Harry B.G. Ganzeboom

The Comparative Measurement of Stratification Indicators

Mini-Course, University of Tallinn
November 17-19 2010

Lecture Day 2

The comparative measurement of occupational status

5. Occupation coding

Harmonizing ages between surveys

- Suppose we would have surveys with two different measures of age, one in categories, and the other one continuous (or different categories).
- The most obvious way to harmonize these two measures is by ***scoring*** all information to plausible points.
- We should (and would) avoid harmonizing by choosing a common denominator.

Harmonizing political parties

- Suppose we would have two different measures of party choice in elections (e.g. two different countries).
- We would (and should) never harmonize this information by scoring it on an underlying dimension. Nor would be harmonize by creating a common classification.
- Rather, we would leave the information as it is and leave it to the analyst to process it.

Harmonizing education and occupation

- Occupations are like age:
  - Use a common (detailed) classification
  - Occupational hierarchies are the same the world over, so “one size fits all”.
- Educations are like political parties:
  - Leave the information as it is, country specific.
  - Leave it to the analyst to process this information.
- The difference arises because occupations are generated by a universal process (the division of labor), while educations are institutionally organized.

Occupational classifications

- The backbone of occupational measurement is to code occupations using a standard classification.
- Country-specific standard classifications exist – they are often produced for census purposes.
- Since 1958: the International Standard Classification of Occupations [ISCO]: 1958, 1968, 1988 and now 2008.
- Although there are profound differences in how classifications are organized, it makes remarkable little difference which one you work with.

ISCO

- ISCO is produced by the ILO (International Labour Organisation). See their website.
- ISCO-58 had little application in social surveys.
- ISCO-68 became a tool of classification in some social surveys. One source of its popularity is Treiman (1977). A by-product of the SIOPS scale.
- ISCO-88 is now almost universally applied in comparative social surveys (such as ESS, PISA, ISSP, EVS, SHARE).
- ISCO-08 is coming up!

Outline of ISCO-88

- ISCO has about 580 groups to code occupations.
- These groups are organized in four levels of aggregation by a hierarchical digit system.
- ISCO-88 has a 600 page manual (on the ILO website).
- The Introduction to the ISCO-88 manual (not on the ILO website) is a useful piece to read.
- ISCO-88 is skill-oriented, but it does not really show.

Major groups

- 1000 Managers
- 2000 Professionals
- 3000 Associate Professionals
- 4000 Clerical Workers
- 5000 Sales and Service Workers
- 6000 Skilled Agricultural Workers
- 7000 Craft Workers
- 8000 Machine Operators
- 9000 Elementary Occupation
- 0100 Military, all ranks

The hierarchical digit system

- 1000 Legislators, Senior Officials and Managers
  - 1100 Legislators and Senior Officials
    - 1110 Legislators
    - 1120 Senior Government Officials
    - 1140 Senior Official, Special Interest Orgs.
      - 1141 Political Party Officials
      - 1142 Economic Interest Org Officials

The devil is not in the detail

- Major groups
  - Sub-major groups
    - Minor groups
      - Unit groups
- Major groups are far more important than more detailed groups.
- Coders should really have the major and sub-major groups in their heads.
- However, adding the last two digits is not so much work, when you do it right.

Major problem in ISCO-88

- Where to code farmers?
- How to code managers and supervisors?
- What is the difference between 7000 and 8000?
- How to code crude occupations like “skilled worker”?

Do's and Don'ts of occupation coding

- I have produce a manual of Do's and Don'ts in occupation coding (see course readings).
- Some important rules:
  - Use multiple coders.
  - Let them work independently on slightly overlapping random parcels.
  - Code in two round: (A) first two digits (B) second two digits.
  - Use an MTMM model to estimate random and systematic coding error.

Estimating coder quality

- Coding is just another course of error in data. It is reasonable to assume that most coding error is random error.
- How can we know the quality of coding? Standard answer: double coding.
- Double coding has two major problems:
  - It is twice as expensive as single coding.
  - What do we do if two codes disagree?

Estimating coder quality without double coding

- If you use multiple coders (DO!), you can use the elementary MTMM model to compare coder quality.
- Strictly speaking we do not even have to do overlapping coding, but is helps if we have some overlap.
- An MTMM model will then lead us to a estimate of random and systematic coder error.

