Harmonizing education for comparative research With special reference to ESS

Harry B.G. Ganzeboom
Department of Social Research Methodology
Free University Amsterdam

Mannheim, ESS-QEM, October 15-16 2009

Harmonizing age ...

- Define a common standard: years fully expired since year of birth in AD counting.
- Convert all information (detailed ages, crude ages, birth years) to this common standard.
- Problems (examples):
  - Sometime unclear whether information is age or year of birth.
  - Open categories in crude ages.
  - Korean ages start at 1.
  - Taiwanese calender starts at 1911 = 00.

→ Despite all these problem we would still harmonize to years of age starting at 0.

Harmonization political parties

- Don't
- Don't translate the names of political parties.
- Don't group political parties into major blocks.
- Do: Provide information about the nature of party programs.
- Do: Present parties in order of left/right dimension.

→ We should never harmonize this information.

Occupation and education

- Occupations are like age
- Educations are like political parties.

Differences with occupations

- Educations are planned institutions – they vary between societies and historically within societies by human design. The details can be quite complex, even within a certain society.
- Fortunately, the information provided in surveyas is often quite limited in detail.
- Fortunately, the information has a strong ordinal character along a single underlying hierarchy.

Common denominator harmonization

- The most elementary system of harmonization:
- It compresses (distorts) the available information. You induce (aggregation) error.
- The more education systems you have to harmonize:
  - The cruder your classification will be (e.g. tertiary or not).
  - The more error you will induce.

→ So this is not a good idea!

Measurement error

- Two kind of measurement error:
  - Systematic bias (invalidity)
  - Random error (unreliability)
- Measurement error arises in education measures just as in other measures.
- Random measurement error can be estimated and corrected by repeating the true score (parallel indicators).
- Systematic error can be estimated and corrected by repeating the error (parallel indicators on correlated constructs = MTMM designs).
- Generally, random errors are more influential than systematic errors.

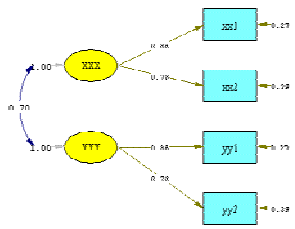
Aggregation error

- An obvious form of error that arises in education (in particular if constrained by common denominator harmonization) is aggregation error == collapsing of heterogeneous categories.
- Aggregation error arises primarily as random error.
- Simulation shows: aggregation error become substantial at 5 categories or less. It can be perfectly estimated and corrected with a multiple indicator design.

Aggregation error: simulation

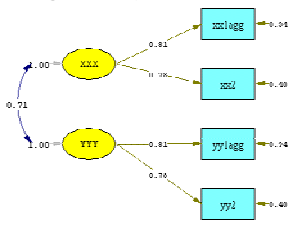
r	0.710
10 groups	0.676
6 groups	0.660
5 groups	0.650
4 groups	0.631
3 groups	0.593
2 groups	0.493

Only random error …



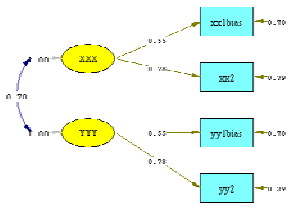
Chi-Square=0.17, df=5, P-value=0.99943, RMSEA=0.000

With aggregation error (5 categories) …



Chi-Square=0.47, df=5, P-value=0.99328, RMSEA=0.000

With systematic bias…



Chi-Square=[illegible], df=5, P-value=[illegible], RMSEA=[illegible]

Harmonizing education

- There is no need for it. For most purposes using rank scores within countries / cohorts is fine.
- It cannot be done: education is strongly locally arranged. Educational programs are like political parties.
- If you do it, do not use a common denominator approach with a limited set of categories, but scale towards an underlying dimension.

Scaling education

- Get unidimensional rank-order, using:
  - Institutional information about program sequence / requirements
  - Scale by validation criteria (occupation, education of spouse, parents)
  - Standardization (Z- or P-scores) will get you a within-group metric.
- Choose anchor point between countries / educational system to establish a common metric.
- Calibration: Re-express the scores of anchor points in the within-group metric in the common metric. You need at least two anchor points.
- Insert other educational programs into this common metric in between anchor points, while respecting order by (broken) linear interpolation).

