How multi-national studies in education affect your school and the nation

Raymond J. Adams

Dean’s Lecture: September 6 2005
Bivariate data with one variable subject to nonresponse – the bivariate normal case

\[ \theta = (\mu_1, \sigma_{11}, \mu_2, \sigma_{22}, \sigma_{12}), \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \]

\[ y_i = (y_{i1}, y_{i2})^T, \quad \mu = (\mu_1, \mu_2)^T \]

\[ L(\theta \mid Y_{obs}) = \int \prod_{i=1}^{n} P(y_{i1}, y_{i2} \mid \theta) \prod_{i=n_1+1}^{n} P(y_{i1}, y_{i2} \mid \theta) dY_{mis} \]

\[ \propto |\Sigma|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \right\} \]

\[ \times \sigma_{11}^{-(n-n_1)/2} \exp\left\{ -\frac{1}{2 \sigma_{11}} \sum_{i=n_1+1}^{n} (y_{i1} - \mu_1)^2 \right\} \]
Programme for International Student Assessment: PISA

- Overview
- Background
- Issues and validity threats
- Some outcomes
What is PISA?

Programme for International Student Assessment

Funded by the Organisation for Cooperation and Development (OECD)

International data collection has been managed and led from here in Australia since its inception
Developing PISA

- 1995: OECD countries sought
  - Comparative assessment of yield of education systems
  - Better basis defining educational standards

- 1997: OECD countries adopted strategy and financial framework to...
  - Monitor a broad range of curricular and cross-curricular outcomes...
    ...within a comparative framework established collaboratively by countries
The PISA Survey Cycle

- A comprehensive assessment of the yield of education systems
  - including and beyond the curriculum
- Comparable outcome measures
  - that can guide policy decisions and resource allocations
- A strong substantive and multicultural core
- Insights into the mix of factors which contribute to the development of knowledge and skills
  - and how these factors operate similarly or differently across countries
Participation

- **2000**
  - OECD 28, non-OECD 4
  - PISA+, 11 non-OECD
  - 170,000 students

- **2003**
  - OECD 30, non-OECD 12
  - 250,000 students

- **2006**
  - OECD 30, non-OECD 28
  - 350,000 students (estimated)
### OECD countries
- Australia
- Austria
- Belgium
- Canada
- Czech Republic
- Denmark
- Finland
- France
- Germany
- Greece
- Hungary
- Iceland
- Ireland
- Italy
- Japan
- Korea
- Luxembourg
- Mexico
- Netherlands
- New Zealand
- Norway
- Poland
- Portugal
- Slovak Republic
- Spain
- Sweden
- Switzerland
- Turkey
- United Kingdom
- United States

### Partner countries in PISA 2003
- Brazil
- Hong Kong-China
- Indonesia
- Latvia
- Liechtenstein
- Macao-China
- Russian Federation
- Serbia and Montenegro
- Thailand
- Tunisia
- Uruguay

### Partner countries in other PISA assessments
- Albania
- Argentina
- Azerbaijan
- Bulgaria
- Chile
- Colombia
- Croatia
- Estonia
- Israel
- Jordan
- Kazakhstan
- Kyrgyz Republic
- Lithuania
- Macedonia
- Peru
- Qatar
- Romania
- Slovenia
- Chinese Taipei
## PISA survey cycle

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading literacy</td>
<td></td>
</tr>
<tr>
<td>Mathematical literacy</td>
<td></td>
</tr>
<tr>
<td>Scientific literacy</td>
<td></td>
</tr>
<tr>
<td>Background information</td>
<td></td>
</tr>
</tbody>
</table>
### PISA survey cycle

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading literacy</td>
<td></td>
</tr>
<tr>
<td>Mathematical literacy</td>
<td></td>
</tr>
<tr>
<td>Scientific literacy</td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td></td>
</tr>
<tr>
<td>Attitudes to learning</td>
<td></td>
</tr>
<tr>
<td>Background information</td>
<td></td>
</tr>
</tbody>
</table>

**Dean’s Lecture**
**Sept 6 2005**
## PISA survey cycle

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading literacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical literacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific literacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes to learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem solving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background information</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### PISA survey cycle

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2003</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading literacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical literacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific literacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitudes to learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem solving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer-Based Science</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background information</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Basic Methodology

