Measuring and Modeling Level of Education in Comparative Research

Heike Schroeder & Harry B.G. Ganzeboom
Free University Amsterdam

December 13 2010

Hong Kong University of Science and Technology

Current state of affairs (ESS, ISSP)

- Three ways to measure education in comparative surveys:
  - ISCED: International Standard Classification of Education.
  - Duration: years of education (in full-time equivalents) of school-leaving age.
  - CS [country specific] measures.
- In ESS (and ISSP) all three ways are included!

ISCED

0 Incomplete / No Primary

1 Complete Primary

2 Lower Secondary

3 Higher Secondary

4 Post-Secondary

5 Lower Tertiary, BA

6 Higher Tertiary, MA and up.

Problems with ISCED

- Is not a detailed classification, but rather crude: in practice it restricts variations to 3 or 4 (out of 7 possible) levels.

  → Aggregation bias
- This problem varies between countries. In ESS in particular Germany and the Czeck Republic have massive numbers in higher secondary.
- There are many practical problems with implementing ISCED in surveys (Schneider).

Duration

- ESS: About how many years of education have you completed, whether full-time or part-time? *Please report these in full-time equivalents and include compulsory years of schooling.*
- ISSP: How many years (full-time equivalent) have you been in formal education? *Include all primary and secondary schooling, university and other post-secondary education, and full-time vocational training, but do not include repeated years. If you are currently in education, count the number of years you have completed so far.*

COMPARE

- GSS: What is the highest grade in elementary school or high school that you have finished and got credit for? How many years of college did you complete for credit?

Problems with duration

- Easy measurable and conceivable metric with relevance for human resource policies.
- However, in tracked education systems the relationship between level of education and duration is unclear.
- In practical applications, there is much confusion on how to phrase / answer the question.
- Empirical evidence (see below) shows that duration does not work well and leads to rather attenuated coefficients.

Problems with CS measures

- CS measures seem not comparable by definition.
- CS measures have very little usage in comparative analyses.
- If used by comparative researchers, CS measure are rendered comparable by resorting to a crude common denominator approach, even cruder than ISCED.

Our approach

- Step 1 ***('measuring level of education')***: Exploit the full detail that is available in CS measures by developing the International Standard Level of Education [ISLED], a uni-dimensional hierarchy of country-specific education categories used in surveys, with a cross-nationally comparable metric.
- Step 2 ***('modeling level of education')***: Exploit the multiple indicator information that is available in ESS and other surveys using a true score model, that diagnoses and corrects all measurement error.
- Although we present these two steps together, they can be taken independently of one another.

Conclusions, preview

- ISLED allows users to compare level of education across space and time with minimal loss (<5%) of information.
- Alternative approaches may incur up to 10-20% loss of information.
- However, it is possible to create perfect (comparative) measurement by including a second independent indicator (duration) and the use of a common factor or MTMM measurement model.

Level of Education

- Level of education can be defined as the value of educational qualifications in society:
  - In outputs: employment, occupational status, earnings, marital status, status of spouse, children, friends;
  - In inputs: education attainment as the outcome of competition between social groups: educations and occupations of parents.
- An appropriately measured Level of Education is strongly predicted by inputs and predicts outputs strongly. When comparing competing measures, we would like to assess them by how they represent the transfer of inputs into outputs.