Advantages of scaling

- Uses and preserves all information.
- You can accommodate all kinds of input information, with various degrees of detail.
- Clear theoretical interpretation: level of education is a single hierarchy
  - that (best) interprets the way social resource lead to societal outcomes;
  - that best expresses value in outcomes and values in access at the same time.
- Metric information that can be used in interval level models, but can also be made discrete.
- Results are very robust against different ways to obtain the scale scores.

However …

- While scaling is better than any other harmonization strategy, it is not perfect.
- Fortunately we can be perfect using multiple indicator measurement.
- Having a second indicator of level of education is enough to estimate and correct random error.
- We have a second indicator: duration! It may not be as good as type of diploma, but is is good enough to obtain true score measurement.

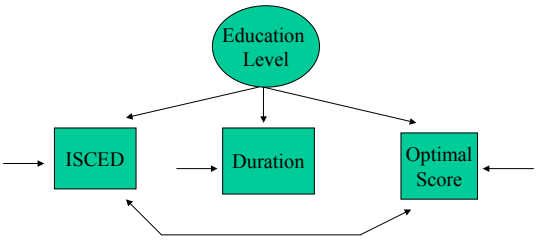
Education in ESS

- Education in ESS is measured in two parallel ways, that are converted into three measures for users.
  - EDUYRS Duration
  - EDLVxx Local measure (varies in detail)
  - EDULVL Local measure post-coded in ISCED (common denominator)
- Most users prefer EDULVL as their indicator.
- Education of spouse, father and mother have been measured, but only in ISCED: EDULVLF, EDULVLM, EDULVLP.
- (A similar approach is used in ISSP.)

Multiple indicator model

- We should not choose between degree and duration, but use them both.
- The three measures can be employed in a multiple indicator model to estimate and correct random error.
- The local measures are rendered comparable by using optimal ordering / scaling.
- Local measure and EDULVL are interpreted as an ordered variable.
- The level of aggregation bias in EDULVL can be estimated by adding the duration as a third indicator.

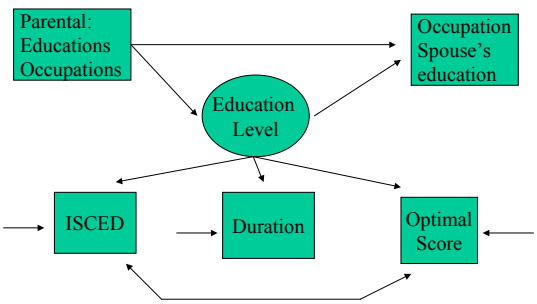
Multiple indicator model



Identification

- The measurement model is not identified in itself.
- But it becomes identified if we add criterion variables:
  - Inputs: (background) variables that produce educational attainment, such as father's and mother's education and occupation.
  - Outputs: (demographic) variables that are produced by education, such as occupation and spouse's education.

Status attainment model



MIMIC model

- Note that we now have an indirect effects model with multiple causes and multiple consequences (MIMIC).
- This gives an elegant interpretation to the measurement of education:
  - The true level of education is the way inputs are transferred into outputs.
  - The best measure of education is the one that minimizes the direct effect of inputs on outputs.
  - Note the similarity with the ISEI construction.

Example: Germany in ESS

- Germany in ESS is an interesting case to look at:
  - German education is often claimed to be an extreme case of a divided system in which a duration measure does not work.
  - In ESS, measurement of German education using EDULVL has gone wrong and leads to very odd results.
  - Note: the problems arise not only because of the question format but also because of incompetent post-processing!

Education in ESS-DE

- In 2004 and 2006, the local education variable has 8 categories, that are effectively reduced to 2-3 categories when expressed in ISCED [EDULVL]. In 2002, the (reconstructed) local measure was more detailed (not shown.)
- The local information is optimally scaled using parental educations, occupations, respondent's occupation and spouse's education as criterion variables → EDUOPT. See Table 4.