- Target population: 15-year-olds in school
- Target sample sizes of 5250
  - 150 schools of 35 students
- Rotated two-hour written tests
- Student questionnaires
- School questionnaires
- Parent questionnaires
PISA Governing Board

- Sets policy objectives and priorities
- Sets budget
- Monitors adherence to policy objectives
- Guides analysis and reporting of results
- Establishes quality standards
- Adjudicates on breaches of quality standards
Separating policy and science

Consortium commissioned to run PISA

Australian Council for Educational Research
National Institute for Educational Research, Japan
Educational Testing Service, USA
Westat, USA
Netherlands National Institute for Educational Measurement
Separating policy and science

Consortium commissioned to run PISA

OECD manages interactions with PGB

National Project Managers in countries
Innovations in PISA

- International, intergovernmental ownership
  - primary focus on public policy issues
  - sustained commitment
  - collaborative development

- Age not grade-based target population

- Substance
  - not common denominator of national curricula
  - not whether students have learned
  - whether they can use what they have learned
  - Criterion-based reporting
PI SA’s Literacy Orientation

- Projective (rather than reflective) approach to instrument development
- NOT meant to be curriculum independent
  - It is about authentic application of what is learned
  - In contrast to assessments that are like instructional exercises
- A broadening, not narrowing notion
Domain definitions include two components:

- Each domain defined, in terms of knowledge and skills needed in adult life, not merely in terms of mastery of the school curriculum.
- Emphasis is placed on the mastery of processes, the understanding of concepts and the ability to function in various situations related to real life.
Domain Definitions: Reading

Reading is often understood as simply decoding, or reading aloud, whereas the intention of this survey is to measure something broader and deeper. The focus is on the application of reading...
Domain Definitions: Mathematics

..the emphasis is on mathematical knowledge put into functional use in a multitude of different situations in varied, reflective and insight-based ways. Of course, for such use to be possible and viable, a great deal of fundamental mathematical knowledge and skills are needed and such skills form part of our definition of literacy.
Example Reading Items

Working-age population
2656.5

In labour force
1706.5 64.2%

Not in labour force
949.9 35.8%

Employed
1578.4 92.5%

Unemployed
128.1 7.5%

Full-time
1237.1 78.4%

Part-time
341.3 21.6%

Seeking full-time work
101.6 79.3%

Seeking part-time work
26.5 20.7%

Not seeking full-time work
318.1 93.2%

Notes
1. Numbers of people are given in thousands (000s).
2. The working-age population is defined as people between the ages of 15 and 65.
3. People “Not in labour force” are those not actively seeking work and/or not available for work.
Example Reading Items

Retrieving Information

How many people of working age were not in the labour force? (Write the number of people, not the percentage.)
Example Reading Items

Interpreting Text

What are the two main groups into which the working-age population is divided?

A  Employed and unemployed.
B  Of working age and not of working age.
C  Full-time workers and part-time workers.
D  In the labour force and not in the labour force.
Example Reading Items

Reflection on text

The information about the labour force structure is presented as a tree diagram, but it could have been presented in a number of other ways, such as a written description, a pie chart, a graph or a table.

The tree diagram was probably chosen because it is especially useful for showing

A changes over time.
B the size of the country’s population.
C categories within each group.
D the size of each group.
Example: Mathematics Item

If a discount card has the following sequence of Free offers, give reasons to explain whether the shop will be offering the same discount value as the card in the photograph.

1 2 Free 4 5 6 7 Free 9 10
Wrong? Well, I’ve never been good at music questions...
Threats to validity

- Item selection
- Samples
- Translation
- Implementation fidelity
- Motivational issues
Item selection - how it is done

- Test development teams in Norway, Germany, Netherlands, Australia and Japan
- Call on item submissions from all participants
- Items conform to rigorously debated framework
- Items are reviewed, trialled, reviewed and used
Item selection- does it matter? (only if you care about your ranking)

- Item selection does influence rankings
- Item selection does influence trends
- Not in (as yet) predictable ways
  - Language
  - Curriculum
National Rating Categories

- Curricular match
- Match to PISA objectives and framework
- Interest level
- Cultural relevance
- Sensitivity concerns
- Translation or adaptation problems
- Coding/Marking problems
- Priority for inclusion
Ratings of Priority for Inclusion

