Comparisons Between PISA and TIMSS – Are We the Man with Two Watches?

Dougal Hutchison

Ian Schagen

National Foundation for Educational Research

"The man with one watch always knows what time it is – the man with two watches is never quite sure" (Anon)

OUTLINE OF PRESENTATION

- 1. Introduction
- 2. Frameworks of TIMSS and PISA
- 3. Sampling
- 4. Test items
- 5. Use of IRT
- 6. Background information
- 7. Survey of co-ordinators
- 8. Conclusions and recommendations

One view

- ***"For those doubters who constantly seek to run down (our education performance), we now have the OECD/PISA study – the biggest ever international study of comparative performance of 15-year-olds in 32 countries – which shows UK fourth in science, seventh in literacy and eighth in mathematics. Only Finland and Canada are consistently ahead of the UK – and major countries like Germany, Italy and Spain are well behind".*** **(Normington, 2002).**

Another view

- ***'It is particularly incredible because in the previous year a far more authoritative study- the third international mathematics and science study, conducted by the respected International Association for the Evaluation of Educational Achievement,- put the UK 20th out of 41 countries '*** **(Gibb, 2002).**

COMPARE AND CONTRAST

- **TIMSS (Year 8 Study only)**
- **PISA**

IEA/ TIMSS

- *'the three major international comparisons of mathematics attainment (carried out by) the IEA ( ) have had a greater influence on education world-wide than any other single factor during the last 50 year period'. Brown (1999)*

Aims of TIMSS

- International variations in mathematics and science curricula
- Training of teachers in science and mathematics
- The influence of textbooks
- The course content that is actually taught
- The effectiveness of different instructional practices
- Students' achievement
- The attitudes and opinions of students and teachers
- The role of technology
- Participation rates in pre-university courses
- The effect of tracking, streaming and other practices
- International 'league tables' not mentioned!

Aims of PISA

- ***'The prime aim of the OECD/PISA assessment is to determine the extent to which young people have acquired the wider knowledge in reading literacy, mathematical literacy and scientific literacy that they will need in adult life.'*** **(OECD, 2004, 4).**
- ***'provides insights into the factors that influence the development of the skills at home and at school and examines how these factors interact and what the implications are for policy development'*** **(OECD, 2003, 10).**

Apparent difference in aims

- **TIMSS:** ***"What science have you been taught and how much have you learned?"***
- ***PISA: "What can you do with the science you have been taught?"***

Aims

- **TIMSS: Inspecting what is happening in the classroom with a microscope**
- **PISA : Waiting to see what comes out**

TIMSS and PISA

- **PISA : 3 year cycle**
  Maths and science literacy (reading)
  Pupil and school info

- **TIMSS: 4 year cycle**
  Maths and Science (PIRLS)
  Pupil, school and teacher info

SAMPLING

Sampling - countries

- **TIMSS: Anyone who wants to take part and can afford it**
- **PISA: OECD countries plus some 'partner' countries**

Organizati
Economic
Cooperati

Sampling - schools

- Both operate tightly controlled sampling schemes within countries
- Probability proportional to size (PPS) stratified sampling
- Some allowance for excluding a few schools from population before sampling
- Fixed (and rather arbitrary) limits on response rates in order to be allowed into main analysis
- No allowance for countries which can weight responses to reduce bias

Sampling - pupils

- **TIMSS: Population = all students enrolled in the upper of the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing**
- **PISA: Population = all 15-year-old students attending educational institutions located within the country, in grades 7 and higher.**