Optimal scaling

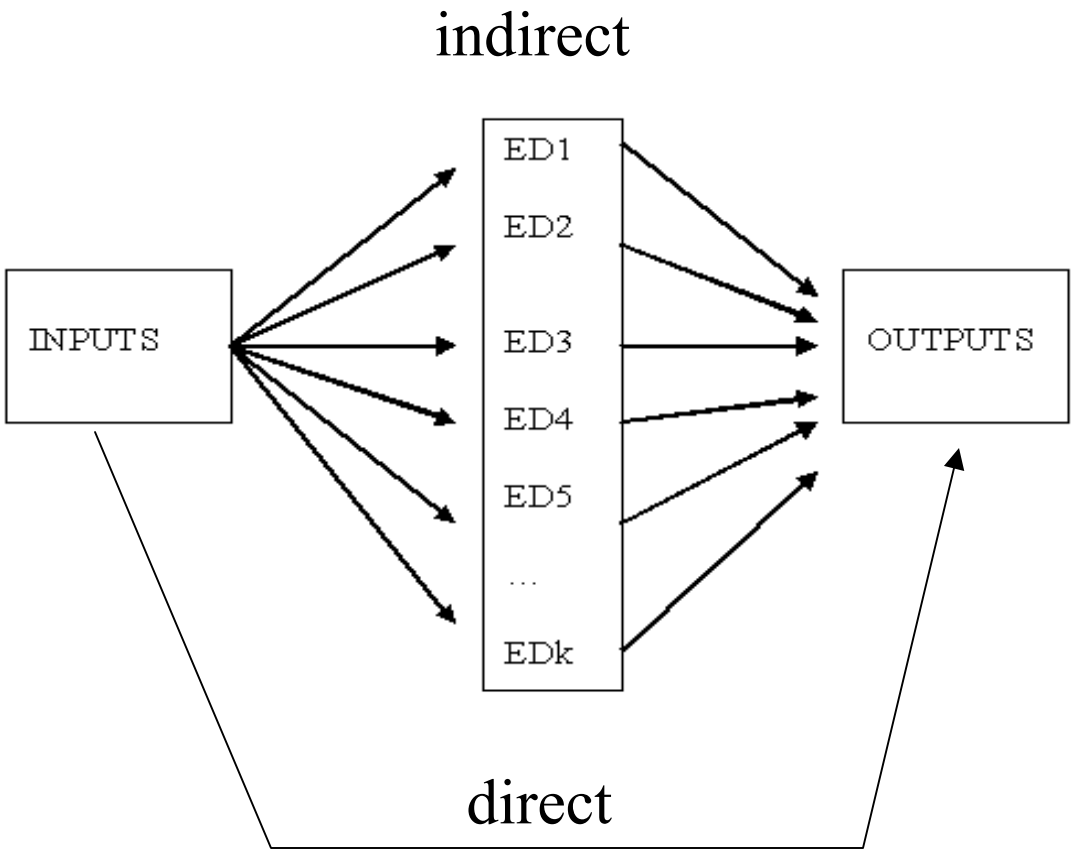
- We take it one step further: scale educational qualifications in such a way that education mediates between social backgrounds into status attainments.
- Minimize direct effects, maximize indirect effects.
- The idea is the same as used to construct ISEI, but now:
  - Country specific measures;
  - Multiple inputs, multiple outputs.

Methodological justification

- In indirect effects models, the proportion of the total effect that goes to the indirect effects, depends upon the quality of measurement of the intermediate variables.
- Of you measure the mediator badly, all goes to the direct effect.
- We turn this argument around: by optimizing the mediator, we remove measurement error.
- Note that there is an optimum: in no way you can manipulate the data so that no direct effect would remain!

Step 1: Measuring level of education by optimal scaling

MIMIC model of LED



Scaling procedure

- We scale educational categories in such a way, that:
