Similarities and differences between groups of countries concerning relative weaknesses and strengths

GA meeting, Wien October 2014

Svein Lie

University of Oslo

ENHET FOR KVANTITATIVE UTDANNINGSANALYSER

ekva

The role of achievement test items

- A test consists of numerous test items that together represent a good measurement of the actual competency.
- This collection of items must ensure high validity as well as reliability:
  - Validity: good coverage of the decided framework concerning types of competencies, item formats, difficulties, etc
  - Reliability: ensuring low measurement errors, giving good estimates of their magnitude
- A high quality test score can then be calculated by combining data from many items into a score
  - How to do that depend on the detailed measurement model

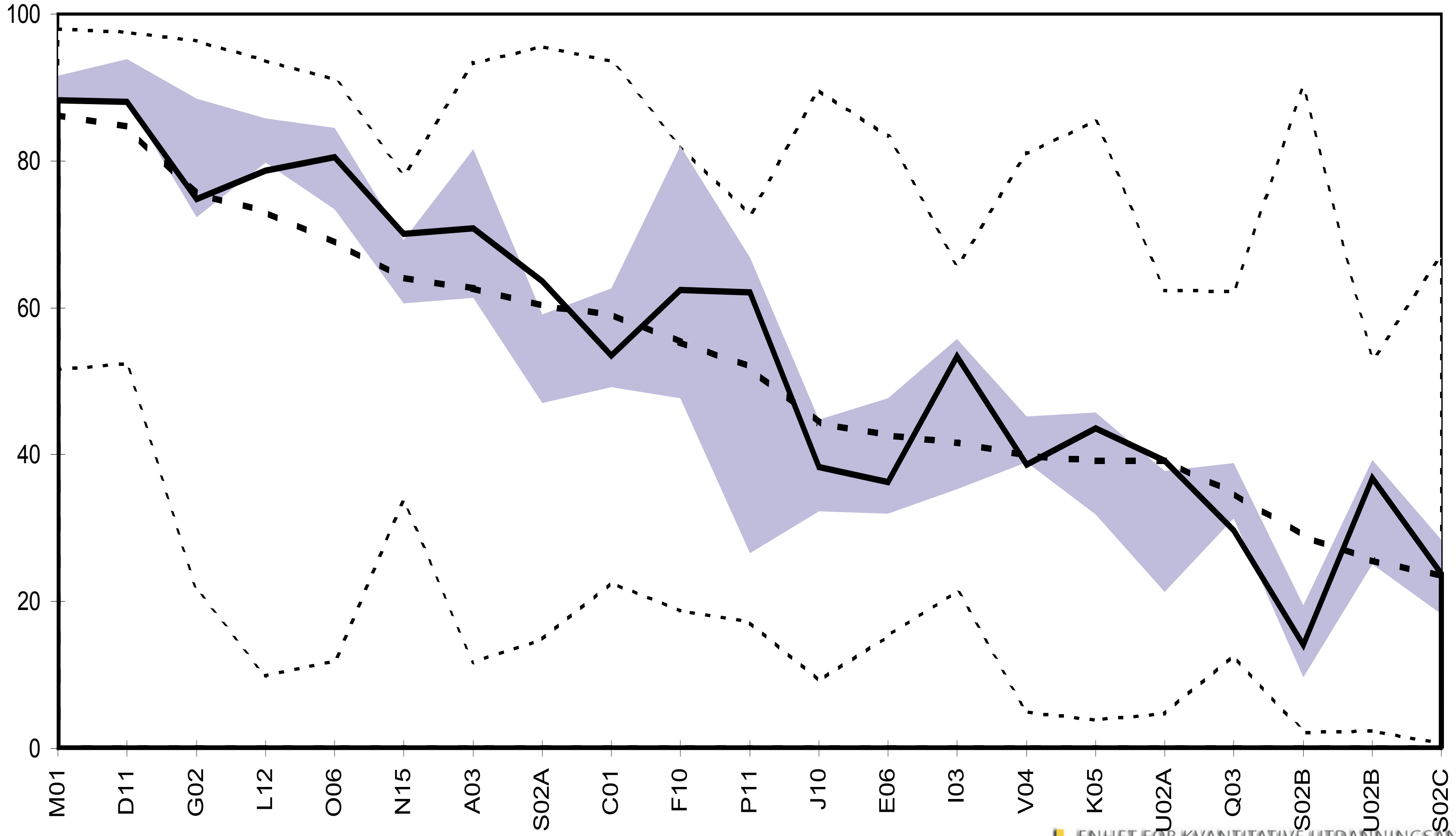
A test item perspective

- From response to item score:
  - for MC items: *which* alternative chosen
  - for constructed (open) response items: *what* was written, type of correct or wrong response
    - For many items reports include this information
    - Two-digit scoring rubrics
- From item scores to scale score
  - Only ”sum” of p-values ”counts”
  - Valuable to investigate also the ”90 % lost” information
  - Not only random noise!!!