The excercise

- In the ESS, the occupations of Estonian fathers and mother are not coded.
- Kadri has done part of it, but some 1120 occupations remain uncoded. I have divided this up in (overlapping) parcels of 175 occupations.
- Each of you does one parcel.
- You are not allowed to communicate, except with people from your own university (Tallinn versus Tartu).
- Have it ready by midnight.

6. Construction of ISEI

SEI

- Duncan (1961) developed the most famous of all SEI indexes.
- He was faced with the problem that prestige surveys had provided scores for some occupations, but not all.
- He then developed a score for all occupation using mean education and mean income as predictor variables for their prestige.
- SEI then is the expected prestige for an occupation, given the mean education and mean earnings of its incumbents.
- Note that Duncan also took SEI as the status measure for occupations for which the prestige was known!

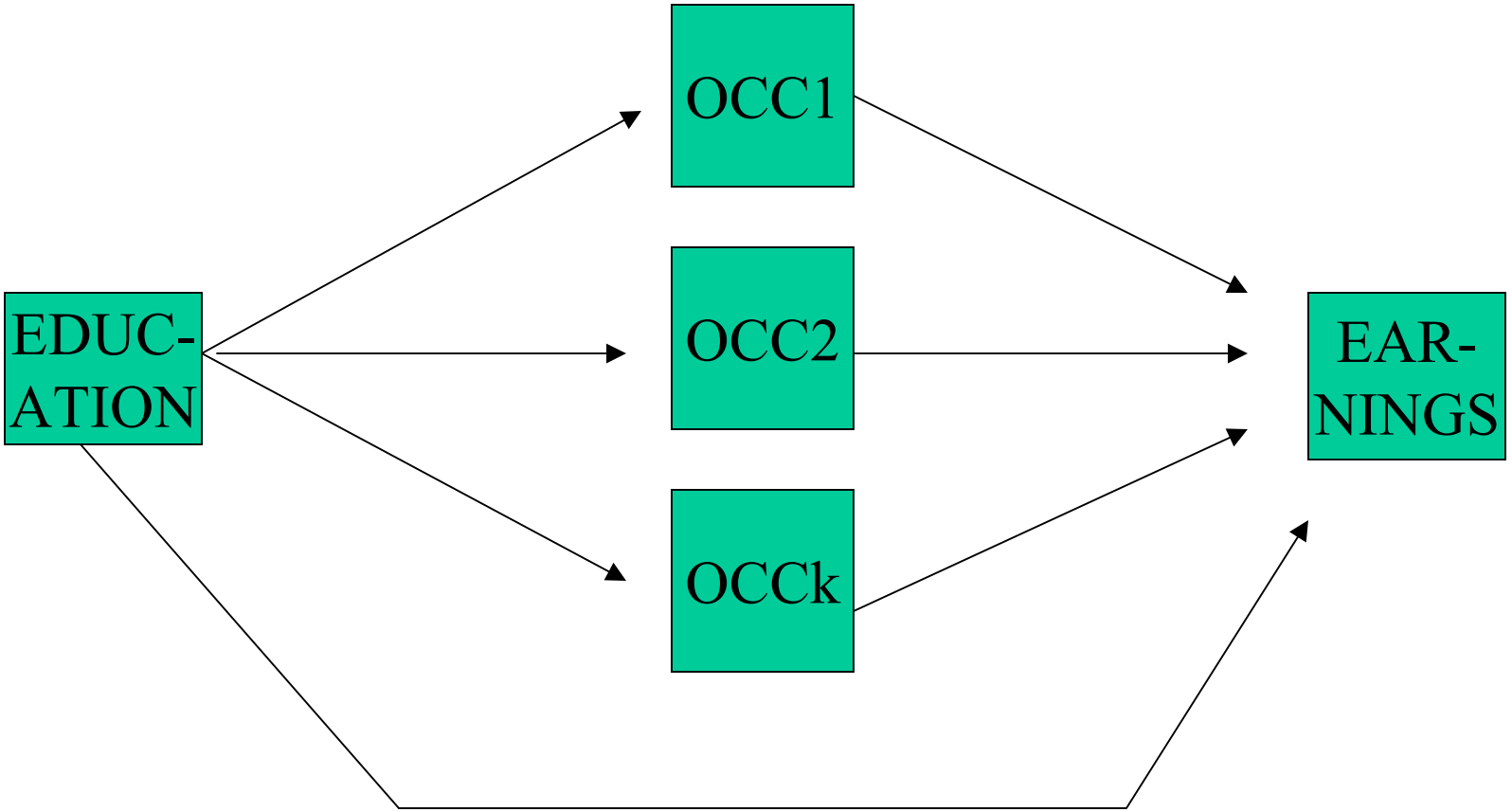
SEI and prestige

- Because of its construction by Duncan (1961), prestige and SEI are often mixed up.
- They are correlated, but not identical.
- When compared, SEI turns out the better measure of occupational status.
- Featherman & Hauser (1976): SEI is not a approximation of prestige, its is rather the other way around.
- Most of the differences, but not all, occur in the position of farmers.

ISEI

- ISEI is the International Socio-Economic Index of occupational status, created for ISCO-68 and ISCO-88 (ISCO-08 coming soon).
- The construction does not use prestige as a criterion variable.
- In stead, scores are derived from an indirect effects model, in which occupation mediates the effect of education on earnings. The weights used maximize the indirect and minimize the direct effect.
- Justification: this minimizes measurement error in occupational scaling.

The indirect effects model



Some details about ISEI construction

- No prestige used!
- Parental and spouse occupation are not used – so excellent for validation.
- Most of the work is in organizing the detailed groups. Criteria:
  - N > 21 cases
  - Represented in at least 2 countries.
- Data: for ISCO-68 and ISCO-88 around 70.000 men with complete data from some 17 countries. For ISCO-08: around 200.000 men and women from 30 countries (ISSP).