  - The indirect effects of inputs on output via education are maximized; and …
  - … the direct effects of inputs on outputs are minimized.
- Procedure is similar to the one used to construct the ISEI (International Socio-Economic Index of occupational status).
- Inputs: father's and mother's education and occupation.
- Outputs: respondent's occupation and spouse's education.
- Scaling procedure can be characterized as “cause- and effect proportional scaling”.

Defining a metric

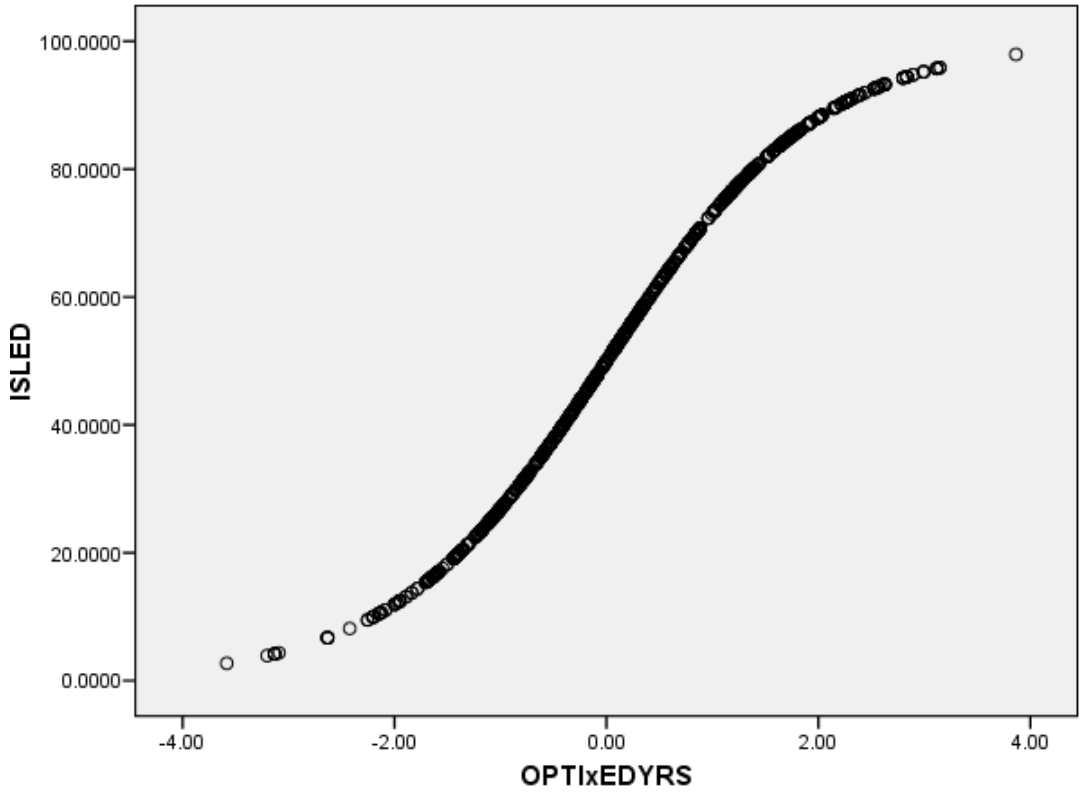
- Optimal scaling leads to ***best*** measurement (by a single indicator), but not to ***comparative*** measurement. We need a comparable metric.
- Options considered and rejected:
  - Define metric in terms of the criterion variables used to develop the scaling. I.e. assume that similar occupations and educations (of fathers, mothers, spouses) have the same value everywhere,
  - Two calibration points, *viz.* primary completion (6) and secondary completion (12).
  - Percentile score in a pan-european distribution (using the criterion variables).