Looking for interesting secondary information from test items

- Two examples:
  - Investigating student conceptual understanding by investigating responses to one or a few items
    - A number of such studies has been carried out at EKVA
  - Looking for meaningful patterns in the item residuals (remains when the score aspect has been taken out)
    - An in-depth study of relative strengths and weaknesses for countries

Norway in the ”Nordic river” (excl. Finland) p-values items in TIMSS 1995



ENHET FOR KVANTITATIVE UTDANNINGSANALYSER

ekva

Calculating p-value residuals

- How much better or worse is a national p-value than expected, considering
  - the international difficulty of the item
  - and the overall score for the country?

Looking for patterns in the residual p-value matrix

- Possible patterns of countries
  - pairs or groups of countries showing similar strengths and weaknesses
- or of items
  - groups of items that tend to discriminate similarly between countries
- Tools: correlations, cluster analysis etc.

In practice

- Very simple procedure
  - Start with a p-value matrix: item by country
  - Average p-values calculated both for countries and for items
  - Subtract actual item average as well as country average to get country/item residuals
  - How much better or worse did a particular country score than expected from the country's overall score and the item's overall difficulty?

Countries	A	B	C	Mean
Item 1	45	60	32	45,7
Item 2	57	80	53	63,3
Item 3	38	65	38	47,0
Mean	46,7	68,3	41,0	**52,0**

Countries	A	B	C	Mean
Item 1	-1,7	-8,3	-9,0	-6,3
Item 2	10,3	11,7	12,0	11,3
Item 3	-8,7	-3,3	-3,0	-5,0
Mean	0,0	0,0	0,0	**0,0**

Countries	A	B	C	Mean
Item 1	4,7	-2,0	-2,7	0,0
Item 2	-1,0	0,3	0,7	0,0
Item 3	-3,7	1,7	2,0	0,0
Mean	0,0	0,0	0,0	0,0

Correlations of p-value residuals between Norway and all other countries. Mathematics. TIMSS 1995

TOP 10		BOTTOM 10	
		….	
Sweden	.68	Kuwait	-.25
Iceland	.55	Israel	-.25
Denmark	.47	Singapore	-.30
Germany	.40	Thailand	-.34
Switzerland	.37	Korea	-.34
Scotland	.36	Russia	-.35
New Zealand	.36	Hong Kong	-.37
England	.35	Bulgaria	-.38
Netherlands	.30	Iran	-.43
Australia	.29	Romania	-.44

