status attainment model. Based on two sociological theories, status attainment and positional goods theory, education level is defined as an intervening variable in an indirect effects model, in which social backgrounds determine social outcomes via education level. Within this model optimal scale scores are generated by minimizing the direct effects of social background on social outcomes, while at the same time maximizing the indirect effects that are mediated by education.

As a result each and every education category found in the country-specific ESS variables receives a score value. The scores are made comparable by calibrating on the duration variable, the alternative indicator of education level that is also contained in the ESS data. After applying an anti-logistic transformation on standardized scale scores, ISLED scores range between 0 and 100 and can be used as a continuous indicator of level of education. Subsequently, ISLED is compared with the two indigenous ESS-measures, the duration measure and the ESS common denominator harmonization, a five-category version of the International Standard Classification of Education [ISCED-97], using the same status attainment model. Alternating the respective education measures as single indicators in the model reveals that ISLED outperforms both ESS standard measures. Next, in order to further improve the measurement quality of the
education variable, the potential of modelling is explored. Instead of using single indicators, now two indicators are combined in latent variable models. This type of modelling further improves measurement quality by means of correcting random measurement error.

CHAPTER 3
In chapter 3 the ESS data of Round 5 (ESS-R5) are analysed. In this round ESS introduced new, much more detailed education variables. Not only the country-specific variables were adapted and contain more detail, but also the harmonized variables. In particular, a new variable EDULVLb was introduced, which is based on a new version of the International Standard Classification of Education [ISCED] 2011. This new variable contains as many as 27 categories. In practice they are not all used for all countries, but in any case the variable is much more detailed than its predecessor (EDULVLa), which only contained five categories. As EDULVLb is a harmonized variable, its ISLED-scale values are now homogeneous across countries.

Applying optimal scaling to the country-specific as well as to the new and the old ESS harmonisations yields as many as five different ISLED-scaled variables. This makes it possible to isolate the improvement brought about by the scaling procedure from that caused by a difference in detail. Moreover, in ESS-R5 the education variables are available in the same detailed format for multiple persons (respondent, partner, parents), which can be scaled in the same way. This makes it possible to establish the cumulative effect each indicator has on the regression coefficients in the structural model. In line with expectations, the analyses reveal that, generally speaking an increase in detail (number of categories) leads to an improvement in measurement quality. Consequently, the country-specific ISLED performs best again, but now closely followed by the new detailed harmonized variable based on ISCED-2011 and the compact new harmonization EISCED. The old harmonisations turn out to be much weaker. All categorical variables surpass the duration measure in quality. The quality differences are reflected in the regression coefficients within the status attainment model. The better the education indicator used, the smaller the direct effects of parental background on social outcomes, the larger the indirect effect that runs via education and the larger also the explained variance in all dependent variables.

CHAPTER 4
In chapters 2 and 3 ISLED is tested with the same data and within the same model used to derive it. In chapter 4 ISLED is validated on fresh data, namely
the Dutch data of six rounds of the International Social Survey Programme (ISSP-NL, 2003-2008). These data contain the required variables and also have the advantage of providing two independent education measures not only for the respondent, but also for respondent’s partner. For this reason, the data are not only suitable for a validation of ISLED, but make it possible to carry out two additional analyses. First, it is now possible to not only correct for random measurement error, but also for correlated error. Second, by applying a model first introduced by Saris & Andrews (1991), the measurement coefficients that indicate the measurement quality of a given measure, can now be decomposed into a validity and a reliability part.

ISLED-scores are assigned to the ISSP country-specific variables based on the universal ISLED based on ISCED-2011 as developed in Chapter 3. The newly derived variable is compared with the indigenous ISSP measures, a harmonized qualification measure and a duration measure. The results are clearly in favour of ISLED, which scores best on all accounts. It contains the lowest amount of random and no correlated measurement error and consequently has the highest validity as well as reliability. Our analyses furthermore reveal that the ISSP harmonization performs remarkably well, while the duration measure turns out to be of much lower quality.

