

The cross-national validity of background measures in PISA

Cees Glas & Khurrem Jehangir

Members of Core B

Consortium responsible for background questionnaires

Topics to be discussed

- Short overview of methodology for the analysis of the background questionnaires
- Research questions following the field trial for cycle 2009
- Results of field trial data analysis for cycle 2009

Background Questionnaires (BQs)

- Student questionnaire
 - School questionnaire
 - Parent questionnaire
 - ICT questionnaire
-
- BQs have stand-alone items and scales
 - The presentation pertains to the scales

Example of scale

Q Which of the following are in your home?

- a) A desk to study at
- b) A room of your own
- c) A quiet place to study
- d) A computer you can use for school work
- e) Educational software
- f) A link to the Internet
- g) Classic literature (e.g. <Shakespeare>)
- h) Books of poetry
- i) Works of art (e.g. paintings)
- j) Books to help with your school work
- k) A dictionary
- l) A dishwasher
- m) A <DVD> player

Example of a scale

Q How often do you read these materials because you want to?

- a) Magazines
- b) Comic books
- c) Fiction (novels, narratives, stories)
- d) Non-fiction books
- e) Newspapers
- f) Manuals, instructions, directions
- g) Tables, graphs, diagrams, maps
- h) Science fiction
- i) Subtitles on television

Research questions

Using the field trial data, two research questions were investigated:

- Do more complicated IRT models attribute to the precision of the ordering of countries? The conclusion was negative: the GPCM was not chosen over the PCM.
- Does the use of country-specific item parameters to model cultural bias attribute to the precision of the ordering of countries? The conclusion was that was worth pursuing for at least one scale (the scale “Liking of Reading”)

Models for scaling the BQ

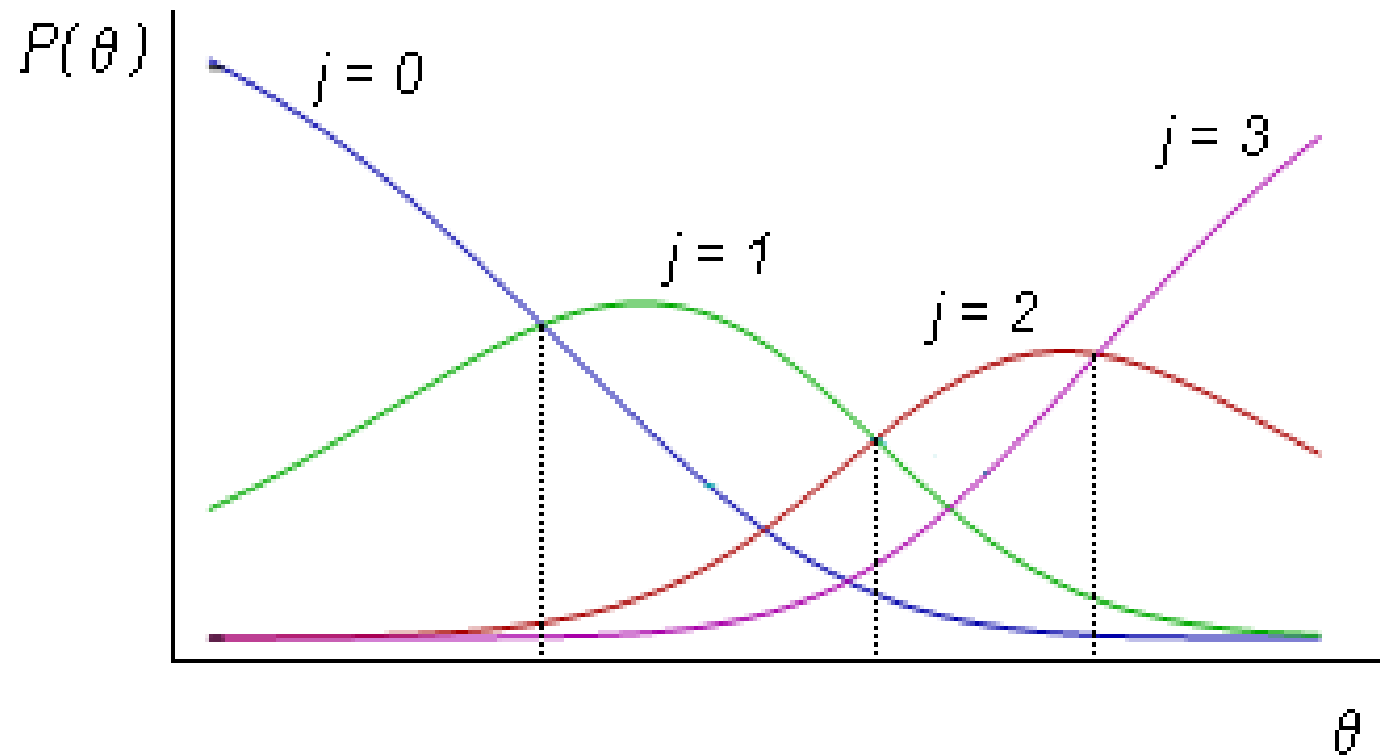
Partial Credit Model:

$$P(X_{nij} = 1 | \theta_n) = \frac{\exp(j\theta_n - \beta_{ij})}{1 + \sum_{h=1}^m \exp(h\theta_n - \beta_{ih})}$$

Generalized Partial Credit Model

$$P(X_{nij} = 1 | \theta_n) = \frac{\exp(\alpha_i j \theta_n - \beta_{ij})}{1 + \sum_{h=1}^m \exp(\alpha_i h \theta_n - \beta_{ih})}$$

Partial Credit Model



Research Question 1: PCM or GPCM

- Analysis 1: MML estimates per scale, per country to evaluate fit using observed and expected item scores
- Analysis 2: MML estimates per scale for all countries simultaneously to assess the ordering of the latent means of countries

Analysis per scale, per country

Scale IC08, Norway

Item	LM	df	Groups: 1		2		3		
			Prob	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
1	0.11	2	0.95	1.67	1.68	2.39	2.38	2.68	2.68
2	4.41	2	0.11	0.76	0.80	1.37	1.37	1.86	1.81
3	4.76	2	0.09	1.85	1.80	2.42	2.49	2.75	2.74
4	6.59	2	0.04	2.14	2.12	2.84	2.84	2.94	2.96
5	1.56	2	0.46	1.70	1.70	2.55	2.53	2.77	2.79

Scale IC08, Slovenia

Item	LM	df	Groups: 1		2		3		
			Prob	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
1	1.87	2	0.39	1.86	1.85	2.45	2.48	2.79	2.77
2	8.61	2	0.01	1.27	1.22	1.73	1.84	2.32	2.24
3	4.98	2	0.08	1.90	1.86	2.53	2.56	2.81	2.82
4	10.65	2	0.00	2.34	2.31	2.81	2.86	2.99	2.97
5	9.49	2	0.01	1.79	1.76	2.49	2.47	2.75	2.79

Fit GPCM within countries: residuals above 0.10 in 55 countries

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
ST20A	0	0	0									
ST20B	0	0	0	0	1	3	0	0				
ST20C	0	0	0	0	0							
ST21	6	0	0	0	0							
ST24	0	0	0	0	0	0	0	0	0	0	0	0
ST25	3	6	10	2	10	4	6	1	21			
ST26	0	6	0	3	0	0	0	8	0	0	1	
ST35	0	0	1	0								
ST37	0	0	0	2	0							
ST40	0	0	0	0	2	0						
ST41	1	0	0	3	4	0						
ST42	0	0	0	0	0	6	0	0	0			
ST43	0	0	0	0								
ST44	0	0	0	0	0	0	0	0				
ICT4	0	0	0	0	0	0	0	0	0	0		
ICT5	6	0	0	3	0	0						
ICT6	4	0	0	0	0	0	0	0	0			
ICT8	4	0	0	0	0	0						
ICT10	0	0	0	0	1							