Comparative metric chosen

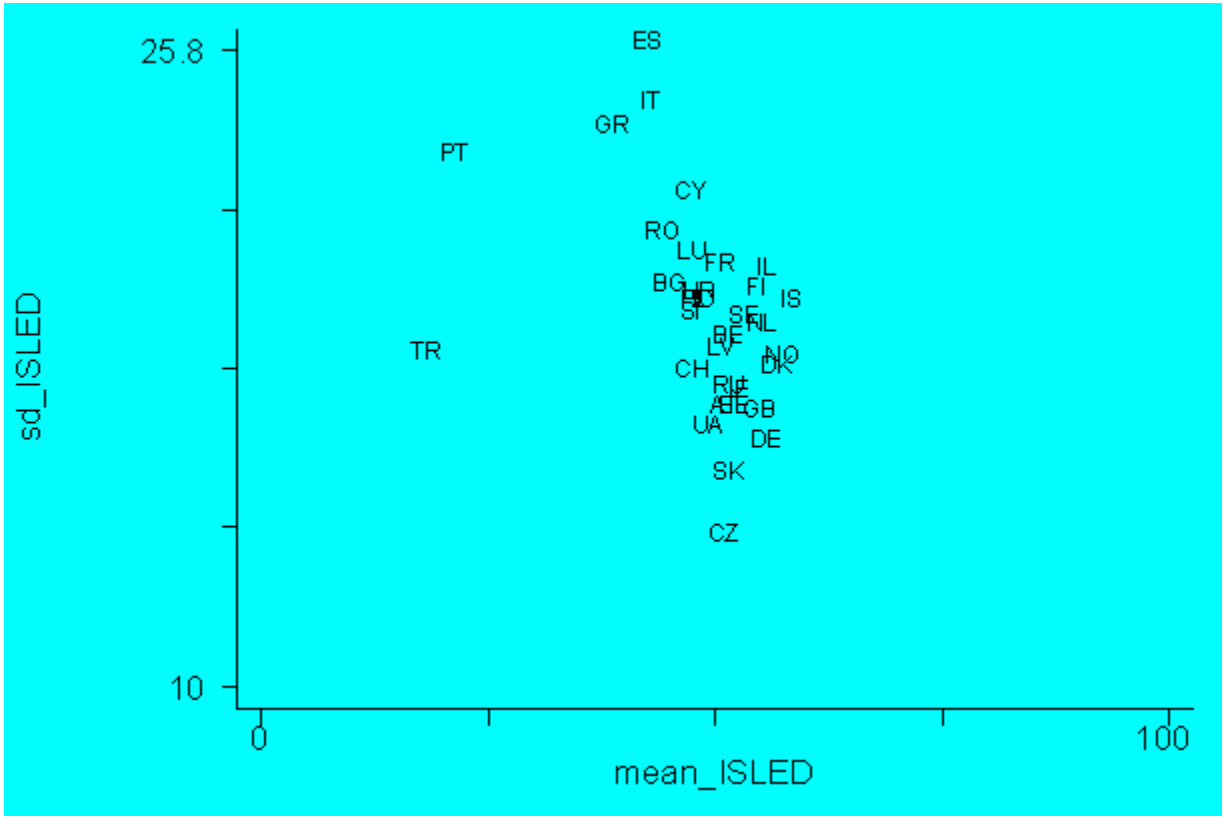
- We define a comparable metric by **equalizing** mean and dispersion between optimal scale and the duration measure (in over-all standardized terms Z).
- Then project back into 0..100 metric using anti-logistic transformation: ISLED=100*(exp(Z)/(1+exp(Z))).