Some more details

- The actual optimization algorithm was invented by Jan de Leeuw (Leiden/UCLA) and is explained in GGT (1992). It requires iterations, but can be done in SPSS.
- Input (education) and output (earnings) are harmonized within countries. For education: Z-scores. For earnings: Z-scores of logged income.
- The final metric is between 10 and 90. Unclear how the details were done.
- Although ISEI-68 and ISEI-88 are extremely similar (> 0.95), they still have unique systematic information. Averaging lead to better models than the two indicators on their own. This means that there are errors in the two scales!

7. Harmonizing and scaling education

Harmonizing education

- Don’t. Like with political parties, you become unhappy by harmonization.
- Even translation has it problems. My recommendation: give program titles both in the original and English language.
- Common denominator harmonization is often used in data and in data analysis. This leads to extremely simple representations of educational stratification (e.g. elementary, secondary, higher).

ISCED

- ISCED: International Standard Classification of Education.
- Produced and maintained by OECD: 1978 and 1997 versions. Revision coming up.
- Although ISCED looks like ISCO, it is not the same:
  - Limited number of countries (OECD).
  - Brings together national educational structures, as they were at the time of construction (1997).
  - ISCED's digit system is not truly hierarchical, in the sense that leading digits aggregate the more refined detail very well. The first digit is duration oriented, the second digit is level oriented.

ISCED (first digit)

- No or incomplete primary
- Primary
- Lower secondary
- Higher secondary
- Post-secondary, non tertiary
- Lower tertiary (BA and MA)
- Higher tertiary (PhD)

ISCED (other digits)

- A,G Academic, general orientation / destination
- B,V,P Vocational orientation / destination
- C Other

- There are also distinctions by duration of the program and position in the national qualification system.
- All of this for programs in OECD countries in 1997.
- The classification is in such small print that it is unreadable.

