Reporting: achievement, associations with contexts, and interpretation

Ursula Schwantner, ACER, Centre for Global Education Monitoring (GEM)
Main topics

• Purposes of large-scale assessments and reporting
• Reporting on results:
  – Student achievement
  – Contexts
  – Associations between achievement and contexts
• Interpreting results
• Challenges of evidence-based reporting
Purposes of large-scale assessments

- Evaluating and monitoring the **quality** of the education system
- Ensuring **equity**
- Exercising **accountability**

Tobin, Lietz, Nugroho, Vivekanandan, & Nyamkhuu, 2015
Purposes of reporting

- Describing outcomes (cognitive outcomes, attitudes, values, beliefs), e.g. current state, trends over time, learning growth
- Comparing different groups of students, e.g. sub-groups that may experience disadvantage
- Exploring relationships between outcomes and contexts, e.g. influence of socio-economic factors
Data analysis and reporting plan

• Description of the policy and research questions
• Description of the analysis (purpose and processes)
• Types of reports and target audience
• Timelines for analysis, report writing and release
• Resources required
Data analysis and reporting plan

- Informs the assessment design
- Identifies methodological issues
- Reveals gaps in data collection
- Provides information about the kind of findings/messages that will be possible
- Supports the development of a dissemination strategy and better use of assessment data in education policy and practice
REPORTING ON...

Achievement

Contexts

Associations between achievement and contexts
Achievement
The learning domain

- Curriculum and literacy orientation of the assessment domain

*Education that empowers students to take their abilities beyond the classroom to other theatres of daily activity, and beyond schooling to work and other areas of adult life has literacy orientation* (Turner, 2014).
Literacy orientation

• Focus of the learning domain on understanding and application of knowledge, skills and procedures

• Literacy oriented assessments assess the domain in its complexity

Turner, 2014
Numeric scales
Scaling

• Scaling models are driven by different theories.
• Most commonly used in large-scale assessments of learning are scaling models based on Item Response Theory (IRT). Less often scaling models are based on Classical Test Theory (CTT).
Described proficiency scales

• One of the most powerful ways to report student achievement is through a scale which describes comparative abilities.

Turner, 2014
Described proficiency scales

Students are typically able to demonstrate the skills at and below their ability level.

On average, boys in ‘Region X’ are performing at level 8 on this scale.

On average, students in ‘Region X’ are performing at level 7 on this scale.

On average, girls in ‘Region X’ are performing at level 6 on this scale.

Set up equation and solve it in a real life situation. Explain the information shown in a complex graph.

Solve word problems requiring two mathematical processes. Calculate the length and area of parts of a circle. Solve algebraic equations where two or more steps are required.

Use percentages and ratios to solve problems. Convert units of measurement for area and volume. Understand information from a statistical graph with grouped data.

Find missing angles in shapes. Understand the order of mathematical operations. Calculate the volume and surface area of standard 3D objects (cuboid, cylinder).

Solve simple word problems. Distinguish between simple shapes. Find the value of a simple algebraic expression. Write ratios using small numbers in their simplest form.
Level 2 is considered the baseline level of mathematical proficiency that is required to participate fully in modern Society (OECD, 2014, p. 68).
Ways of reporting achievement

• Averages and distributions
• Against national or international performance standards or benchmarks
• In the learning domain, in sub-domains, or item-by-item
• For subgroups of students (e.g. gender, socio-economic background, language background)
• Against other background variables

Forster, 2001
Mathematics PISA 2012: means for OECD-/EU-countries

- No sig. difference to AUT
- sig. higher than AUT
- sig. lower than AUT

Schwantner, Toferer & Schreiner, 2013
Means and standard deviation for 65 participating countries (PISA 2012)