→Means and dispersion of ISLED distribution are proportional to that of duration, but expressed in a 0..100 metric.

ISLED metric



ISLED: Means and SD's



Data: ESS

- We have merged the data from R1234 of ESS, 32 countries, 2002-2008.
- Effective sample: men and women aged 25-74 with valid education measures, N= 150750.
- Criterion variables:
  - **Inputs** are father's and mother's educations and occupations.
  - **Outputs** are respondent's occupation and spouse's education.
- Metrics of the criterion variables: ISCED (0-6) for educations and ISEI (10-90) for occupations.
- Altogether, there are over 1100 country-specific educational categories used, of which over 500 are unique.

Problems with the ESS data

- Some countries have very detailed CS measures: LU, CS, DE and NL, others have very little or no detail: AT, FI, GB.
- Many countries have changed their CS measure between rounds, usually by splitting categories, but sometimes also by adopting entirely new education schemes (EE).
- Criterion variables are also heterogeneous in detail and quality, in particular father's and mother's occupation.

Heterogeneity within countries

- To deal with within-country heterogeneity in CS classification, we have reorganized the ESS codes into a hierarchical digit system, in which splits are represented by a second digit.
- Optimal scores of higher-level categories are obtained by averaging over optimal scores of lower-level splits.

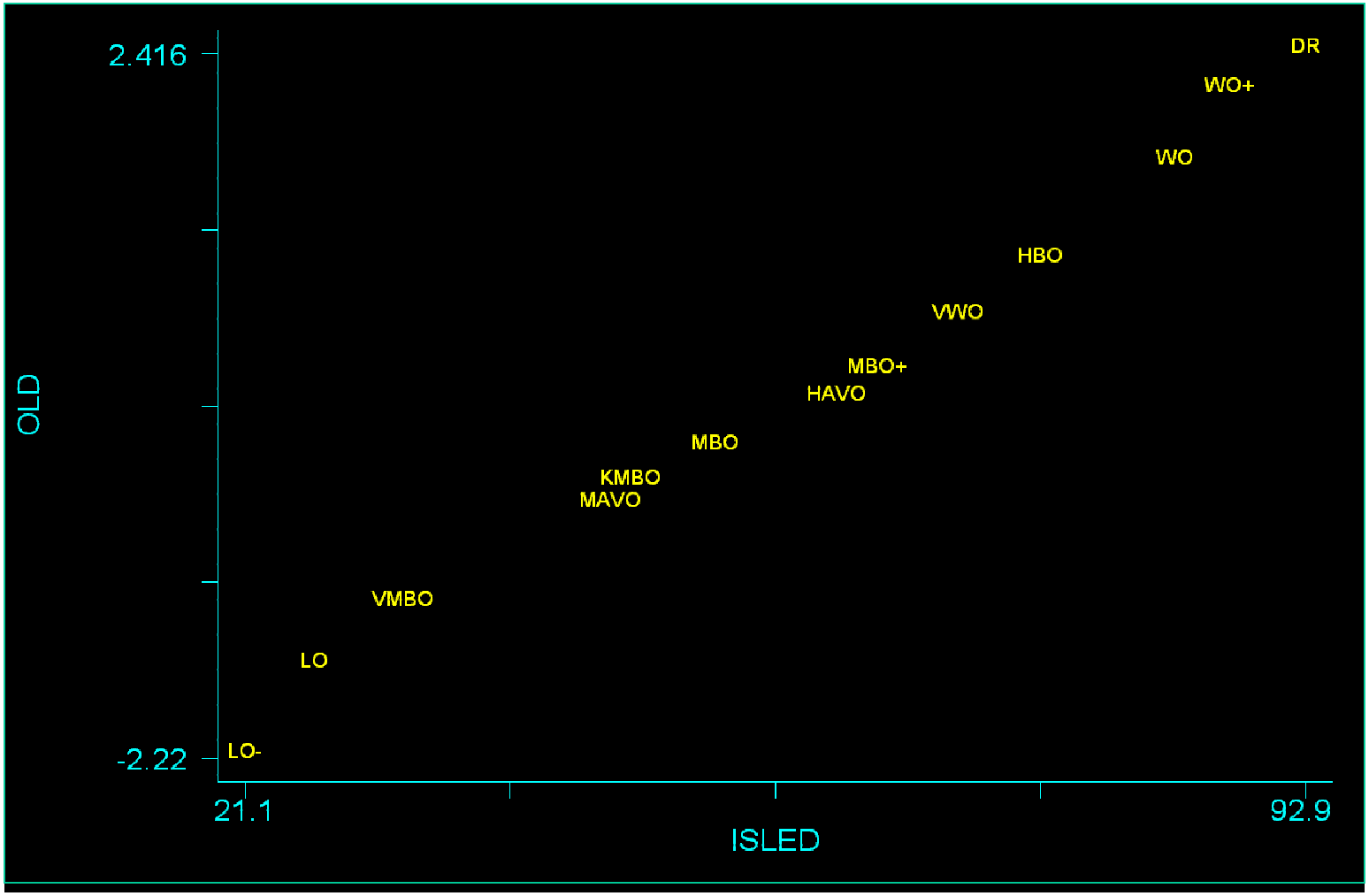
Optimization

- Similar to the procedure used for ISEI scores for detailed occupations (De Leeuw).
- Algorithm:
  - Take unweighted average of standardized inputs and standardized outputs (correlation: 0.85).
  - Find optimal relative weight for inputs (0.61) and outputs (0.39).
  - Optimal weight is found by systematic search. Stop at minimal direct effect of inputs on outputs.
  - No age adjustment used.
- Note that the weights are the same in all countries, but the inputs and outputs (and correlation structure) are different.
- Re-standardize the result within countries and then re-express using country specific means and standard deviations of the duration measure.

ISLED (Netherlands)

EDUCXX	**OPTI**	**ISLED**
LO-	-1.81	21.1
LO	-1.53	25.8
VMBO	-1.19	31.8
MAVO	-0.48	45.9
KMBO	-0.42	47.2
MBO	-0.14	53.0
MBO+	**0.38**	**63.9**
HAVO	**0.25**	**61.2**
VWO	0.67	69.4
HBO	1.01	75.0
WO	1.67	84.1
WO+	2.03	87.8

ISLED-scaling (NL)



Results Step 1

- Appendix shows all optimal scale scores (within country standardized) and ISLED by country and category.
- See: www.harryganzeboom.nl/isled.htm
- This is ready to use:
  - For analysis using ESS data
  - For other European data after assigning ISLED to CS education codes.

Step 2: Modeling level of education using a multiple indicator model

Modeling measurement error

- Social structure indicators (like education) are prone to measurement error, just as any other variable.
- Classical measurement theory provides us with tools to diagnose and correct measurement error ***(disattenuation)***:
  - Multiple indicators to cover somewhat different aspects of underlying concepts.
  - SEM (single common factor) model will diagnose and correct **random** measurement error.
  - MTMM (correlated residuals, multiple common factor) model will also diagnose and correct **systematic** measurement error.