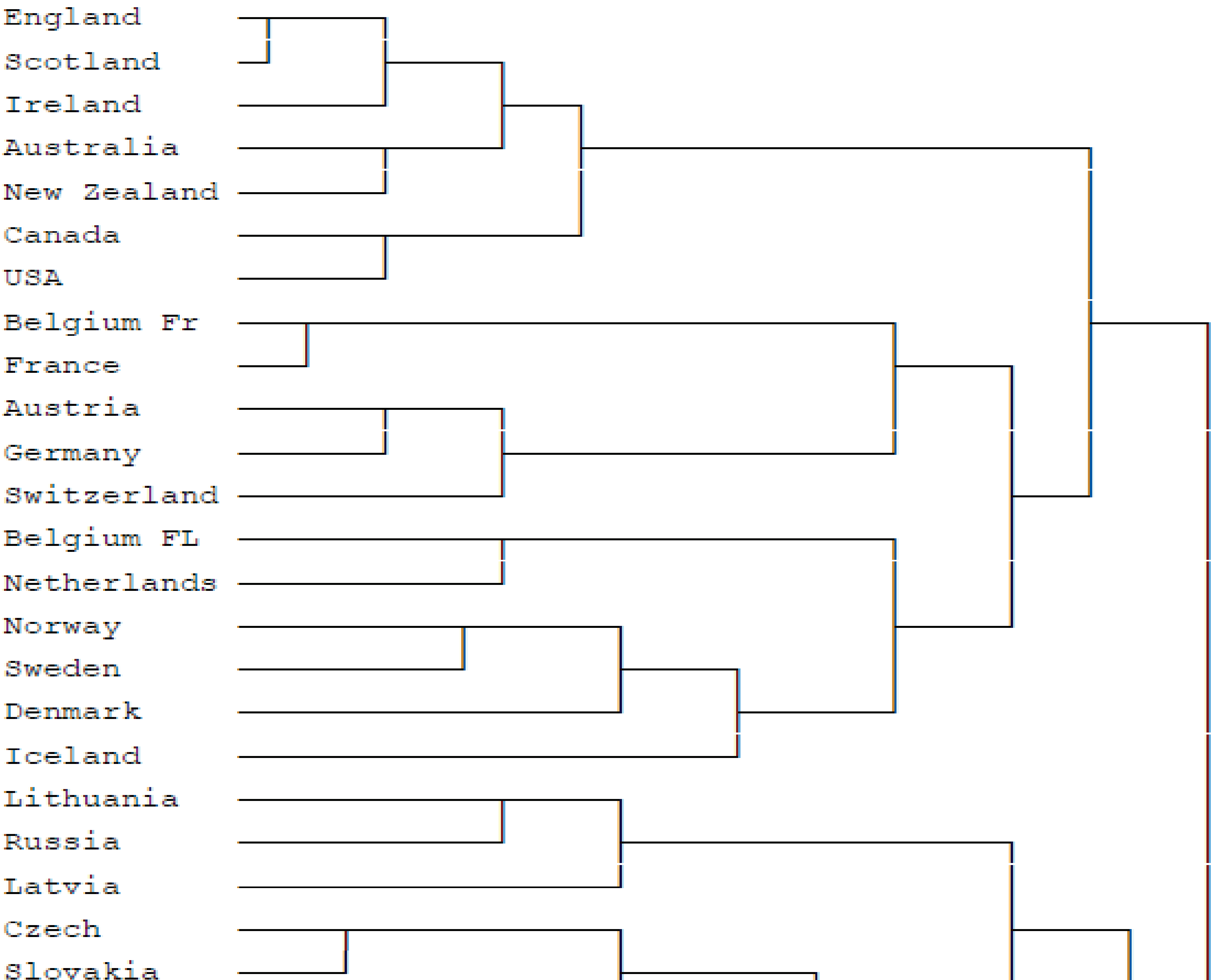
Clusters with labels and reliability TIMSS Science 1995

- East Asia (.52): Hong Kong, Japan, Korea, Singapore, Thailand
- East Europe (.68): Bulgaria, Czech rep, Hungary, Latvia, Lithuaenia, Romania, Russia, Slovac rep, Slovenia
- English-speaking (.88): Australia, Canada, England, Ireland, New Zealand, Scotland, USA
- North Europe (.73): Denmark, Iceland, Norway, Sweden, Belgium (Fl), Netherlands, Switzerland
- South Europe (44): Cyprus, Greece, Portugal, Spain