Scales in IRT analysis

Label	Scale	N	R(rank)	R
ST20	Which of the following are in your home?	16	0.989	0.990
ST21	How many of these are there at your home?	5	0.972	0.981
ST24	Reading Attitude	14	0.988	0.991
ST25	Like Reading	9	0.981	0.983
ST26	Online Reading Activities	11	0.989	0.995
ST35	School Climate	4	0.905	0.915
ST37	Classroom Climate	5	0.911	0.922
ST40	Classroom Engagement in Reading	6	0.942	0.942
ST41	Stimulate Reading Engagement	6	0.958	0.954
ST42	Teacher Structuring Strategies	9	0.994	0.997
ST43	Instruction Teachers Strategies	4	0.905	0.911
ST44	Activities in School	8	0.992	0.997
IC04	Home Usage of ICT	9	0.988	0.991
IC05	ICT for School Related Tasks	5	0.951	0.962
IC06	Use of ICT at School	10	0.952	0.969
IC08	ICT Competency in Different Contexts	5	0.912	0.934
IC10	Attitude Towards Computers	4	0.910	0.912

N : number of items;

R(rank) : rank correlation order of countries obtained using PCM and GPCM

R : correlation between country means on scale obtained using PCM and GPCM

Research Question 2: Cultural Bias

- Analysis 1: MML estimates per scale, over all countries to evaluate fit using observed and expected item scores
- Analysis 2: MML estimates per scale over all countries simultaneously to assess the ordering of the latent means of countries

Cultural bias: country-specific item parameters

Assumption: a number of items in the scale is free of cultural bias

Some items measure the same construct over countries but have a different metric

Example: the number of cars owned by a family is a proxy for wealth, but the scales are different in New York (Amsterdam) and Texas (Twente)

Solution: different weights, i.e. item parameters

Scales in IRT analysis with country-specific item parameters

Label	Scale	PCM		GPCM	
		R(rank)	R	R(rank)	R
ST25	Like Reading	0.823	0.664	0.881	0.765
ST26	Online Reading Activities	0.972	0.976	0.979	0.979
ST42	Teacher Structuring Strategies	0.979	0.987	0.988	0.993
ST44	Activities in School	0.940	0.950	0.950	0.963
IC04	Home Usage of ICT	0.989	0.989	0.991	0.991

R(rank) : rank correlation between order of countries obtained using PCM and GPCM
R : correlation between country means on scale obtained using PCM and GPCM

Rank order and mean scale level of countries on the scale
ST26 for the GPCM and PCM with and without country-
specific item parameters

	Rank Order				Mean on Latent Scale			
	PCM Without	PCM With	GPCM Without	GPCM With	PCM Without	PCM With	GPCM Without	GPCM With
ALB	17	17	22	20	-0.129	-0.200	-0.157	-0.219
AUS	38	41	41	41	0.140	0.285	0.128	0.185
AUT	53	54	52	53	0.539	0.823	0.541	0.744
AZE	33	31	26	24	0.089	0.109	-0.066	-0.115
BGR	10	13	14	15	-0.336	-0.413	-0.267	-0.326
CAN	44	50	49	49	0.225	0.422	0.276	0.366
CHE	14	11	7	7	-0.291	-0.422	-0.484	-0.640
COL	9	9	12	13	-0.350	-0.447	-0.302	-0.362
CZE	12	10	13	12	-0.317	-0.438	-0.297	-0.385
DEU	39	36	33	33	0.148	0.210	-0.001	0.000
DNK	48	47	44	43	0.268	0.377	0.171	0.217

Conclusions

- Do more complicated IRT models attribute to the precision of the ordering of countries? The conclusion was negative: the GPCM was not chosen over the PCM.
- Does the use of country-specific item parameters to model cultural bias attribute to the precision of the ordering of countries? The conclusion was that was worth pursuing for at least one scale (the scale “Liking of Reading”)