Multiple indicator measurement model

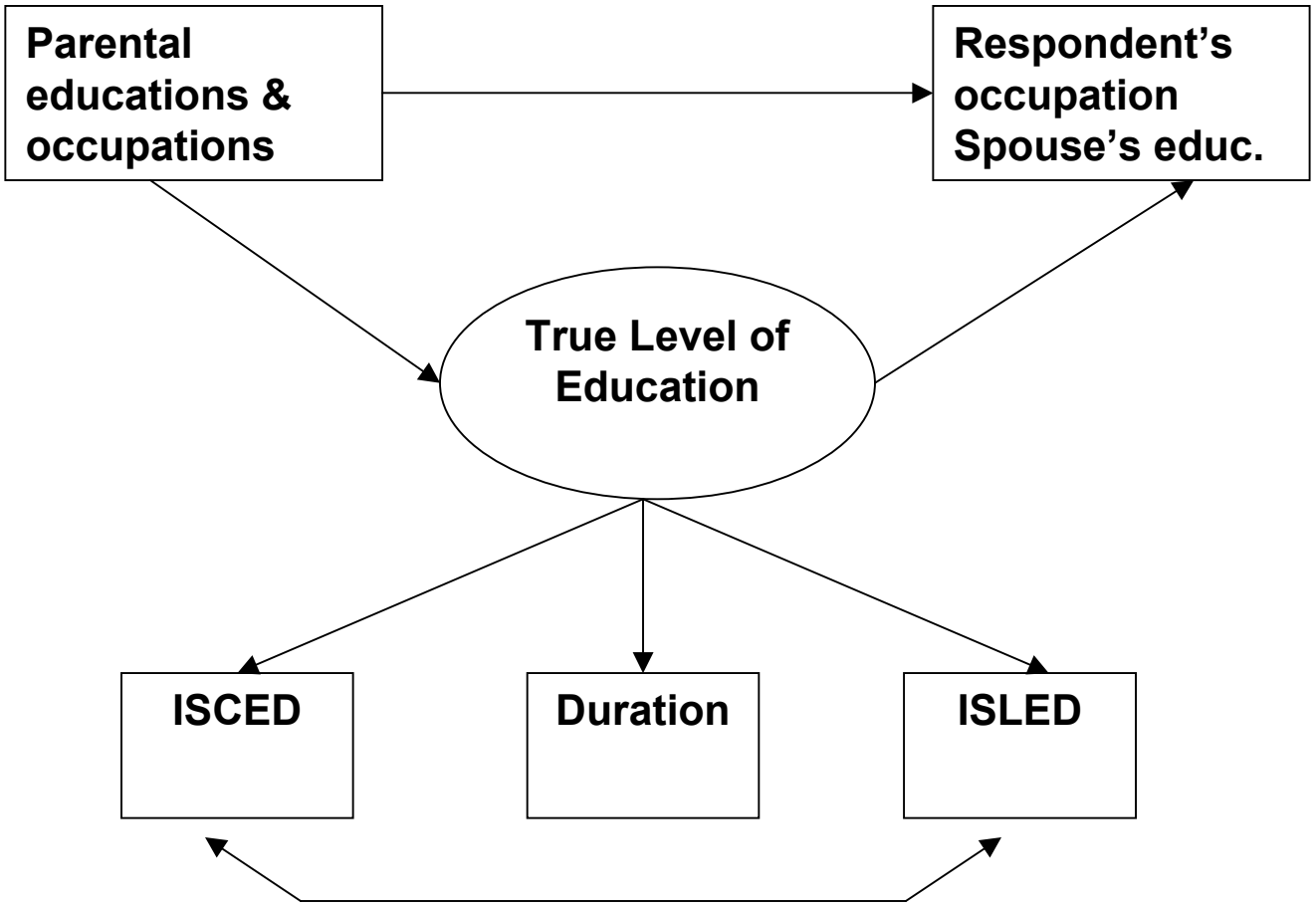


Table 2: Model parameters for ALL COUNTRIES ESS Round 1-2-3-4						
N=127,681						
ALL	**1**	**2**	**3**	**4**	**5**	**6**
	EDDUR	**ISCED**	**ISLED**	**1 & 2**	**2 & 3**	**1, 2 & 3**
A. Structural models						
EDUCATION R.						
FEDUC	0.169	0.184	0.188	0.201	0.200	0.200
MEDUC	0.139	0.138	0.143	0.152	0.151	0.151
FOCC	0.144	0.140	0.156	0.160	0.167	0.164
MOCC	0.103	0.078	0.121	0.043	0.127	0.125
R2	**0.201**	**0.210**	**0.239**	**0.259**	**0.269**	**0.266**
SPOUSE'S EDU						
FMEDUC	0.165	0.152	0.143	0.127	0.127	0.127
EDUC	0.351	0.394	0.407	0.452	0.447	0.448
R2	**0.301**	**0.324**	**0.329**	**0.356**	**0.352**	**0.353**
OCCUPATION R.						
FMOCC	0.094	0.078	0.056	0.043	0.033	0.036
EDUC	0.491	0.545	0.589	0.623	0.639	0.635
R2	**0.330**	**0.374**	**0.406**	**0.434**	**0.446**	**0.443**
B. Measurement models						
ISCED	1			0.913		0.915
Duration		1		0.856	0.845	0.848
ISLED			1		0.947	0.945
C. Fit statistics						
Chi-square	2326.5	1220.0	821.3	603.4	596.1	1401.0
RMSEA	0.051	0.037	0.030	0.020	0.019	0.025
Standardized parameters. # = non-significant.						

Results Step 2a

- ISLED outperforms ISCED on average by 4%
- The improvement is more dramatic for countries with detailed CS measures and hardly noticeable for countries with no or uninformative CS measures.
- ISLED outperforms Duration by 15%.
- Any combination of two independent measures (ISCED-EDDUR or ISLED-EDDUR) reaches basically the same dis-attenuated structural coefficients.