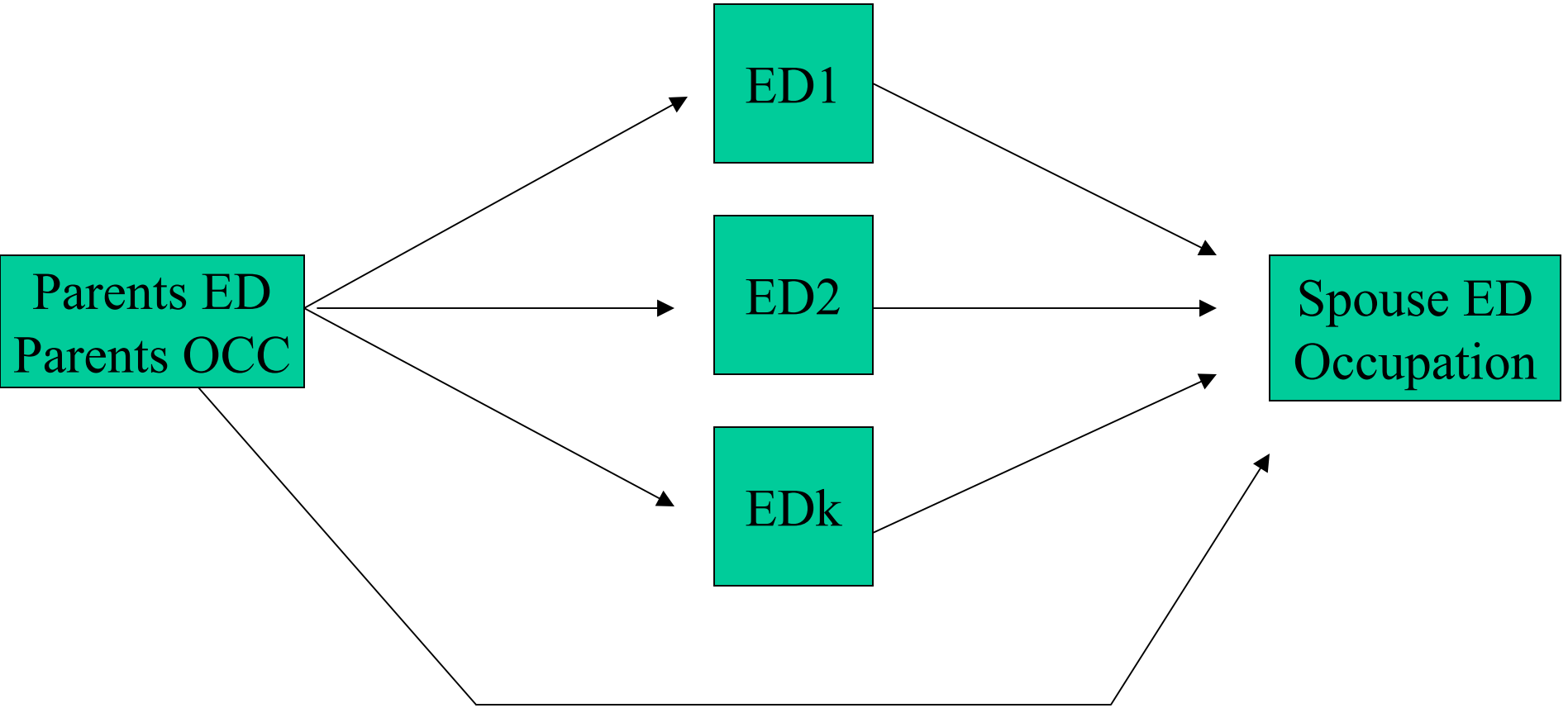
ISCED disasters

- Disasters happen when projects decide to code/use only the first digit of ISCED (ESS, PISA, IALS and more).
  - In many countries only 2-3 categories are distinguished.
  - Some of those categories can be as large as 70%.
  - It remains unclear how the coding or classification was done.
- ISSP does a better job; ESS and EVS have moved to country-specific more detailed measurement.

ISLED

- ISLED == International Standard Level of Education. Schroeder & Ganzeboom, 2010.
- Level of education: the value of education. This can be found in two ways:
  - By looking at returns to (effects of) education in the labor market (and marriage market): effect proportional,
  - By looking at access to different programs for different social background groups (father's and mother's education and occupation): cause proportional.
- Together these make for an indirect effects model similar as was used for ISEI.