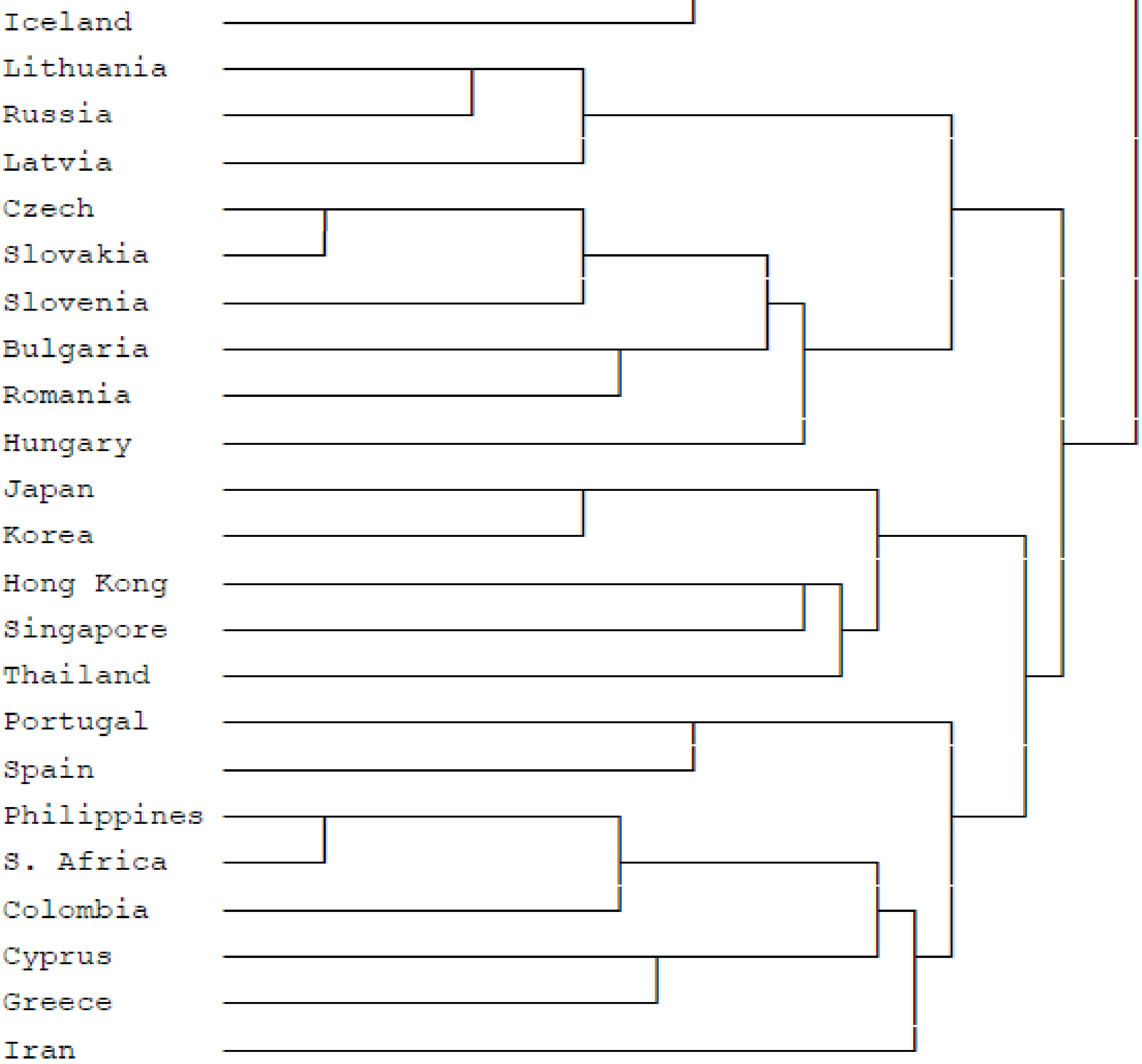
Cluster analysis:

- A method to cluster variables by agglomeration of cases into ever larger groups based on similarities
- Various measures of similarity, e.g.
  - Correlations (most common), «Distance» etc
- Various rules for combining groups, e.g.
  - Internal cohesion, External isolation