Optimal scale scores EDLVDE

Table 4: Optimally scaled levels of education in Germany, 2004, 2006

	2004	2006
2 Left school without school leaving qualification / kein Abschluss	-2.06	-1.67
3 Secondary general school leaving certificate / Hauptschule	-1.04	-0.88
4 Intermediate school leaving certificate / Realschule	-0.03	0.00
5 Higher educ. entrance qual., but not f. university / FH-Reife	0.10	0.57
6 General higher education entrance qualification / Abitur	0.95	1.34
7 Technical college / Fachhochschule	0.83	1.00
8 University degree; PhD/Uniabschl.; Doktortitel	1.98	1.89

Correlations ESS-DE

- Tables 5a and 5b (omitted) show the status attainment correlations for ESS-DE, both for the pooled files and the separate three years.
- Note that both EDUOPT and EDUYRS correlate higher with the other variables than EDULVL.
- Table 6-7 reports on fit and coefficients of various multiple indicators LISREL models.

Comparison to single indicator models 1-3

- To appreciate the gain by using multiple indicators, it is important to compare:
  - Explained variance in educational attainment
  - Explained variance by educational attainment.
- Note that the gains are appreciable, despite the rather good quality of all the education measures.

Results ESS-DE (1)

- Single indicator models (1-2-3) for education all result in downwardly biased effects on and by education.
- The three-indicators model (4) confirms that there is considerable loss of information when using ISCED [EDULVL].
- However, note that the loss in EDUOPT is still 9-10%.
- Also note that duration [EDDUR] is still a reasonable indicator of the true score, and better than EDULVL.

Results ESS-DE (2)

- If we restrict the model to two indicators of education (always involving EDDUR) (models 5-6):
  - The point-estimates remain virtually unchanged.
  - This is also true when we combine EDDUR with EDULVL ("two bad measures").
- However, note the T-values and the loss of statistical power when dropping one indicator.

Measurement models

Table 7: Structural and measurement effects in status attainment model, using different specifications. Coefficients from standardized solutions. GERMANY 2002-2006.

Measurement models						
FISEI	0.881	0.88	0.881	0.881	0.881	0.881
FCRUDEX	0.866	0.866	0.865	0.866	0.866	0.866
MISEI	0.822	0.821	0.822	0.821	0.822	0.82
MCRUDEX	0.841	0.842	0.842	0.841	0.841	0.842
EDUOPT	1	-	-	0.913	0.906	-
EDULVL	-	1	-	0.744	-	0.755
EDDUR	-	-	1	0.825	0.814	0.848

Education

Table 7: Structural and measurement effects in status attainment model, using different specifications. Coefficients from standardized solutions. GERMANY 2002-2006.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Education						
FEDUC	0.128	0.152	0.116	0.146	0.137	0.152
	-6	-6.6	-5.2	-6.5	-6.1	-6.3
MEDUC	0.062	0.106	0.087	0.075	0.062	0.089
	-2.7	-4.3	-3.6	-3.1	-2.5	-3.4
FOCC	0.275	0.163	0.257	0.293	0.308	0.278
	-9.6	-5.4	-8.7	-9.8	-10.1	-8.3
MOCC	0.178	0.07	0.115	0.178	0.195	0.143
	-5.7	-2.1	-3.6	-5.4	-5.9	-4
R2	0.287	0.157	0.226	0.329	0.345	0.294

Occupation

Table 7: Structural and measurement effects in status attainment model, using different specifications. Coefficients from standardized solutions. GERMANY 2002-2006.

Occupation						
FOCC	0.135	0.228	0.18	0.081	0.068	0.104
	-5.9	-9.7	-7.7	-3.5	-2.2	-3.5
MOCC	-0.021	0.039	0.017	-0.048	-0.058	-0.027
	-0.9	-1.6	-0.7	-2.1	-2.2	-1.1
EDUC	0.543	0.402	0.459	0.638	0.655	0.606
	-31.2	-24.2	-26.7	-30.1	-12.8	-11.8
R2	0.371	0.3	0.325	0.437	0.444	0.423

Spouse's Education

Table 7: Structural and measurement effects in status attainment model, using different specifications. Coefficients from standardized solutions. GERMANY 2002-2006.