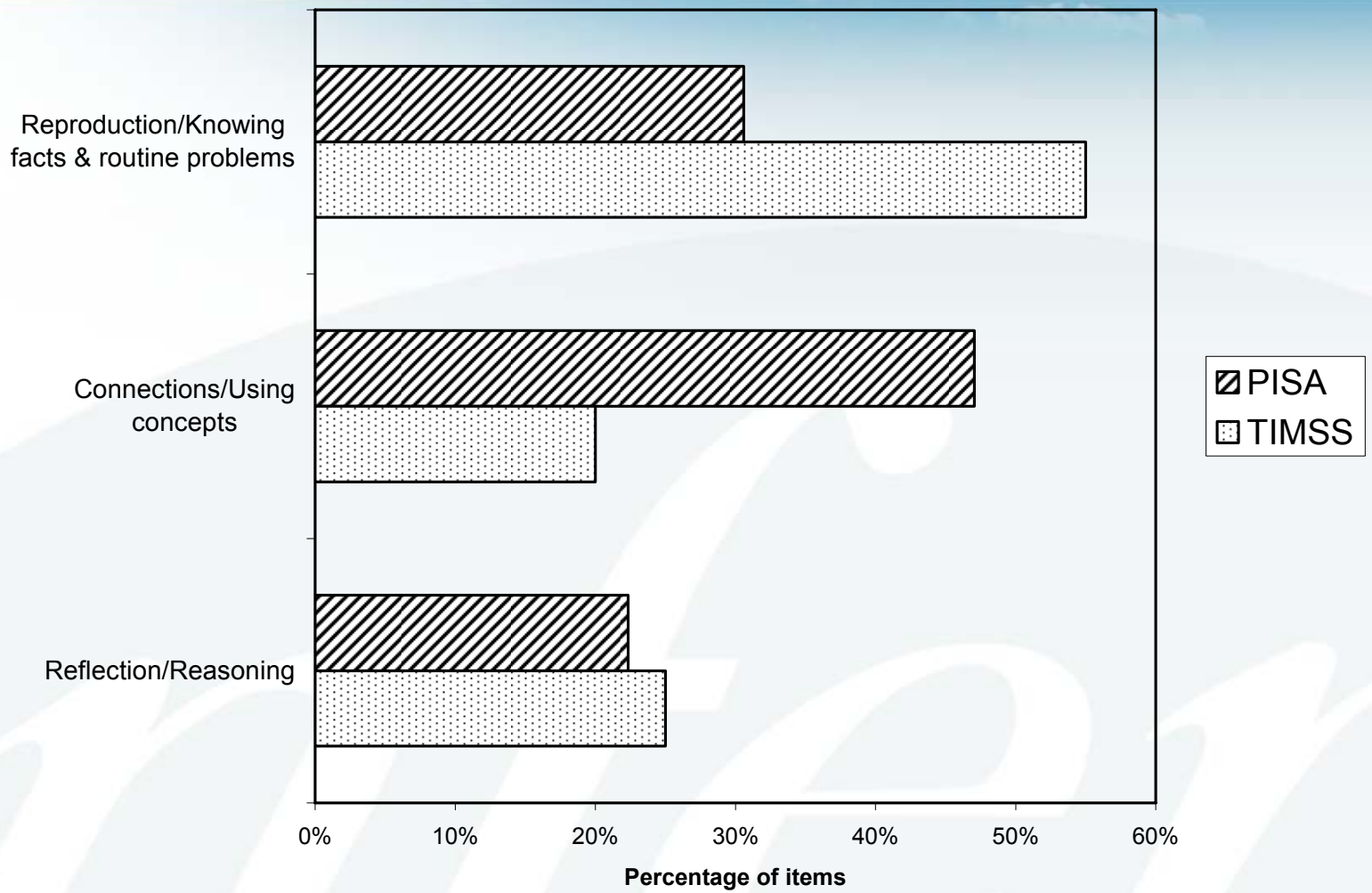
ITEMS

- **PISA Claims to be different**

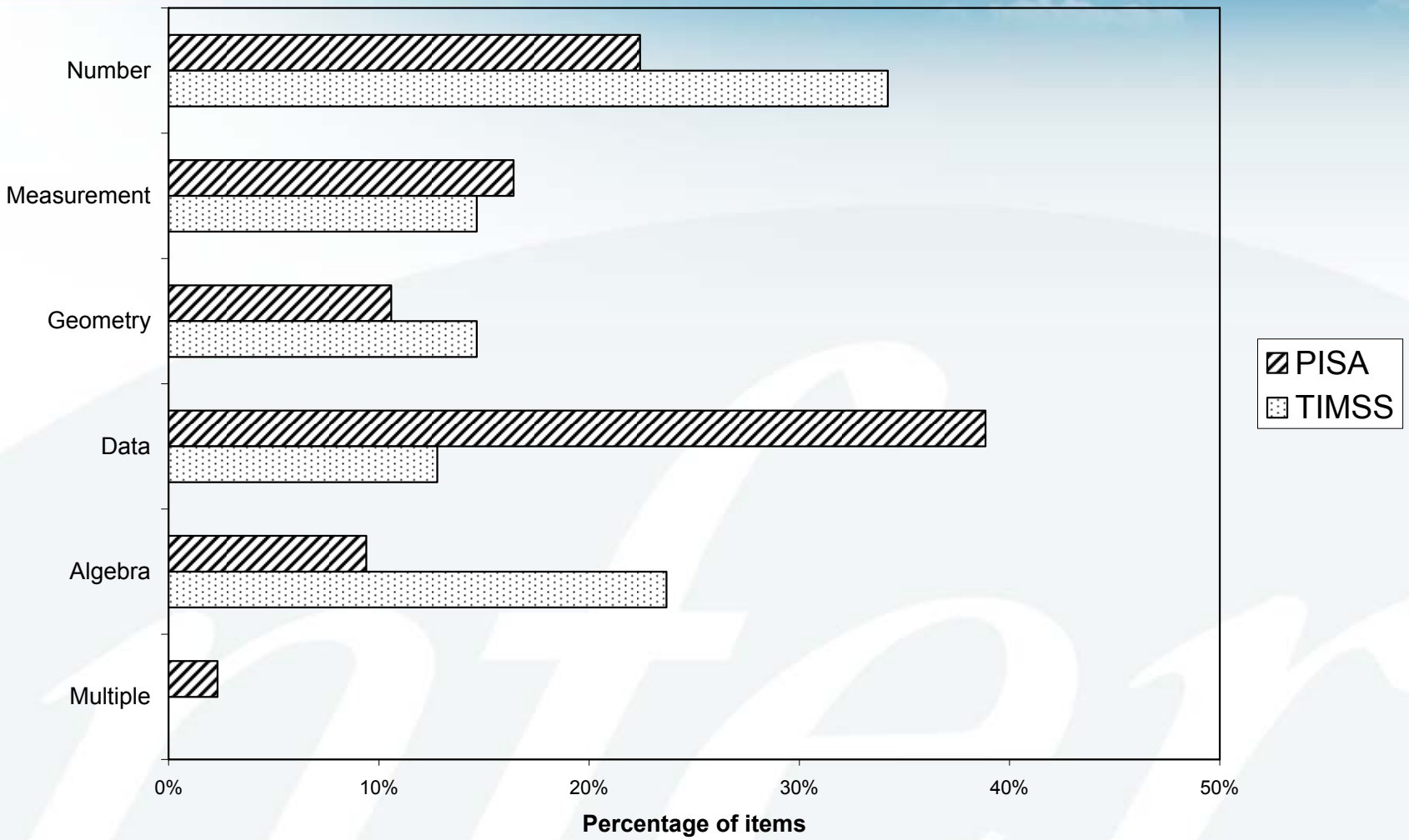
BUT IS IT?

- **You can’t make a three piece suite into a carrot simply by calling it a carrot**