The indirect effect (MIMIC) model



Differences with ISEI

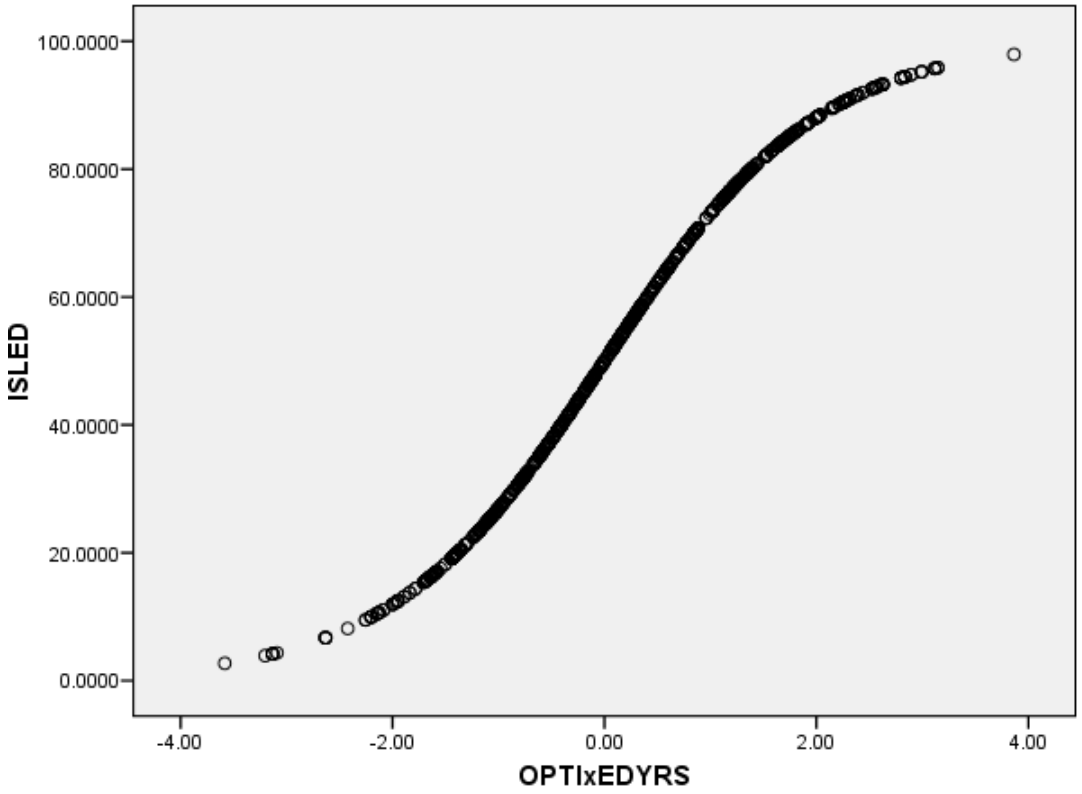
- We now have multiple inputs (4) and multiple outputs (2).
- Standardization of criteria over all countries
- Unit groups are nationally specific, even different between studies (ESS rounds).
- Minimum N > 21.
- Algorithm is the same.
- Data: ESS Round 1-4.

Defining a metric

- Optimal scaling leads to best measurement (by a single indicator), but not to comparative measurement. We need a comparable metric.
- 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 into 0..100 metric using anti-logistic transformation: ISLED=exp(Z)/(1+exp(Z)).

→Means and dispersion of ISLED are proportional to duration.

