Gender differences in Variability and Extreme Scores

ECER, Berlin, 15 sept. 2012

Ariane Baye
Dominique Lafontaine
Christian Monseur

University of Liège Belgium
Gender differences in an international context

- International comparative surveys (such as IEA and PISA) of student achievement are useful tools for estimating the magnitude of the gender gap.
Gender differences in an international context

- Review of literature
  - Reading: gender gap in favour of girls, at all ages, larger among the youngest students, both in achievement and attitudes
  - Physical sciences and Math: gender gap in favour of boys, both in achievement and attitudes, increasing with the age
Gender differences in an international context

- The magnitude of the gender gap may depend on:
  1. The test content
     1. Test content: reading, reflect upon and evaluate, spatial representation, physical science
     2. Test/Item format: smaller on non-continuous texts, females better on constructed-response
  2. The study design
     1. Target population: in math, differences only in high school
     2. Type of sample: representative sample vs. particular sample
  3. The statistics performed: effects sizes / central tendency statistics
Purpose of the study

- Exploring gender differences in reading, mathematics and science since the 1990s using international large-scale survey data (IEA and PISA)

Hypothesis:

(1a) gender differences at the extreme tails of the distribution may vary substantially compared to gender mean differences

(1b) those differences will be in favour of males at the higher end of the distribution in mathematics and sciences, and to the disadvantage of males at the lower end of the reading distribution

(2a) the supposed greater variability of males will vary according to the domain

(2b) methodological choices such as the study design will influence the results
Method

- Data:
  - 6 IEA studies:
    - PIRLS 2001, 2006
  - 4 PISA cycles:

=> 1,393 cases, one case corresponding to an assessment of one domain in one population in one country
Method

- **VARIABILITY**
  - Gender Differences in Standard Deviation
    - Females SD – Males SD (negative and significant > males’ sd is significantly higher than females’ sd)
  - Gender Variance Ratio
    - Males variance/Females variance

- **EXTREME SCORES**
  - Gender differences at p. 5, 10, mean, 90, 95 with effect sizes
    - (Males mean- Female mean)/ Pooled standard deviation
## Gender differences in variability

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mal S</th>
<th>Mal NS</th>
<th>Fem NS</th>
<th>Fem S</th>
<th>VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>285</td>
<td>61%</td>
<td>34%</td>
<td>6%</td>
<td>0%</td>
<td>1.15</td>
</tr>
<tr>
<td>Math.</td>
<td>554</td>
<td>44%</td>
<td>47%</td>
<td>9%</td>
<td>0%</td>
<td>1.12</td>
</tr>
<tr>
<td>Sciences</td>
<td>554</td>
<td>48%</td>
<td>43%</td>
<td>8%</td>
<td>0%</td>
<td>1.14</td>
</tr>
<tr>
<td>Primary</td>
<td>327</td>
<td>33%</td>
<td>57%</td>
<td>9%</td>
<td>0%</td>
<td>1.11</td>
</tr>
<tr>
<td>Secondary</td>
<td>1066</td>
<td>54%</td>
<td>38%</td>
<td>8%</td>
<td>0%</td>
<td>1.14</td>
</tr>
<tr>
<td>IEA</td>
<td>779</td>
<td>37%</td>
<td>53%</td>
<td>11%</td>
<td>0%</td>
<td>1.12</td>
</tr>
<tr>
<td>OECD</td>
<td>614</td>
<td>63%</td>
<td>32%</td>
<td>4%</td>
<td>0%</td>
<td>1.16</td>
</tr>
<tr>
<td>Mean</td>
<td>1393</td>
<td>49%</td>
<td>43%</td>
<td>8%</td>
<td>0%</td>
<td>1.13</td>
</tr>
</tbody>
</table>
Results Variability

- Gender Differences in Variability
  - Standards deviations: in 92% of the 1,393 cases, male standard deviation are larger than female ones
  - The difference is statistically significant in 49% of cases
  - Females standard deviation is higher than males one in 8%, but this difference is significant for only 2 of the 1,393 cases
  - Greater SD for males varies according to the domain, education level, IEA and PISA
Results

Variability

- Gender Differences in Variability
  - Male/female variance ratio is 1.13 on average: male variance is 13% higher than female variance.
Results

Variability

• Variance:
  • Almost no exception to the higher male variance ratio
  • The differences between domains, educational levels, organisations and surveys are quite slight, except for the difference between PISA and PIRLS in reading
Results

Extreme scores

- Note: effect sizes, Cohen’s (1977) categorisation:
  - 0.20-0.50: small
  - 0.50-0.80: medium
  - More than 0.80: large
## Gender diff in extreme scores

<table>
<thead>
<tr>
<th></th>
<th>$N$</th>
<th>P5</th>
<th>P10</th>
<th>Mean</th>
<th>P90</th>
<th>P95</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td>285</td>
<td>0.45</td>
<td>0.44</td>
<td>0.34</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Math.</strong></td>
<td>554</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.06</td>
<td>-0.13</td>
<td>-0.15</td>
</tr>
<tr>
<td><strong>Sciences</strong></td>
<td>554</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.07</td>
<td>-0.15</td>
<td>-0.16</td>
</tr>
<tr>
<td><strong>Primary</strong></td>
<td>327</td>
<td>0.11</td>
<td>0.09</td>
<td>0.01</td>
<td>-0.05</td>
<td>-0.06</td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
<td>1066</td>
<td>0.12</td>
<td>0.10</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.09</td>
</tr>
<tr>
<td><strong>IEA</strong></td>
<td>779</td>
<td>0.05</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.11</td>
<td>-0.12</td>
</tr>
<tr>
<td><strong>OECD</strong></td>
<td>614</td>
<td><strong>0.21</strong></td>
<td><strong>0.20</strong></td>
<td>0.10</td>
<td>0.00</td>
<td>-0.02</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>1393</td>
<td>0.12</td>
<td>0.11</td>
<td>0.03</td>
<td>-0.05</td>
<td>-0.07</td>
</tr>
</tbody>
</table>
| **Positif Effect Size in favour of females** | &n...
Conclusion

- Extreme tails
- The gender differences at the extreme tails of the distribution are often more substantial than the gender differences at the mean, which may suggest the need to rethink education policies for low-achieving boys in reading and for high-achieving girls in mathematics and science.
- Variance
- The “greater male variability hypothesis” is confirmed, although our results suggest that this may depend on the test content, the study design and the computed statistics.
To go further

- Systematically explore the effect of the test composition in terms of format of questions
Thank you

ariane.baye@ulg.ac.be