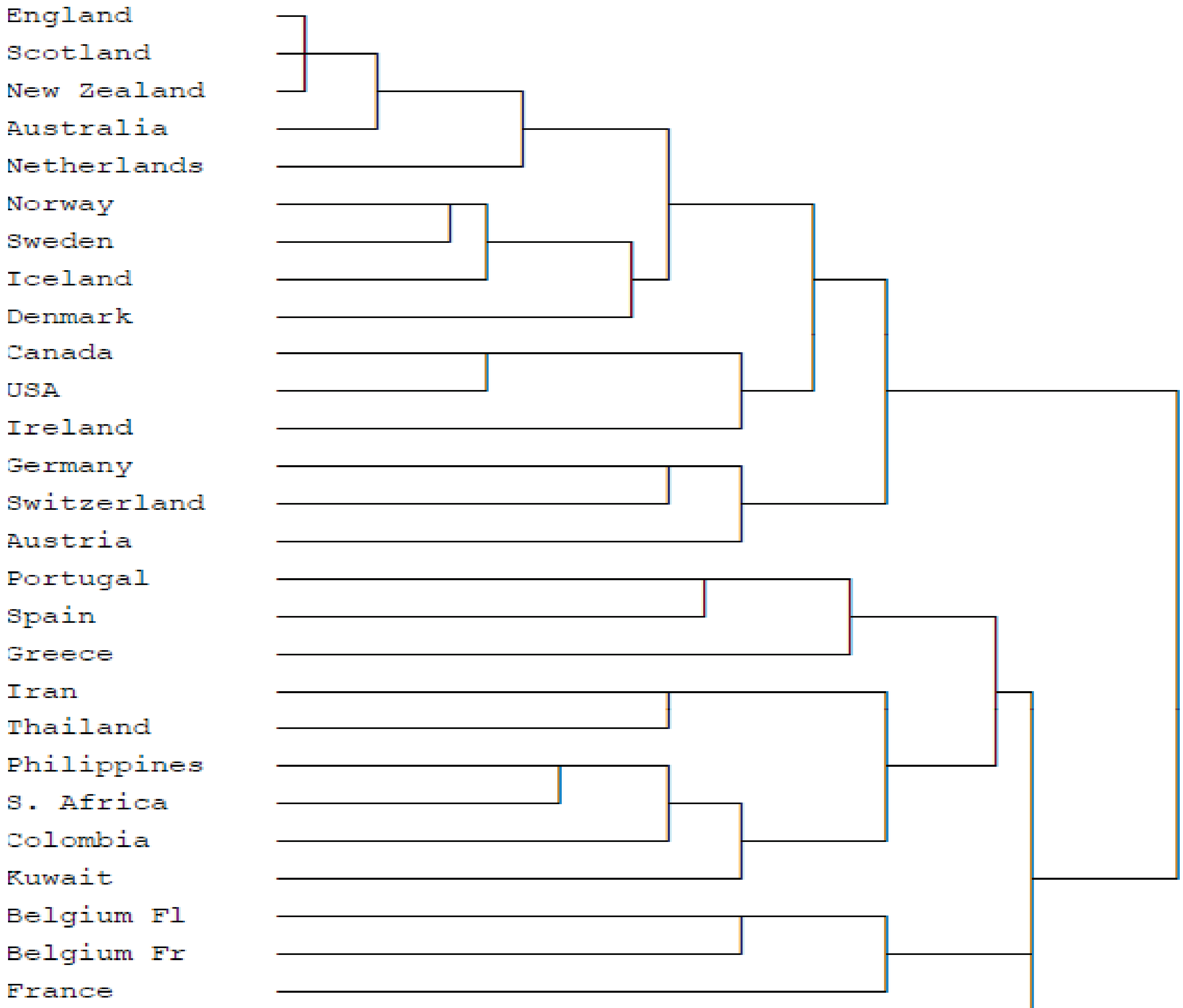
TIMSS Science 2003 (split into halves)