Spouses Education						
FEDUC	0.069	0.07	0.077	0.146	0.059	0.056
	-3.1	-3	-3.4	-2.5	-2.6	-2.4
MEDUC	0.129	0.113	0.116	0.111	0.115	0.089
	-4.9	-4.6	-4.7	-4.6	-4.7	-4.3
FOCC	0.102	0.153	0.122	0.075	0.066	0.086
	-3.3	-5	-4	-2.4	-2.1	-2.7
MOCC	-0.081	-0.045	-0.058	-0.093	-0.101	-0.079
	-2.5	-2.1	-1.8	-2.8	-3	-2.4
EDUC	0.302	0.226	0.252	0.362	0.371	0.35
	-15.1	-12.3	-13.1	-15.9	-10.9	-10.2
R2	0.173	0.153	0.158	0.196	0.193	0.194

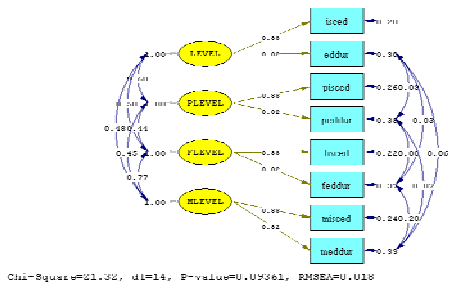
Brittannia rules the waves

- GB is another notorious problematic case in educational measurement, due to the complexity (and lack of standardization) in the system.
- Unlike the Germans, the Brits often solve their problems by resorting to duration (leaving age).
- In ESS-GB the measurement is too crude to be acceptable to the general data.
- However, in ESS Round 2 the Brits have done it right: measure duration for respondent, spouse, father and mother!

ESS-GB (Round2)

- A multiple indicator model (MTMM) shows:
  - Measurement models for all actors can be constrained to be identical
  - Measurement coefficients are 0.86 (EDULVL) and 0.82 (duration).
  - Appreciable correlated error in duration measures, in particular for father-mother.
  - Error-corrected structural correlations look fine.
- This is the best way to measure education in comparative research!

England, Round2



Chi-Square=21.32, df=14, P-value=0.0561, RMSEA=0.018

Recommendations (1)

- Encourage countries to ask about all educations in (more) detail, using a single showcard.
  - Countries that presently ask only 6-7 alternatives or less (Austria, Great Britain, Bulgaria, Finland, Cyprus, Greece, Israel) should be strongly encouraged to ask more.
  - Countries that use multiple, sequential questions (Germany, Estonia), should be encouraged to combine these into a single showcard.
- Be as detailed as possible on the EDLVxx showcard, but present the alternatives in clear hierarchical order, i.e. by (expected) level of occupation.
- The question should be about highest completed / currently attended education.

Recommendations (2)

- The crucial element of improvement of measurement of education is to have a second parallel indicator of level of education. This condition has been fulfilled by the measurement of duration (EDUYRS), as it presently has been included for the respondent. This can be used to estimate and correct random measurement error.
- Years of education should be added for spouse, but preferable also for father and mother to allow for modelling systematic errors.

Recommendations (3)

- Provide category titles in national language (romanized) with the appropriate national abbreviations. The labels should be immediately recognizable to a local expert. Do not use ‘translations’ like “grammar school” or “high school”.
- Classify students by the education of current enrollment.

Recommendations (4)

- ISCED(ES) [Schneider] will be a good harmonization frame, but it should not be a harmonization harness. There is no need:
  - To provide alternatives in each and every category
  - To leave out alternatives that do not fit well with the harmonization frame.
- Rather, the emphasis should be on the value or level of each nationally appropriate alternative. So only combine educational categories that are truly identical.

Recommendations (5)

- Revise the EDULVL in earlier rounds according to ISCED(ES).
- Add EDLVDE for Round 1. Have EDLVDE in all rounds reconstructed by someone who understands the German education system and how to operate SPSS.

Recommendations (6)

- Appoint an expert group that concentrates on social background variables. These are the variables that everybody uses! Urgent attention is required for education, income, occupation and immigration status.
- Make all country-specific variables recognizable via a suffix, also in the earlier rounds.
- Put all the country-specific variables into a single cross-national file, that can be downloaded with the cross-national data.
- If countries deviate from the general format in a variable (in particular in EDULVL or HINCTNT) do not demote this variable to the country specific file, but simply add a suffix and keep it in the main file.