- On-balance judgment about suitability of the item for inclusion
  - Should combine, curricular relevance, interest, relevance to the PISA framework, cultural relevance, sensitivity issues, and technical matters
- ‘4’ shows moderately high priority and ‘5’, shows that the item is one regard as having highest priority for inclusion.
## Items with Priority for Inclusion

Ratings => 4

<table>
<thead>
<tr>
<th>Country</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>43</td>
</tr>
<tr>
<td>Greece</td>
<td>52</td>
</tr>
<tr>
<td>Austria</td>
<td>36</td>
</tr>
<tr>
<td>Hungary</td>
<td>47</td>
</tr>
<tr>
<td>Belgium: Flanders</td>
<td>20</td>
</tr>
<tr>
<td>Ireland</td>
<td>1</td>
</tr>
<tr>
<td>Belgium: French</td>
<td>33</td>
</tr>
<tr>
<td>Iceland</td>
<td>10</td>
</tr>
<tr>
<td>Brazil</td>
<td>13</td>
</tr>
<tr>
<td>Italy</td>
<td>26</td>
</tr>
<tr>
<td>Canada</td>
<td>40</td>
</tr>
<tr>
<td>Japan</td>
<td>12</td>
</tr>
<tr>
<td>Czech Rep.</td>
<td>60</td>
</tr>
<tr>
<td>Korea</td>
<td>7</td>
</tr>
<tr>
<td>Denmark</td>
<td>70</td>
</tr>
<tr>
<td>Norway</td>
<td>44</td>
</tr>
<tr>
<td>Spain</td>
<td>22</td>
</tr>
<tr>
<td>New Zealand</td>
<td>11</td>
</tr>
<tr>
<td>Finland</td>
<td>29</td>
</tr>
<tr>
<td>Portugal</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>67</td>
</tr>
<tr>
<td>Sweden</td>
<td>87</td>
</tr>
</tbody>
</table>
Rank Comparisons: Overall versus Favourites

Overall Rank

Rank Based on Favourites

Rank on favourites worse than overall rank

Rank on favourites better than overall rank

Norway 10 versus 13

Korea 3 versus 9

Dean's Lecture Sept 6 2005
Range of ranks based upon different country’s preferred sets

Low rank

High rank

Dean’s Lecture
Sept 6 2005
Poland ranked 7 on Iceland’s items
Range Of Ranks Based upon Different Country’s Preferred Sets

Hungary ranked 29 on Japan’s items
Average Reading Item Percentage Correct: Non-English versus All Items

Percent Correct on non-English

Percent Correct Overall
Change in reading rank: Non-English sourced versus English sourced items

- Ranked higher on English items
- Ranked higher on non-English items
Summary

- PISA adopts a range of strategies to ensure cultural fairness
  - An internationally set and agreed direction
  - Diverse sources for materials
  - Extensive national review
  - Extensive analyses to validate the approach
Samples- how does it happen?

- Random sample of schools selected by the consortium
- Students are randomly selected from sampled schools using PISA developed software
- Response rate standards and exclusion standards to protect results from bias
  - Troublesome countries are Netherlands, USA, UK, NZ and Australia
Samples - does it matter? (only if you are Dutch, British or Austrian)

- Netherlands excluded from PISA 2000
  - Extremely low response rate
- UK excluded from PISA 2003
  - Low response rate and non-credible change in results
- Austria had a decrease from 2000 to 2003 caused by broadening coverage of 15 year olds
  - Addition of apprenticeship students
Translation- how does it happen?

- Provision of two source versions - English and French
- Countries translate
  - Two independent translations followed by an independent reconciliation of these two versions
- International review by tri-lingual verifiers
- National finalisation based upon verification and trial data
Translation- does it matter? (only if you think these are a problem)

- True or false ➔ Real or fake
- Offer not valid for marathon sessions ➔ Offer not valid when the marathon race is on
- It does not necessarily provide proof of a crime ➔ It is not necessary to prove the accused guilty
- You will be more interesting if you are not like everyone else ➔ It will be more fun if you don’t like other people
Implementation fidelity- how does it happen?

- Heaps of tedious manuals
- Independent test administrators who are required to read an inane script
- Strict confidentiality requirements
- International and national checks on marking
- International and national checks on data credibility
Implementation fidelity- does it matter? (only if you get caught)

- The role of the forensic psychometrician
- Some of the give-aways
  - 100% agreement between markers
  - % correct from less developed countries exceed % correct of Japan or Finland
- OECD said ‘Watch country X’
Motivational issues

- Variation in student motivation across cultures is often raised as a threat
  - Usually unsupported by data
- Examples of empirical research
  - German (Baumert & Demmrich, 2001)
  - Swedish (under review)
  - Australia & Germany (Butler & Adams)
The Effort Thermometer

How much effort did you invest?
The Effort Thermometer

How much effort did you invest?