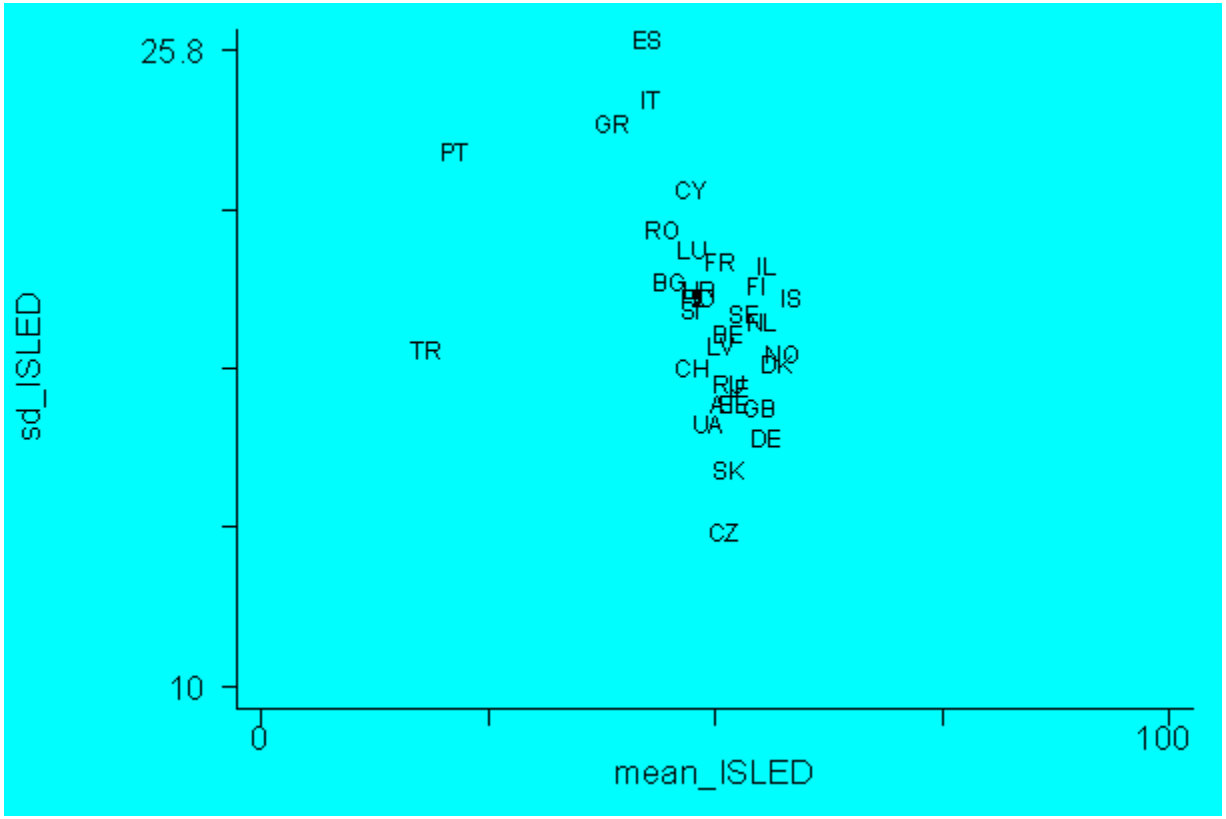
ISLED metric



Measuring and Modeling Level of occupational status

Measuring and Modeling Level of Education status

ISLED: means and sd's



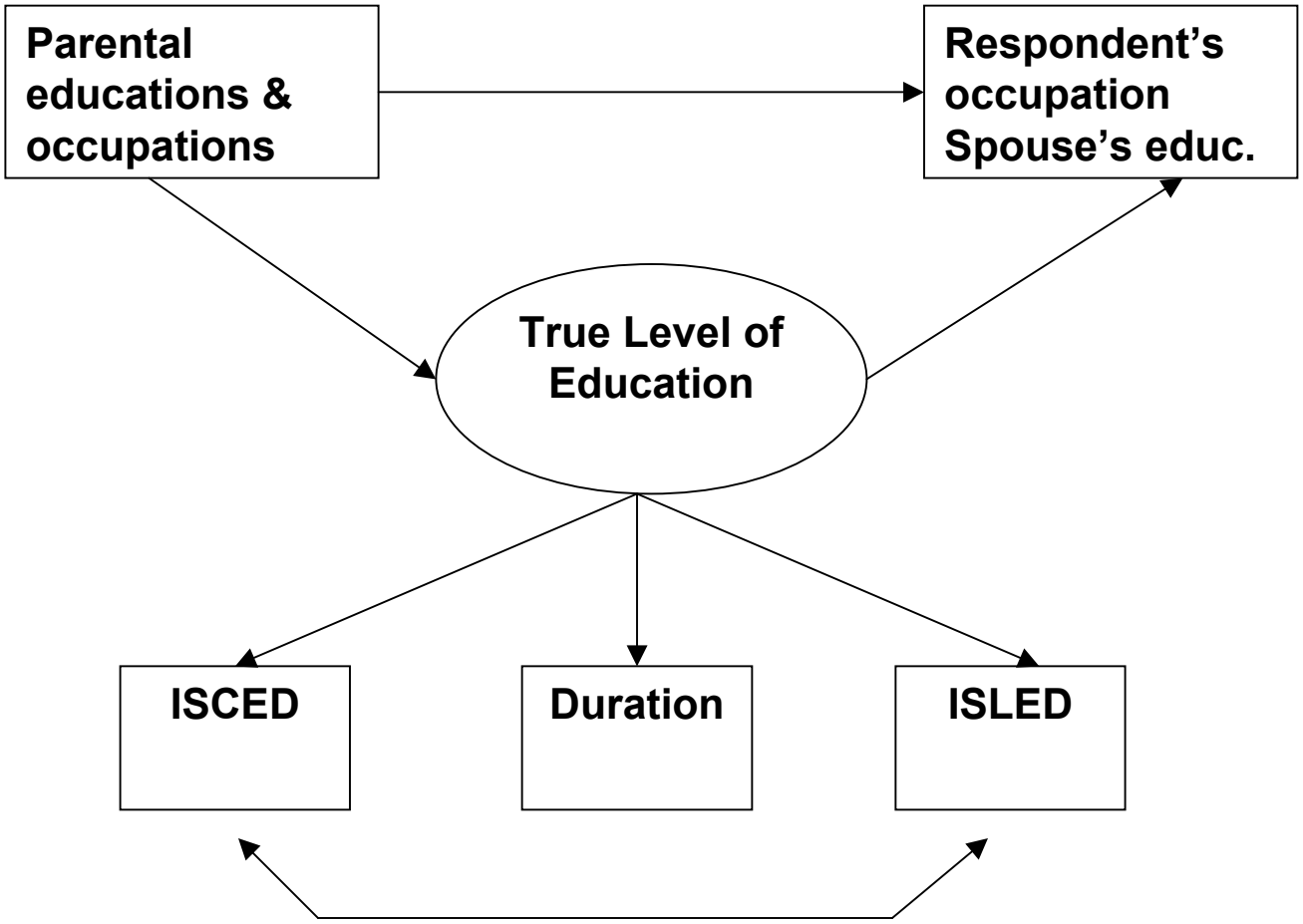
Summary of country-specific education categories and ISCED levels LUXEMBOURG, ESS Round 1-2								
Cat	**Country specific education category**	**N**	**Cat**	**ISCED**	**N**	**Opti**	**ISLED A**	**B**
0	No qualification	43	0	Not completed primary education	43	-1.35	5.6	8.3
1	Primary school	331	1	Primary 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

8. Double indicator validation

Double indicator validation

- ISEI and ISLED seem good ideas, but the result may not be perfect. They contain errors, if only because the underlying occupation and education codes contain errors.
- How do we know: using double indicator measurement and an MTMM model.
- Double indicator measurement can be generated:
  - For occupations using a detailed (open) and crude (closed) question.
  - For education using qualification and duration as a double indicator.