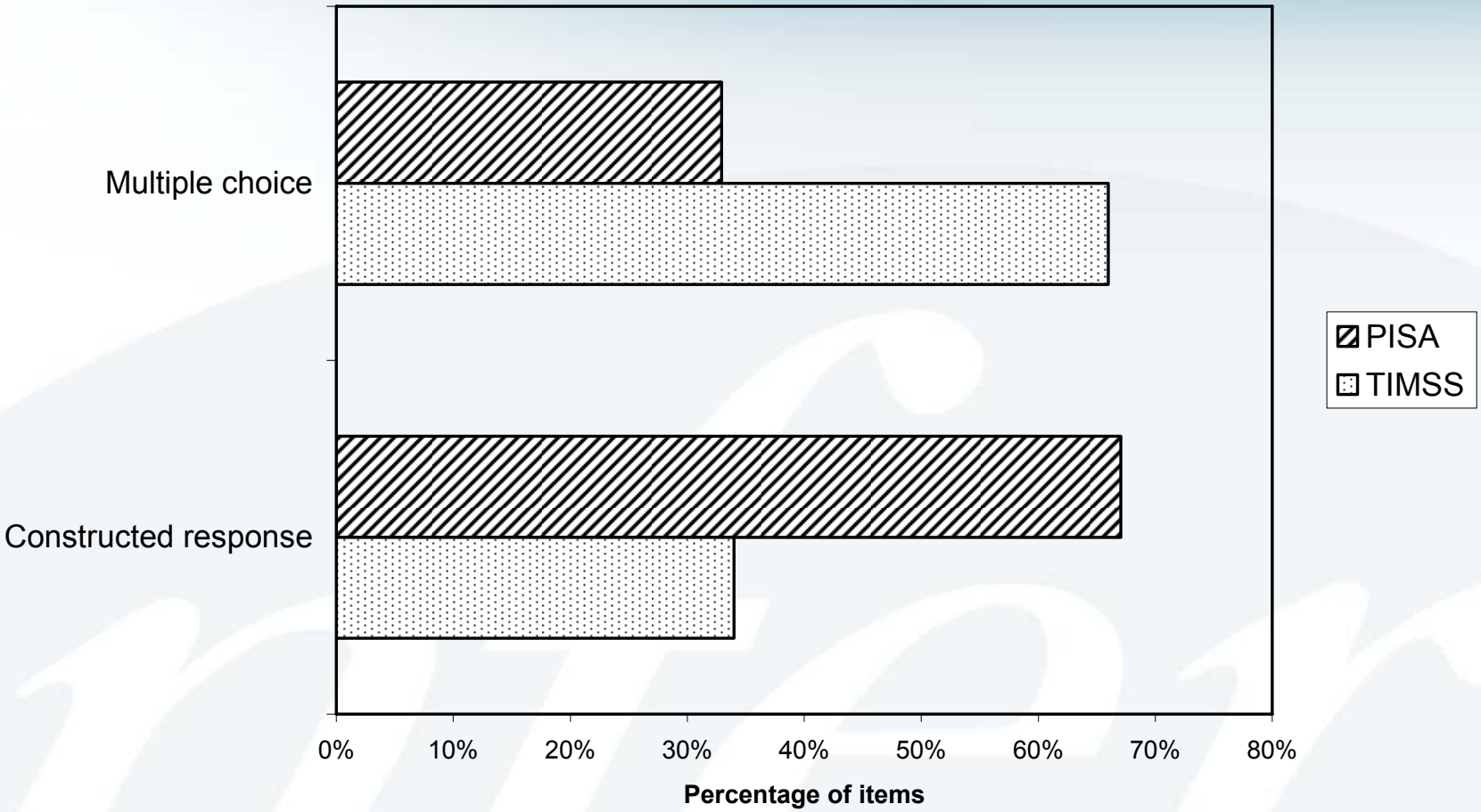
PISA and TIMSS 2003 Mathematics Cognitive/Competency Domains



PISA and TIMSS 2003 Mathematics Content Domains



PISA and TIMSS 2003 Mathematics Item Types



Broad Conclusions from Comparison of Mathematics Items

- TIMSS emphasises items which require the reproduction of facts or standard algorithms, while PISA majors on items which demand connections between existing knowledge.
- TIMSS has a larger number of items focusing on number and measurement, while PISA items are more evenly spread across their content domains.
- A majority of TIMSS items are multiple choice, while a majority of PISA items are constructed response.
- PISA items: multiple questions to one stem

Conclusions from Comparison Study in England

It is the quantity of reading that marks PISA out, not the complexity of the language, which is similarly unfamiliar in both the international studies. The high reading demand of questions in PISA is often accompanied by a relatively lower demand in the mathematics or science required. This reflects the lower level of mathematics or science that students can apply in new contexts as opposed to very familiar ones.

(Ruddock et al., 2006, p.123).

USE OF IRT

TIMSS & PISA IRT Modelling and Scaling

	TIMSS Grade 8 2003	PISA 2003
IRT Model	2/3 parameter (PARSCALE)	1 parameter (CONQUEST)
Separate scales	5 mathematics 5 science	4 mathematics 1 reading 1 science 1 problem solving
Metric (500,100)	Linked to 1995 survey	Reading & science: 2000 Maths & problem solving: new 2003

Issues relating to modelling assumptions

- To what extent do the different modelling assumptions in PISA and TIMSS affect the comparability of results from the two studies?
- How sensitive are these study results in general to modelling assumptions?