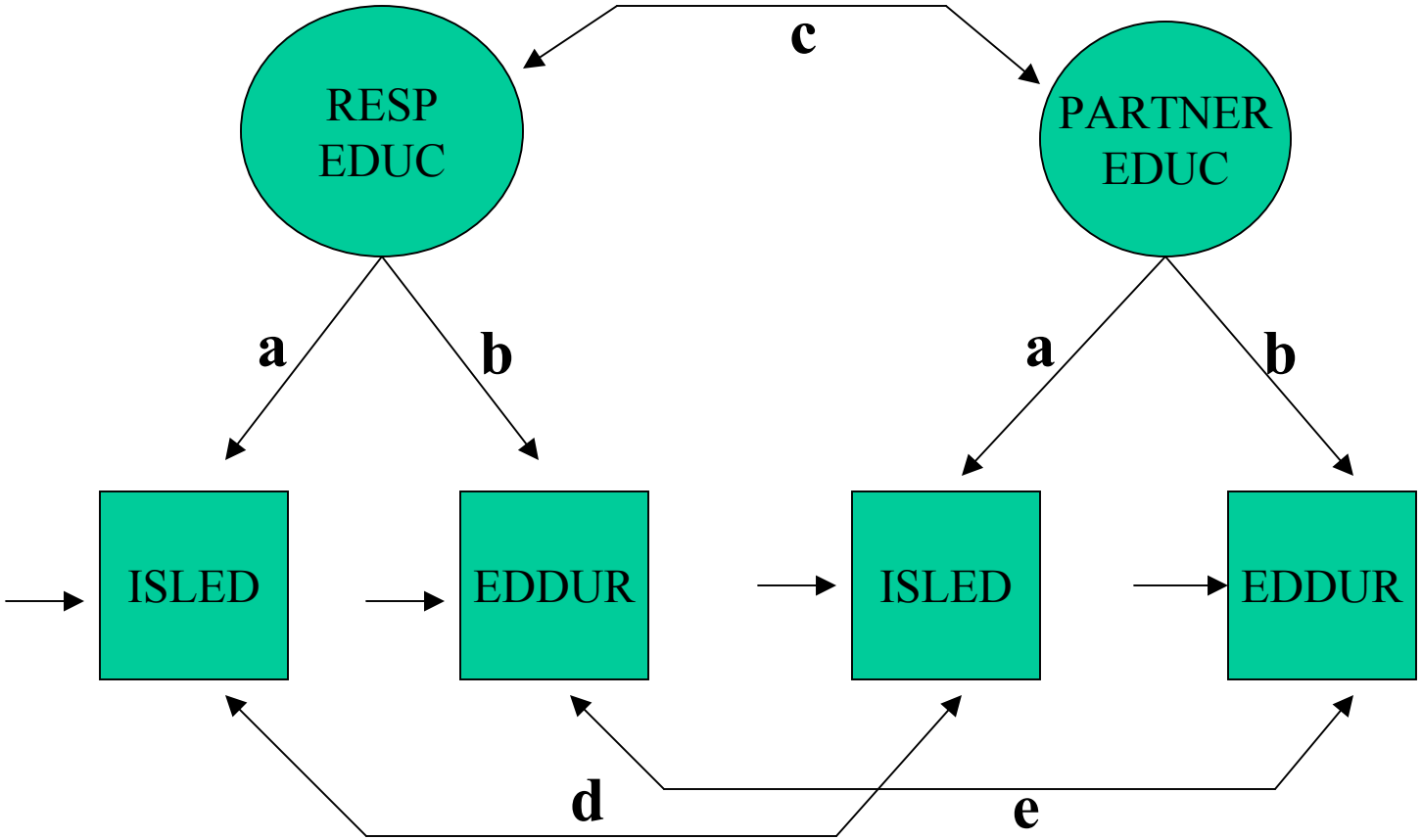
Step 2b: Fresh data, systematic error

- Validation is more appropriately done on fresh data: ISSP-NL data 2002-2008 (six surveys).
- With this data we can also address systematic measurement error (=invalidity), i.e. how indicators systematically represent underlying constructs.
- An important kind of systematic error is correlated error, i.e. errors that arise in every instance of measuring education: MTMM model.

MTMM

- MTMM: Multiple Trait, Multiple Methods.
- Proposed in 1950s by psychometricians, now mostly applied in attitude research to model response-tendencies. SEM with confirmatory multiple factor analysis.
- How do you diagnose random error? By repeating the measurement → multiple indicators.
- How do you diagnose systematic error? By also repeating the error → multiple traits.
- MTMM combines these two.
- We can make MTMM models work for us by repeating the same measurement model on educations of different persons.