- Double indicator models will give us an estimate of random error. If we repeat the measurement over multiple construct, we can also estimate and correct systematic error using MTMM.

Multiple indicator measurement model



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.						

Validation in ISSP-NL (6 rounds)

- In ISSP-NL we collect education of respondent and spouse with double indicators: highest qualification and duration.
- Unlike ESS, we can now estimate a full MTMM model.

ISLED Netherlands

	Oud (3 rondes	Nieuw (4 rondes	ISLED
LO-	-2.221	-1.8075	21.1
LO	-1.627	-1.5265	25.8
VMB	-1.228	-1.1861	31.8
MAV	-0.572	-0.4779	45.9
KMB	-0.422	-0.4244	47.2
MBO	-0.196	-0.1411	53.0
MBO	**0.312**	**0.3828**	**63.9**
HAV	**0.132**	**0.2459**	**61.2**
VWO	0.668	0.6749	69.4
HBO	1.040	1.0105	75.0
WO	1.681	1.6693	84.1
WO+	2.160	2.0270	87.8
DR	2.416	2.7567	92.9

occupational status

ISLED-NL applied in ISP

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,	8	HBO	10	75.0
11	WO, TH, EH	9	WO	11	84.1
12	Post-doctorale opleidingen			12	87.8
13	Promotieopleidingen			13	92.9