**Chapter 5**

While in chapter 4 ISLED was validated on fresh data, the validation was restricted to one country, the Netherlands. In chapter 5 ISLED is once more submitted to the test, but this time using data from the ISSP Social Inequality IV module from 2009 (ISSP Research Group), for all European countries that participated. In order to have a benchmark to evaluate ISLED against, first the country-specific ISSP education variable was optimally scaled, using first and current occupation, as well as parental occupations as criterion variables. Moreover, the universal ISLED-scores based on ISCED-2011 were applied to the country-specific ISSP variables. Subsequently, ISLED is compared with the same indigenous ISSP education measures as in chapter 4, as well as the optimized ISSP-variable. As this latter variable was derived on the same data, it is not surprising that it yields the best results. However, it is closely followed by ISLED, which not only proves to be perfectly adequate to use on fresh data, but outperforms the two indigenous ISSP education indicators for all European countries. It must be admitted that the difference between ISLED and the ISSP-harmonization DEGREE is very slight and only shows up in the third decimal. The difference between ISLED and its continuous rival, however, is substantial.
Table 1.1: Summary of chapters

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Goal</th>
<th>Data</th>
<th>Method</th>
<th>Results</th>
</tr>
</thead>
</table>
| Chapter 2 | Introduction of two methods to improve measurement quality of education variable:  
  - optimal scaling  
  - latent variable modelling | - ESS R1-4 (2002-2008)  
  - 35 European countries  
  - Sample: 25-74, no students  
  - N=150,567  
  - Education variables: EDLVXX, EDULVLa, EDUYSR  
  - Criterion variables:  
    Input: Father’s and mother’s education and occupation  
    Output: Respondent’s occupation, partner’s education | - optimal scaling of country-specific ISLED-scores  
  - latent variable modelling  
  - comparing different single indicator and latent variable models | - country-specific ISLED scores for 35 countries (online appendix)  
  - ISLED improves measurement quality and yields better results than indigenous ESS education indicators  
  - latent variable modelling optimizes measurement quality and yields best possible results |
| Chapter 3 | Comparison of ESS education measures | - ESS R5 (2010)  
  - 25 European countries  
  - Sample: 18-74, no students  
  - N=41,264  
  - Education variables: EDLVa/b/c/dXX, EDULVLb, EDULVL, EDULVLa, EISCED, EDUYSR  
  - Criterion variables:  
    Input: Father’s and mother’s education and occupation  
    Output: Respondent’s occupation, partner’s education | - ISLED scores are developed for 5 harmonized variables as well as for country-specific variables of R5  
  - comparing different single indicator and latent variable models | - country-specific ISLED for new more detailed country-specific ESS education variables  
  - universal ISLED scores based on new detailed harmonized education variable based on ISCED-2011 are presented, which can be used for any survey or country  
  - ISLED scores for remaining old as well as new ESS harmonisations |
| Chapter 4 | - Validation of ISLED on Dutch ISSP data  
  - correction for systematic measurement error  
  - distinction validation and reliability | - ISSP-NL 2003-2008  
  - Sample: 25-74, no students  
  - N=5,732  
  - Education variables: DEGREE, EDUCYRS  
  - Criterion variables:  
    Input: Father’s and mother’s education and occupation  
    Output: Respondent’s occupation, partner’s education | - applying Dutch ESS-derived country-specific ISLED scores to Dutch ISSP country-specific variable  
  - MTMM-model  
  - correction for systematic measurement error  
  - comparison of 3 single-indicator models and latent variable model  
  - comparing different single indicator and latent variable models | - ISLED can be used with these data without any problem  
  - ISLED produces best results:  
    * highest reliability and validity  
    * lowest amount of random measurement error  
    * no systematic error (duration measure does)  
  - ISSP-harmonization is second best  
  - Duration measure is poorest measure by all standards |
| Chapter 5 | - Validation of ISLED on cross-national ISSP data | - ISSP, Social Inequality IV module (2009)  
  - Sample: 25-74, no students, European countries only  
  - N=25,999  
  - Education variables: DEGREE, EDUCYRS  
  - Criterion variables:  
    Input: Father’s and mother’s occupation  
    Output: first and current occupation respondent | - applying ISLED coding to ISSP country-specific variables, based on ISCED-2011  
  - comparing different single indicator and latent variable models | - ISLED can be applied to international ISSP data without any problem  
  - ISLED is on average best indicator  
  - ISSP-harmonization is second best  
  - Duration measure is poorest measure by all standards |
In a further step we compare the effect of the measurement quality of the education variable on the coefficients in the structural model. Contrasting the worst (duration as single indicator) with the very best model (double indicators) reveals that the difference in explained variance in the three dependent variables may be considerable, ranging from 3.5% in current occupation, via 5% in education to as much as 9% in first occupation.