Basic MTMM model



Identification of the MTMM model

- The basic MTMM model is actually not identified for case of the 2*2 educations * measures (despite 5 effects with 6 correlations).
- However, the effects become identified in an elaborated model with criterion variables.

Implementing the ESS-NL values in ISSP-NL

ESS		ISSP			ISLED
1	Niet voltooid lager onderwijs			1	21.1
2	Lager (speciaal) onderwijs	1	Lager onderwijs	2	25.8
3	LBO, HHS, LTS, LHNO, VMBO-b, VMBO-k	2	LBO, HHS, LTS, VMBO-b, LHNO, VBO	3	31.8
4	MAVO, ULO, MULO, VMBO-t	3	MAVO, ULO, MULO, VMBO-t	4	45.9
5	KMBO	6	KMBO , leerlingwezen , **BBL-BOL**	5	47.2
6	MBO, **BBL, BOL**	7	MBO	6	53.0
7	MBO plus, K-HBO			**8**	**63.9**
8	HAVO, MMS, VHBO	4	HAVO, MMS	**7**	**61.2**
9	VWO, HBS	5	VWO	9	69.4
10	HBO, Kweekschool, MO, Conserv.	8	HBO	10	75.0
11	WO, TH, EH	9	WO	11	84.1

Measurement models (1)

Measurement models (ESS-NL)			
Model	**1**	**2**	**3**
	EDDUR & ISCED	**EDDUR & ISLED**	**EDDUR, ISCED & ISLED**
Duration	0.788	0.780	0.782
ISCED	0.899		0.899
ISLED		0.929	0.929

Measuremen models (ISSP-NL)			
Model	**1**	**2**	**3**
	EDDUR & ISCED	**EDDUR & ISLED**	**EDDUR, ISCED & ISLED**
Duration	0.790	0.797	0.797
ISCED	0.829		0.826
ISLED		0.942	0.942

Measurement models (2)

	Father's and Mother's education in ISCED			Father's and Mother's Education in ISLED		
Model	**1**	**2**	**3**	**1**	**2**	**3**
	EDDUR & ISCED	**EDDUR & ISLED**	**EDDUR, ISCED & ISLED**	**EDDUR & ISCED**	**EDDUR & ISLED**	**EDDUR, ISCED & ISLED**
UNCORRECTED for systematic measurement error						
Duration	0.828	0.818	**0.819**	0.830	0.816	0.817
ISCED	0.799		**0.801**	0.798		0.801
ISLED		0.921	**0.921**		0.923	0.923
CORRECTED for systematic measurement error (0.01; 0.15)						
Duration	0.816	0.803	0.803	0.818	0.801	**0.801**
ISCED	0.811		0.814	0.809		**0.813**
ISLED		0.938	0.938		0.940	**0.940**

Structural models (1)

ESS-NL	Single Indicators			Multiple Indicators		
R2	1	2	3	4	5	6
	EDDUR	ISCED	ISLED	EDDUR & ISCED	EDDUR & ISLED	EDDUR, ISCED &ISLED
Education	0.210	0.211	0.239	0.278	0.286	0.283
Occupation	0.239	0.326	0.351	0.383	0.393	0.392

ISSP-NL	Single Indicators			Multiple Indicators		
R2	1	2	3	4	5	6
	EDDUR	ISCED	ISLED	EDDUR & ISCED	EDDUR & ISLED	EDDUR, ISCED &ISLED
Education	0.160	0.139	0.221	0.239	0.253	0.251
Occupation	0.241	0.256	0.295	0.342	0.361	0.360

Structural models (2)

ISSP-NL	Single indicators		Multiple indicators
	ISCED	**ISLED**	**EDDUR & ISLED**
Education Respondent			
FMEDUC	0.143	0.155	0.176
FOCC	0.127	0.165	0.166
R2	**0.145**	**0.214**	**0.250**
Occupation Respondent			
FOCC	0.193	0.132	0.078
EDUC	0.547	0.686	0.759
R2	**0.441**	**0.558**	**0.621**
FIML Chi-square	87.0	94.8	116.6
RMSEA	0.021	0.023	0.021

Results step 2b

- ISLED (NL) works just as well in fresh data:
  - 6% loss for ISLED relative to true score
  - 19% loss for Duration
  - 17% loss for ISCED
- Duration suffers from considerable systematic measurement error, ISLED not.

Conclusions, final

- ISLED is ready for use.
- At this point: only for (33) European countries.
- ISLED beats common denominator harmonization and duration by some margin.
- However, nothing beats an MTMM true score model.

Remaining problems

- Same data used for construction and validation.
- Little evidence on the usefulness of the common metric.
- Shoot!