Findings from comparison of 1- and 3-parameter models

- Brown et al (2005) compared results for TIMSS 1995 using both models
- Concluded that measures of central tendency across country are robust to model choice
- Not true for measures of dispersion, especially for less developed countries

Similarities between TIMSS and PISA in Modelling Assumptions

- IRT
- Derivation of pupil scales
- Use of plausible values
- Unidimensionality
- Local independence
- Universality of application

- How would results change if other modelling paradigms were used?

BACKGROUND INFORMATION

Homework example

- It is well known that doing homework improves performance
- We expect that the more homework that is done, the higher the attainment
- We do not find this pattern in our data
- Therefore there must be 2 groups of pupils – higher attainers because they do homework, and those who do homework because they are lower attainers
- There is no direct evidence for this, but it allows us to make the data consistent with our preconceptions

Background Information – What and Why?

- Both studies collect background information on students and schools through questionnaires, while TIMSS also collects teacher information.
- There is an implicit assumption of a causal link between background factors and attainment which can be illuminated through analysis of this data.
- This can sometimes lead to speculation as to why expected relationships are not found.

Why are these correlational analyses likely to be misleading?

- Correlation does not imply causation
- Important missing factors are not included
- Background factors are attached to latest performance – no measure of recent progress
- In summary, cross-sectional studies of this type are of seriously limited explanatory value

How could we get more from this kind of analysis?

- Main influence on pupils' attainment is performance at an earlier time point.
- If we are able to allow for this, we can carry out 'value-added' analysis looking at the relationship between background factors and progress over time.
- This requires the collection of longitudinal data for the same pupils at two separate time points.
- This is logistically harder than the current system, but not completely impossible.

Setting up a Longitudinal Study

- Link TIMSS Grade 4 and Grade 8? (wrong interval between sweeps)
- Run a special 'baseline' study a year or two prior to the main study?
- Matching pupils across studies – some countries (e.g. England) have data which could be used for this
- In other cases class lists would need to be collected and pupils matched by name
- If this could be managed for a reasonable number of pupils in a subset of countries it would add considerable value to the international data

SURVEY OF NATIONAL CO-ORDINATORS

Survey of national coordinators

- We emailed 23 national coordinators whose countries took part in both TIMSS and PISA
- Received responses from 13
- NB: Little published in the way of official comparisons by countries – except England, US and Ireland

Were there any differences in your country between the results of TIMSS and PISA?

Of the 13 respondents,

three considered that their PISA results were better,

one that TIMSS results were better,

four reckoned that there were no or slight differences,

one stated that there were differences without specifying what they were, and

one was unable to answer the question as TIMSS did not survey the entire country.

- One country reckoned that TIMSS had shown a decline in performance and PISA had not
- One offered the comment that while there was no overall difference, the boy/girl differences were larger in TIMSS.