1.6 Results

The most important contribution of this dissertation is, no doubt, the construction of a novel comparative education measure, the International Standard Level of Education (ISLED). ISLED is a theoretically grounded and empirically derived measure and is presented in two editions, one country-specific and one universal, which are derived from different source variables. ISLED is based on a clear theoretical conceptualization of education level as transferring social status from one generation to the next. In particular, education is defined as the mechanism by which parental educations and occupations determine an individual’s occupational status and likelihood to attract a high educated partner. Like duration measures, ISLED is a continuous measure, which is easy to use in statistical analyses. Compared with conventional measures, be it qualification or duration measures, in all the analyses presented here the country-specific ISLED excels as the education measure with the highest measurement quality. In other words, the country-specific ISLED is found to contain the lowest amount of random and correlated measurement error and for the data at hand turns out to be the variable with the highest validity and reliability.

The introduction by ESS in round 5 of a new very detailed harmonized education variable based on ISCED-2011, EDULVLb, made it possible to derive a universal edition of ISLED, which bears even greater potential than the country-specific one. The advantage of this universal ISLED is that it is homogeneous for all countries (the scale scores per level are the same for all countries) and that it can easily be applied to other data. Especially, once country-mappings will become available for ISCED-2011, ISLED-scores can in principle be assigned to just any national education classification in any survey. This makes ISLED the truly international measure of education level it was originally conceived of. The validation study presented here in chapter 4 suggests that this universal ISLED cannot only be a new standard and user-friendly measure of education level, but also that it may also substantially improve results, especially compared with its frequently used continuous competitor, the duration measure. The measurement quality of the
universal ISLED is, moreover, only marginally weaker than that of the country-specific edition.

Apart from ISLED, this dissertation introduces a second method to improve the measurement quality of the education variable: latent variable modelling (Bollen, 1989). This method, which is common practice for the measurement of social attitude variables, is here applied to the measurement of a social background variable, level of education. The method combines two independent education indicators, which makes it possible to correct random measurement error, maximizing the ultimate measurement quality of the education variable. As error correction is not possible with a single indicator, latent variable modelling is a valuable complement to ISLED and sets the standard against which ISLED and other measures must be judged. If, moreover, two independent variables are available for another person, for example the partner, it is possible to not only correct random but also correlated error. Only such full error correction yields truly unbiased regression coefficients in structural models.

The accumulated empirical evidence of all the analyses is summarized in Table 1.2. As the analyses consist of two derivation and two validation studies, it contains two sets of analyses for each of the two datasets used (ESS and ISSP). This yields four sets of measurement coefficients for ISLED as well as measurement coefficients for the respective indigenous education measures it is compared to. Table 1.2 unequivocally singles out the country-specific ISLED as the variable with the highest measurement coefficient, oscillating around 0.95, closely followed by the universal ISLED and EISCED. The table also shows that the measurement coefficients for both editions of ISLED are remarkably homogeneous across surveys. As for the harmonized measures, the ESS-R5 data show that the quality of the measure depends on the amount of detail (the number of distinguished categories) that is retained in the harmonized variable. This effect, however, is not linear. As a comparison between EDULVL and EISCED (both contain seven categories) illustrates, a comparatively small number of additional categories can produce very different results. Apart from the number of categories, the measurement quality of a harmonized variable also turns out to crucially depend on the way national categories are aggregated into the broader categories of the respective harmonized variable. In all analyses it is the duration measure that comes out as the weakest measure. For all education variables compared, it holds that the measurement quality differs across countries. Here too, however, ISLED excels as the most stable variable in the sense that the distribution of the measurement coefficients has the lowest level of dispersion in all the cross-national analyses.
In chapter 4, which is confined to Dutch data, it was possible to correct for correlated systematic error. The results once again favour the country-specific ISLED, which, in contrast to the ISSP duration measure, turns out to be free of correlated error. For the Dutch ISSP-data it was, furthermore, possible to empirically dissect measurement quality into a validity and a reliability part, a type of modelling introduced by Saris & Andrews (1991), which was here for the first time applied to the measurement of a background variable. The country-specific ISLED stands out once again as the measure with both the highest reliability and highest validity.