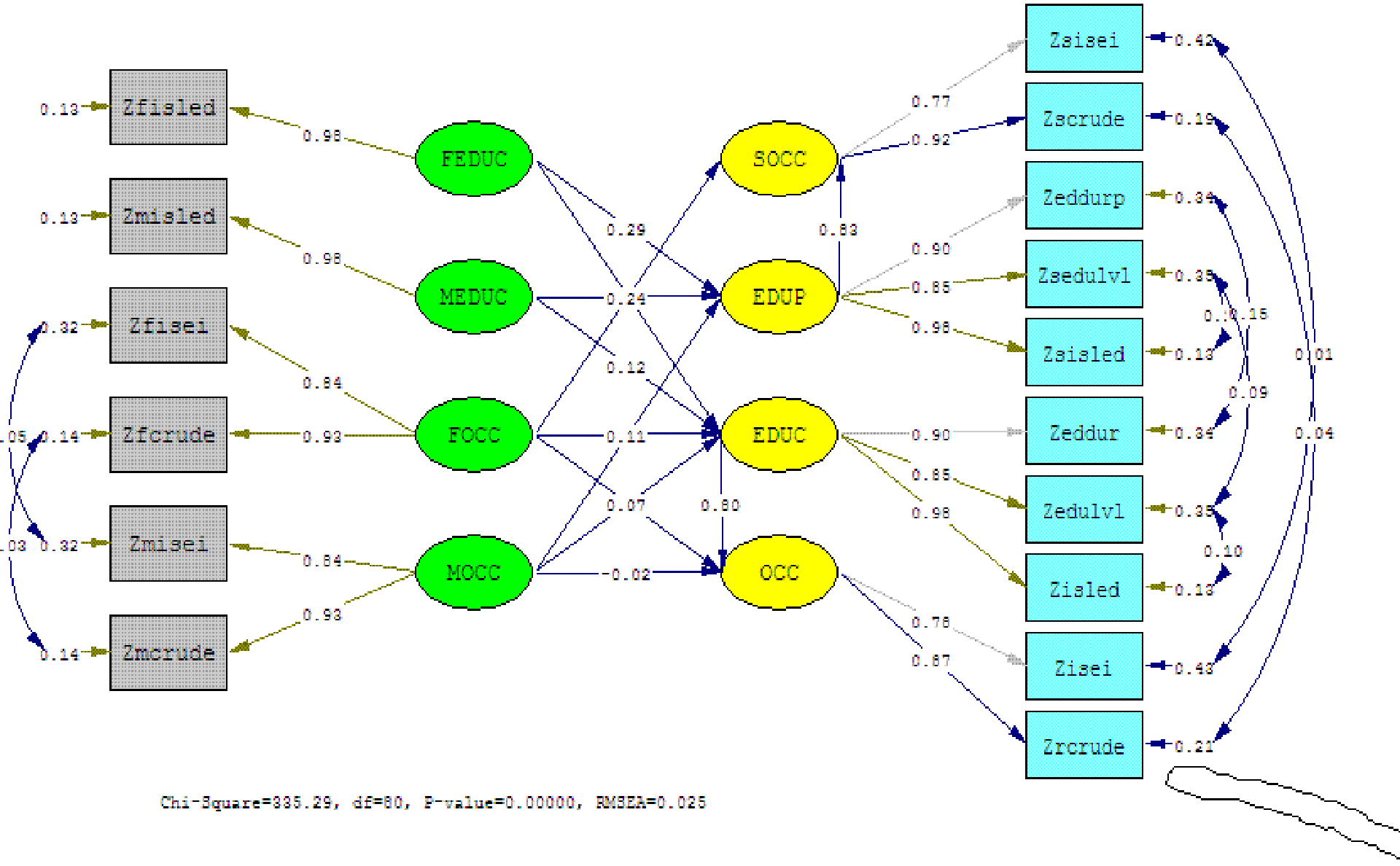
Validation using MTMM ISSP-NL

Correlations for MTMM validation						
	PEDDUR	PISCED	PISLED	EDDUR	ISCED	ISLED
PEDDUR	1					
PISCED	0.672	1				
PISLED	0.758	0.866	1			
EDDUR	0.575	0.464	0.503	1		
ISCED	0.417	0.538	0.496	0.652	1	
ISLED	0.497	0.534	0.592	0.751	0.854	1

Measurement models

Meetmodellen oud (ESS)			
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

Meetmodellen nieuw (ISSP)			
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



occupational status