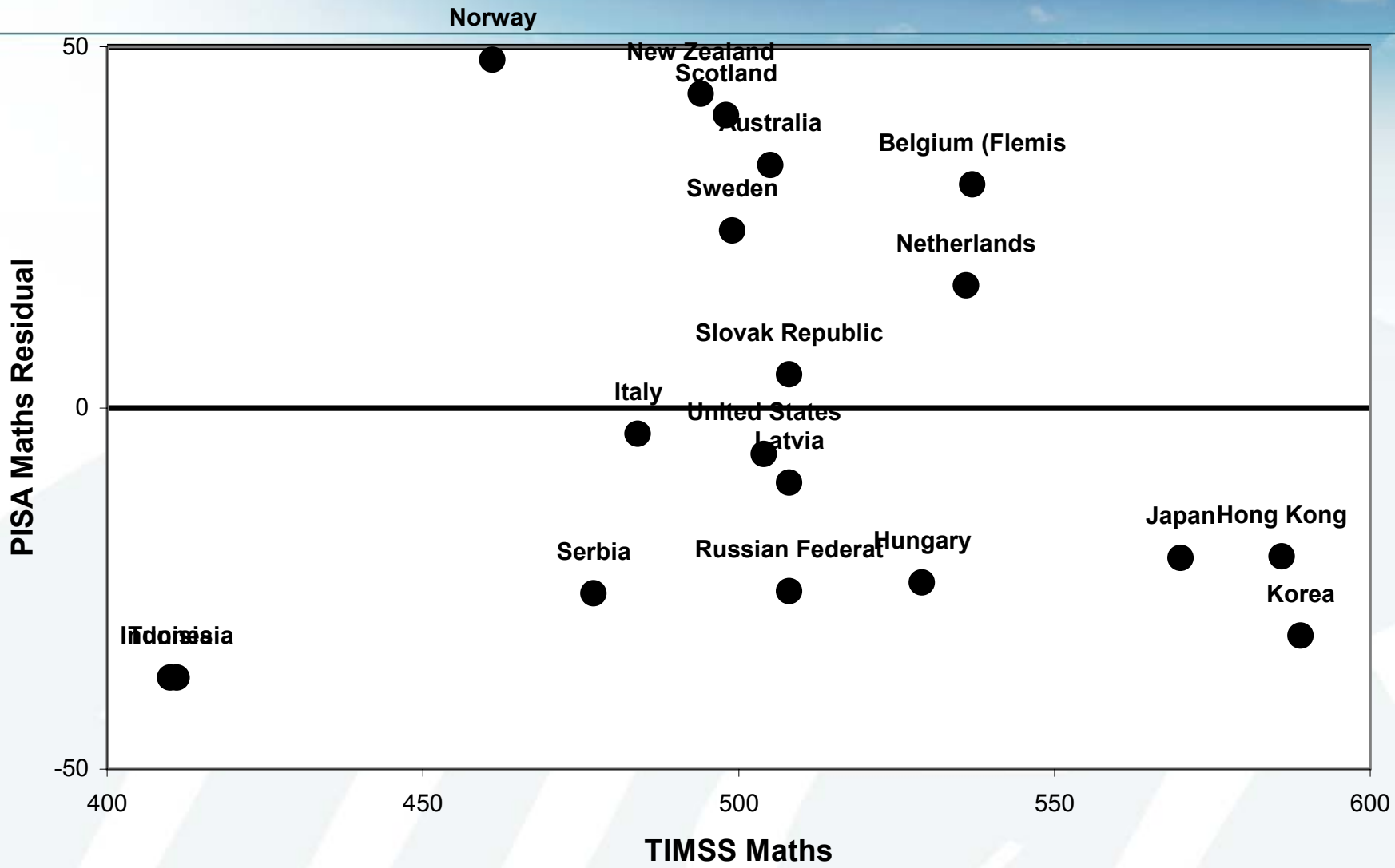
Responses related to perceived differences between PISA and TIMSS

- **Only one country stated that they were surprised by the differences.**
- **None expressed the opinion that differences in the samples were likely to give rise to differences in the results**
- **None considered that differences in the instructions to schools were likely to give rise to differences in the results**
- **None considered that differences in the response rates were likely to give rise to differences in the results**

Possible explanations for perceived differences

- Interaction between age and curriculum
- Effects of specific testing programmes
- Policies on repeating years

Figure 7.3: PISA Maths 2003 Residuals (from TIMSS Maths 2003) vs TIMSS



Pattern of Relative Performance PISA/TIMSS?

- **Broadly, 'first world' nations do better, and 'former Warsaw pact' nations worse, on PISA compared with TIMSS.**
- **Hungary:**

'TIMSS focuses on the curriculum related tasks, while PISA is literacy-based. Hungarian school system is still highly relies on factual knowledge and traditional teaching strategies, so students are relatively good in tasks which are close to their usual classroom tasks, while they meet relatively few literacy-based tasks and they do not know what to do with these'.

***(* Balazsi, 2006)**

CONCLUSIONS AND RECOMMENDATIONS

Recommendations (1)

- **Both IEA and OECD should maintain open access to their methodologies, and encourage criticism and debate from the wider academic community. This should be done in a spirit of openness and willingness to learn and improve, recognising there is not necessarily a 'correct' answer to each technical problem.**
- **More analysis should be carried out on the sensitivity of the main survey results to the exact details of the models used (e.g. 1-parameter versus 2-parameter IRT).**

Recommendations (2)

- The importance of comparing attainment over long time periods should be downgraded, to a maximum of 10 or possibly 12 years.
- The feasibility of introducing a longitudinal study, following the same pupils over a period of years, should be actively investigated.
- There should be careful in-depth investigations of the apparent discrepancies between TIMSS and PISA results at the country level.
- Why two studies?
- Should PISA be looking more at out of school performance?

- Necessarily a shortened version.
- Thanks to Roger Shouse
- Read our paper anyway.