Finally, the analyses presented in chapter 3 may prove to be of additional value as a reference for researchers who set out to analyze ESS education data. Given the wide use of ESS data, this is an important contribution in its own right. The chapter contains descriptions of both the new and the old ESS education variables, including changes between rounds. While it is possible to retrieve this information from the ESS website, the information there is scattered across different files, making it hard for users to put all the puzzle pieces together. This dissertation chapter provides a concise overview of all ESS education variables and, more importantly, a systematic assessment of their respective measurement quality. The analyses warrant the conclusion that the revision of the ESS education variables was highly successful. Both the country-specific source variables and the two new harmonisations are shown to be a major improvement over their predecessors. While it is predictable that the 27-category variable has better measurement quality, the high quality of the seven-category harmonization variable EISCED, is remarkable.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ISLED (country-specific) variable</td>
<td>27</td>
<td>0.949</td>
<td>0.960</td>
<td>0.952</td>
<td>0.941</td>
</tr>
<tr>
<td>ISLED (universal)</td>
<td>n.a.</td>
<td>0.859</td>
<td>0.866</td>
<td>0.782</td>
<td>0.857</td>
</tr>
<tr>
<td>Duration</td>
<td>7</td>
<td>0.892</td>
<td>0.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old ESS harmonization</td>
<td>7</td>
<td>0.892</td>
<td>0.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New ESS-harmonization</td>
<td>7</td>
<td>0.947</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New ESS-harmonization</td>
<td>5</td>
<td>0.902</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISSP harmonization</td>
<td>8</td>
<td>0.931</td>
<td>0.936</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>150,567</td>
<td>41,264</td>
<td>5,732</td>
<td>25,999</td>
<td></td>
</tr>
</tbody>
</table>

NB: For ISLED the number of categories concerns the source variables. For the country-specific variables it varies per country. For the universal ISLED, the source variable is EDULVLb with 27 categories.
1.7 Contentious issues

The research presented here is based on certain assumptions. While there are good reasons for them, I acknowledge that some of the choices made may be questioned. A first contentious issue is the theoretical basis of ISLED. In principle ISLED is the outcome of empirically driven research. As stated above, the analyses are informed by sociological theory, but the choice of theories is not the only one possible. Positional good theory, for example, assumes that there is a single hierarchy of job seekers. But is this assumption justified? One could for example argue that, rather than one, there are several job queues and that job queues do not only depend on the education level, but also on the educational field and the labour market segment. This could imply that job seekers are only in the same queue as long as they are in the same field and aim for jobs in the same labour market segment, in other words that doctors do not compete for jobs with lawyers. In defense of the choices made here, I would argue that even if this were the case, the hierarchy of the most relevant outcome of the status attainment process, income, is unquestionably one-dimensional.

A second point of criticism concerns the empirical findings. It might for example be argued that ISLED scale scores can vary between men and women, and that different scale scores are needed for these two groups. Given, however, that men and women operate in the same labour market, a conscious choice was made not to go along this road. Differences between groups as well as changes over time are difficult to assess if the measure changes as well. For this very reason, it was the explicit aim of this dissertation to produce one standard measure of education that can be applied in different contexts and which can be used to assess the differential effects or determinants of education level for any group or time period.