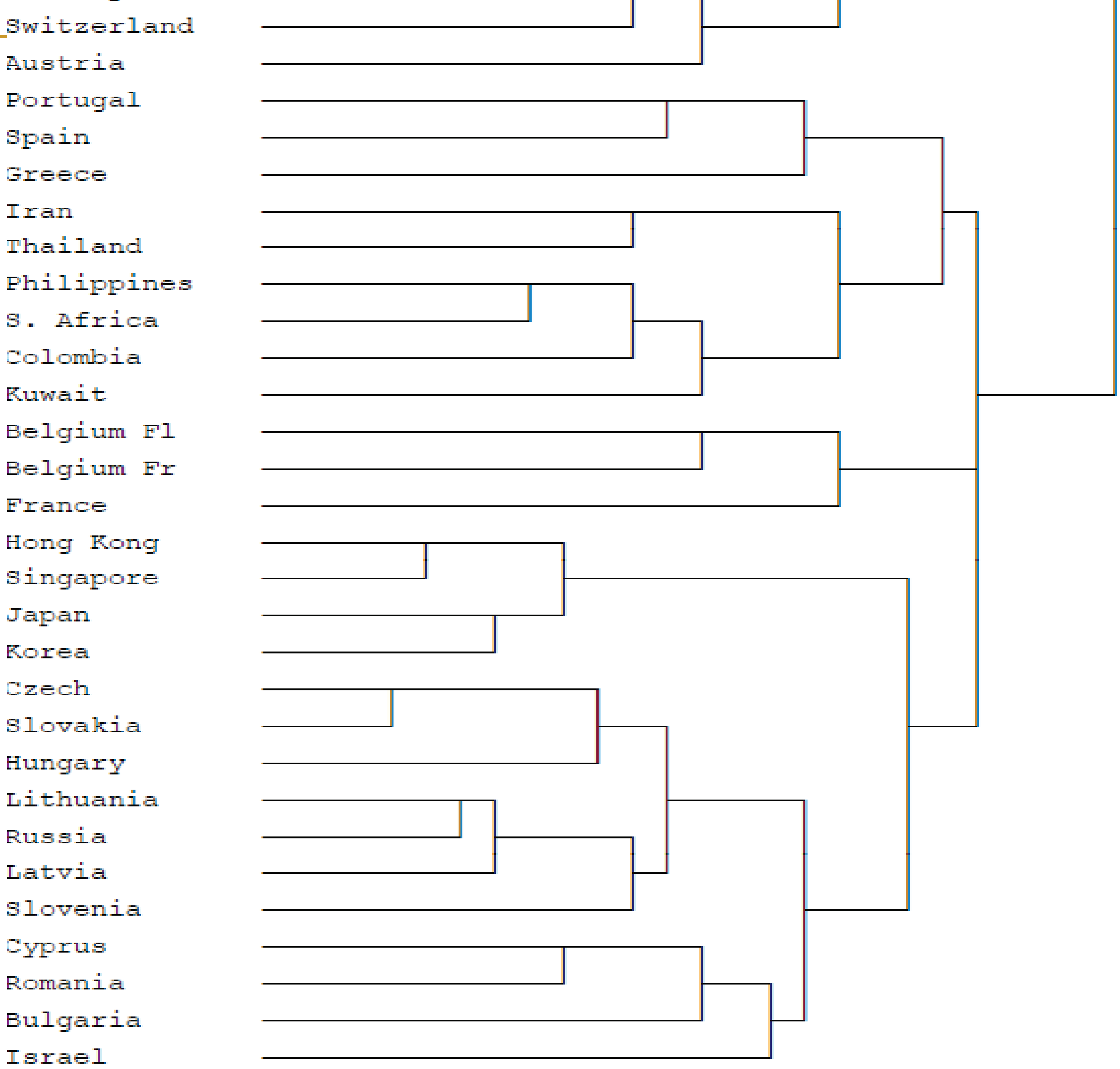


GSANALYSER


Iceland
Lithuania
Russia
Latvia
Czech
Slovakia
Slovenia
Bulgaria
Romania
Hungary
Japan
Korea
Hong Kong
Singapore
Thailand
Portugal
Spain
Philippines
S. Africa
Colombia
Cyprus
Greece
Iran

TIMSS 2003 Math (in halves)




Switzerland
Austria
Portugal
Spain
Greece
Iran
Thailand
Philippines
S. Africa
Colombia
Kuwait
Belgium Fl
Belgium Fr
France
Hong Kong
Singapore
Japan
Korea
Czech
Slovakia
Hungary
Lithuania
Russia
Latvia
Slovenia
Cyprus
Romania
Bulgaria
Israel

PISA science 2006

- Including test items as well as student attitudes towards science

Groups and countries 2006

East Central Europe (former communist)

English-speaking

French/Dutch

German-speaking

Islamic

Latin America

Latin Europe

Nordic

North-East Asia

Why, and So what?

- Actual clustering «mechanisms»: language, geography, political and cultural history, religion, curricular influences etc
- Such groupings of countries invites to focus on:
  - Identifying what constitutes the particularity for each group, and looking for influences whithin and between groups
  - Encouriging regional (groupwise) study reports to better interpretation of national test results concerning important challenges

EKVA

Faculty of Education, University of Oslo

- Norwegian center for administering and running tests and other quantitative studies
  - Internationally: IEA studies, PISA etc
  - National assessments
  - Master- and PhD quantitative projects in science/math/reading education
- International studies:
  - Main task: administrating, running, recording according to given procedures
  - National reporting focusing on Norwegian results/standing in an international context
  - Focusing on comparison with nabour countries
  - Also focusing on trends and links between studies

ENHET FOR KVANTITATIVE UTDANNINGSANALYSER
ekva

Some references

Kjærnsli & Lie:

- *International Journal of Science Education,, 2011/33, p 121-144*
- *Scandinavian Journal of Educational Research, 48/3, p 271-286*
- *IEA Research Conference, Cyprus 2005*