Compared to the situation you have just imagined, how much effort did you put into doing this test?
The Effort Thermometer

How much effort did you invest?

How much effort would you have invested if your marks from the test were going to be counted in your school marks?
Effort

Effort = School mark effort – PISA effort

- Illiterate idealists: negative effort
- Supporters: zero effort
- Realists: positive effort
- Cynics: extreme effort
Means for Effort

<table>
<thead>
<tr>
<th></th>
<th>Australia</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>1.74</td>
<td>2.14</td>
</tr>
<tr>
<td><strong>Note</strong>: higher scores mean lower effort</td>
<td>Significant decline in effort</td>
<td>Significant increase in effort</td>
</tr>
<tr>
<td>2003</td>
<td>1.90</td>
<td>1.85</td>
</tr>
</tbody>
</table>
Means for Effort by gender

Note: higher scores mean lower effort
Australia

2000

2003

Plausible value in Reading

Effort

idealists
supporters
diligent
realists
realists
cynics
Effort by gender for reading

Plausible value in reading

<table>
<thead>
<tr>
<th>Effort</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>idealists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>support realists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>realists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cynics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Female
- Male
Transformed mean scores for reading

Graph showing transformed mean scores for reading in different countries and years, with bars representing female, male, and overall performance. The countries and years include AUS 2000, AUS 2003, DEU 2000, and DEU 2003.
So does effort count?
Dean's Lecture
Sept 6 2005

Boys

Consistent reduction in performance after adjustment for effort

Decrease rather than increase in performance after adjustment for effort

Aus-Raw  Aus-Adjusted  Deu-Raw  Deu-Adjusted
Relatively consistent performances, both with and without adjustment.

Improvement (some, but not all) accounted for by effort.
Summary

- Effort is related to achievement but the relationship is non-linear.
- Effort is less in Germany than Australia but Germany improved its effort investment from 2000 to 2003.
- Effort is less for boys than girls.
- Effort explains some but not all of the differences between the achievement of boys and girls.
- Effort explains some but not all of the improved performance of Germany.
The PISA study reveals why German teenagers have such bad teeth.
PI SA outcomes

- Profiles of knowledge and skills
- Contextual indicators
- Trends and league tables
These students may be able to read, but have not acquired the skills to use reading for learning.

Reading Literacy Level 5
- Evaluating information and building hypotheses
- Drawing on specialised knowledge
- Accommodating concepts contrary to expectations

Reading Literacy Level 1
- Recognise main theme on a familiar topic, make simple connections

Below Level 1
- These students may be able to read, but have not acquired the skills to use reading for learning
PI SA equity outcomes

- Equity is an important focus for PI SA
  - Gender
  - Language background
  - Family structure
  - Socio-economic status
Social background is a powerful factor influencing student performance (Parental occupation, wealth, cultural resources, parental education, family structure, immigrant status).

But poor performance does not automatically follow.
Social Background and Student Performance (PISA 2000)

Student performance in PISA

PISA Index of social background

- Finland
- Japan
- UK
- France
- US
- Italy
- Germany

High performance

High

Low

Performance in PISA

Social Background and Student Performance (PISA 2000)
Equity versus Level: PISA 2003

Increased equity

Above-average level of student performance in mathematics
Above-average impact of socio-economic background

Below-average level of student performance in mathematics
Below-average impact of socio-economic background

OECD mean
Migration and student performance

- % “Non-native” students
- Score “native” students
- Score “non-native” students
- Score “first-generation” students

Performance on reading

% “First generation” students

% “First generation” students
Some General PISA outcomes

- Besides USA the literacy orientation favours English speaking countries
- Eastern European/Germanic education tradition countries are concerned by outcomes less than their expectations
- International studies reinforce the pervasiveness of gender differences in outcomes
- Relative standing of the USA declines with increasing age
And in conclusion – the real value of international studies

- Perspective on education that goes beyond the English language community
  - Research
  - Values
  - Structure
Shall I slice the pizza into four or eight pieces?

Make it four! I could never eat eight!

Thank you.