A third controversial issue concerns the scaling methodology used. To begin with, the choice of criterion variables may be questioned. Other choices would have been possible. One alternative would have been to, like with the International Socio-economic Index of Occupation (ISEI) (Ganzeboom, de Graaf & Treiman, 1992), use single input and output variables instead of several. By the same token, different types of criterion variables could have been chosen. For example ISLED could have been derived using only the occupations of the parents as input variables, rather than their educations and occupations combined. Similarly, different outcome variables are feasible, such as for example income. The variables used now for the derivation of ISLED are all
directly associated with education and include the two variables most strongly
correlated with education level, father’s and partner’s level of education. In other
words, the criterion variables used do pick up the largest part of the variance in
the education variable. Furthermore, this approach integrates the approaches
used in previous scaling attempts, in which either a single income as in cause-
proportional scaling (Smith & Garnier, 1987) or a single outcome as in effect-
proportional scaling (Treiman & Terrell, 1975) were used as criterion variables.
By integrating the two into a cause-and effect-proportional scaling, the impact
of any single criterion variable on the scale scores could be drastically reduced,
countering an important criticism of previous scaling approaches.

A fourth matter of concern is the algorithm used for the derivation of ISLED,
which is rather coarse. Input and output variables are used unweighed, with no
attention being paid to their different level of association with education. The
algorithm used, however, has the advantage of being simple and functional. As
a more refined algorithm would likely have a negligible effect on the results, a
conscious choice was made in favour of parsimony.

A fifth issue of debate is using of the duration measure to make ISLED scale scores
comparable. Here too, alternative options would have been possible (and have
been explored). One alternative option would have been the use of common
anchor points (e.g. end of primary and secondary school). Anchor points,
however, rely on the same common denominator principle as harmonization
and consequently pose the same problems. Anchor points would need to be
reconciled across a large number of countries, when there are no hard and fast
criteria to decide which programmes are of the same level. In order to avoid this
problem, preference was given to the use of the duration measure, which like
the within-country ISLED scale scores, has an empirical basis. The ESS duration
measure, moreover, has been asked in the exact same format in all countries and
rounds, producing a highly consistent and comparable variable.

A sixth and final matter of contention concerns the circularity induced by the
validation model, in particular the test studies reported in chapter 2 and 3, in
which ISLED is compared to other ESS education measures. Here, not only the
validation model is exactly the same as that used for the derivation of ISLED,
but also the same data are used; it may be argued that it is an artifact of the
derivation process if ISLED performs better than the other measures. This
criticism is addressed by including the two validation studies that both use
different data, the ISSP, and at least one of which in part uses some different
variables in the status attainment model.
To sum up, the methodology used for the derivation of ISLED, while being vulnerable to criticisms on a number of points, is the result of a lengthy and thorough process of thought, including many empirical analyses to explore alternative options. It was informed by the state of the art in scaling, continuously revised and adapted and is the best possible choice within the limitations of this study. The theoretical and methodological grounding of ISLED as well as the consistently high measurement quality of the measure evidence the level of care that was used to make it.

1.8 Limitations and suggestions for further research

Despite its merits, it must be acknowledged that the research presented here has a number of limitations. One set of limitations concerns the derivation of ISLED, another its validation. As for the derivation of ISLED, a first important limitation is the number of countries ISLED is available for. Since we used ESS-data to derive it, ISLED is so far confined to European countries. The applicability of ISLED in non-European contexts, however, is within reach. All that is needed is OECD to provide country-mappings for ISCED-2011. It is a matter of time that this will be achieved and that (the universal) ISLED can be applied in a non-European context.

A second limitation concerns the quality of the source variables used to derive the country-specific edition of ISLED. With the benefit of hindsight, it might have been preferable to base the country-specific ISLED on the new, more detailed country-specific variables that were introduced in round 5. This was in fact done in the analyses of chapter 3, but the estimates are based on much smaller samples because for the time being only ESS round 5 data contain these new variables. Future work could redo the analyses combining data of several rounds. However, the results of chapter 3 indicate that an ISLED developed on more detailed source variables would yields only marginally better results (cf. Table 1.1).

A third limitation concerns the two editions of ISLED. Again with the benefit of hindsight, it may have been preferable to focus on the development of universal ISLED scores from the start. Given that the advantage of a country-specific ISLED over the universal ISLED is rather slight for the ESS round 5 data and that the universal ISLED actually produces better results than the country-specific variable derived for rounds 1-4 (cf. Table 1.1) and that, moreover, an application of the universal ISLED-scores to fresh data is much more straightforward, such
a design might have been sufficient to meet the goal of this dissertation. As the appropriate data became available much too late to revise the project, however, this option was not actually feasible. As a result, this dissertation presents two editions of ISLED, which may cause some confusion.

A fourth set of limitations concerns the validation of ISLED. ISLED, it must be acknowledged, still needs much more testing. First, it has so far only been validated with ESS and ISSP-data. Given that ESS-data were also used to derive ISLED, these analyses cannot, as stated above, avoid an element of circularity and critics may say that ISLED’s superiority is merely an artifact of the derivation process. The only independent data-set used for validation is consequently the ISSP. While the ISSP results are encouraging, ISLED must still be tested with different data, such as for example the European Value Survey (EVS). Second, thus far ISLED has only been tested within the status attainment model. While, on a positive note, one of the validation studies at least used some different criterion variables within the status attainment model, more analyses need to be done. Such variables must be strongly associated with education level. Even within stratification theory, the pool of possible criterion variables is not exhausted. An obvious alternative outcome variable would for example be income. Alternative validation criteria outside of status attainment theory could be cultural participation or health. While I am confident that it will, future research still needs to prove that ISLED passes these tests as well.

A fifth and final limitation relates to the duration measure. In the latent variable model two independent education measures have to be available. One of these measures is the duration measure. This measure is needed to make it possible to assess the measurement quality of the first indicator. For latent variable modelling the quality of this second indicator is rather inconsequential. In principle, however, given that duration is probably the most frequently used indicator of level of education, its measurement quality is important in its own right. While the ESS duration measure is based on a single question format that remained stable across rounds, this is not always the case. The exact question formulations of duration questions vary considerably across and even within surveys (cf. ISSP). It may therefore be worth investigating which type of question format actually works best. In other words: whether a question on the length of the educational career yields different results than a question on the school-leaving age. In view of the low measurement quality of the duration variables in all the analysis presented here, such research may ultimately lead to an improvement of the measurement quality of duration measures.
1.9 Recommendations

The results of the analyses presented here have repercussions for both data collection and data analysis. To begin with data collection, the systematic comparison of the various different education measures contained in the ESS and the ISSP has revealed that they differ in a predictable way. As for qualification variables, it generally holds that their quality improves with the amount of detail they contain and that any loss of detail, i.e. any loss in the number of distinguished categories, attenuates regression coefficients in structural models. A first recommendation is therefore that country-specific source questions asked in questionnaires should be as detailed as possible and that these country specific variables should be made accessible for users.

Duration measures, by comparison, while having the advantage of being more straightforward in use, turn out to be more error-loaded than any of the qualification measures and consequently need to be used with some caution. Despite the relatively poor measurement quality of duration measures, they are extremely valuable as the second indicators needed for latent variable modelling. A second recommendation for data collectors is therefore that all comparative surveys should include these two independent questions on education. Ideally this information should be collected not only for the respondent, but also for other persons. Only if this is the case, both random and correlated measurement error can be corrected and accurate regression coefficients obtained.

As far as data analysis is concerned, two methods are proposed to improve measurement quality in the education variable. Both these methods have been shown to improve the results achieved using any of the conventional indigenous measures found in the surveys. To begin with, two editions of ISLED are presented, both of which are user-friendly continuous indicators of education level, which have the potential to considerably improve results. One ISLED is country-specific, the other one universal and both have their own applications. The country-specific ISLED is particularly suited for the analysis of ESS-R1-4 data, where it has been shown to be the best single indicator. While it is in principle possible to apply it to other data as well, here, due to its grounding in ISCED-2011, it is the universal ISLED that deserves preference. In the future especially, once ISCED country-mappings have been renewed, ISLED should be very straightforward to apply. As long as these mappings are still lacking, its applicability may be more awkward and require some background knowledge. In a European context, however, application should be straightforward. A third recommendation is
therefore that given ISLED has the potential to improve results, in a European context at least, it may and should be used.

The merits of ISLED notwithstanding, we have achieved the very best results by means of latent variable modelling. Correction for random (and if possible correlated) error proves to outperform any single indicator, including ISLED. A fourth and last recommendation is therefore that latent variable modelling is applied wherever feasible.