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>QCN</td>
<td>613</td>
<td>101</td>
</tr>
<tr>
<td>SGP</td>
<td>573</td>
<td>105</td>
</tr>
<tr>
<td>HKG</td>
<td>561</td>
<td>96</td>
</tr>
<tr>
<td>TWN</td>
<td>560</td>
<td>116</td>
</tr>
<tr>
<td>KOR</td>
<td>554</td>
<td>99</td>
</tr>
<tr>
<td>MAC</td>
<td>538</td>
<td>95</td>
</tr>
<tr>
<td>JPN</td>
<td>536</td>
<td>94</td>
</tr>
<tr>
<td>LIE</td>
<td>535</td>
<td>95</td>
</tr>
<tr>
<td>CHE</td>
<td>531</td>
<td>94</td>
</tr>
<tr>
<td>NLD</td>
<td>523</td>
<td>92</td>
</tr>
<tr>
<td>EST</td>
<td>521</td>
<td>81</td>
</tr>
<tr>
<td>FIN</td>
<td>519</td>
<td>85</td>
</tr>
<tr>
<td>CAN</td>
<td>518</td>
<td>89</td>
</tr>
<tr>
<td>POL</td>
<td>518</td>
<td>90</td>
</tr>
<tr>
<td>BEL</td>
<td>515</td>
<td>102</td>
</tr>
<tr>
<td>DEU</td>
<td>514</td>
<td>96</td>
</tr>
<tr>
<td>VNM</td>
<td>511</td>
<td>86</td>
</tr>
<tr>
<td>SVN</td>
<td>501</td>
<td>92</td>
</tr>
<tr>
<td>NLD</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>DNK</td>
<td>500</td>
<td>82</td>
</tr>
<tr>
<td>CZE</td>
<td>499</td>
<td>95</td>
</tr>
<tr>
<td>FRA</td>
<td>495</td>
<td>97</td>
</tr>
<tr>
<td>GBR</td>
<td>494</td>
<td>95</td>
</tr>
<tr>
<td>ISL</td>
<td>493</td>
<td>92</td>
</tr>
<tr>
<td>LVA</td>
<td>491</td>
<td>82</td>
</tr>
<tr>
<td>LUX</td>
<td>490</td>
<td>95</td>
</tr>
<tr>
<td>AUS</td>
<td>504</td>
<td>96</td>
</tr>
<tr>
<td>IRL</td>
<td>501</td>
<td>85</td>
</tr>
<tr>
<td>NOR</td>
<td>489</td>
<td>90</td>
</tr>
<tr>
<td>PRT</td>
<td>487</td>
<td>94</td>
</tr>
<tr>
<td>ITA</td>
<td>485</td>
<td>93</td>
</tr>
<tr>
<td>ESP</td>
<td>484</td>
<td>88</td>
</tr>
<tr>
<td>RUS</td>
<td>482</td>
<td>86</td>
</tr>
<tr>
<td>SVK</td>
<td>482</td>
<td>101</td>
</tr>
<tr>
<td>USA</td>
<td>481</td>
<td>90</td>
</tr>
<tr>
<td>LTU</td>
<td>479</td>
<td>89</td>
</tr>
<tr>
<td>CYP</td>
<td>440</td>
<td>93</td>
</tr>
<tr>
<td>BGR</td>
<td>439</td>
<td>94</td>
</tr>
<tr>
<td>ARE</td>
<td>434</td>
<td>90</td>
</tr>
<tr>
<td>KAZ</td>
<td>432</td>
<td>71</td>
</tr>
<tr>
<td>THA</td>
<td>427</td>
<td>82</td>
</tr>
<tr>
<td>CHL</td>
<td>423</td>
<td>81</td>
</tr>
<tr>
<td>MYS</td>
<td>421</td>
<td>81</td>
</tr>
<tr>
<td>JOR</td>
<td>386</td>
<td>78</td>
</tr>
<tr>
<td>QAT</td>
<td>376</td>
<td>100</td>
</tr>
<tr>
<td>COL</td>
<td>376</td>
<td>74</td>
</tr>
<tr>
<td>IDN</td>
<td>375</td>
<td>71</td>
</tr>
<tr>
<td>PER</td>
<td>368</td>
<td>84</td>
</tr>
</tbody>
</table>
Mathematics processes PISA 2012: Means for OECD-/EU-countries

wantner, Toferer & Schreiner, 2013
Distribution of proficiency levels in mathematics (PISA 2012)

Schwartner, Toferer & Schreiner, 2013
Exhibit 4: Distribution of Class 6 mathematical proficiency
Overlapping of top performers in mathematics, reading and science on average across OECD countries

- Reading only 1.9%
- Reading and science 0.6%
- Mathematics only 4.4%
- Mathematics and science 2.3%
- Science only 1.1%
- Reading, mathematics and science 4.4%
- Mathematics and reading 1.5%

Note: Non-top performers in any of the three domains: 83.8%.
Source: OECD, PISA 2012 Database, Table l.2.29.
Mathematical literacy

Exhibit 10 is a description of the proficiency scale for mathematics. Examples are items from the Class 6 assessment. Exhibit 11 presents the scale with illustrated items.

Exhibit 10: Proficiency descriptions for mathematics (Class 6)

<table>
<thead>
<tr>
<th>Level and examples</th>
<th>Proficiency description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 11 and above</strong> (259 and above)</td>
<td></td>
</tr>
<tr>
<td>e.g. Teapot (full credit)</td>
<td></td>
</tr>
<tr>
<td>12-sided shape (full credit)³</td>
<td></td>
</tr>
<tr>
<td>Class 6 students at this level: 0%</td>
<td></td>
</tr>
<tr>
<td><strong>Level 10</strong> (242 to less than 250)</td>
<td></td>
</tr>
<tr>
<td>e.g. Buying walnuts, Carpet turn</td>
<td></td>
</tr>
<tr>
<td>Class 6 students at this level: 1%</td>
<td></td>
</tr>
<tr>
<td><strong>Level 6</strong> (178 to less than 194)</td>
<td></td>
</tr>
<tr>
<td>e.g. Bales of cotton</td>
<td></td>
</tr>
<tr>
<td>Class 6 students at this level: 23%</td>
<td></td>
</tr>
</tbody>
</table>

Students at this level can typically recognise common shape names, and they can use spatial reasoning as part of a counting strategy or to make comparisons involving mathematical properties of objects.

Lumley et al., 2015
Level 6 – 178 to less than 194

Students at this level are performing below the average level achieved by students in their cohort.

86% of Class 6 students demonstrated these and more advanced skills.

They are typically able to recognise common shape names. They can interpret images of familiar objects and use spatial reasoning, for example to devise and apply a counting strategy for stacked objects. They can interpret graphical representations and use them to make comparisons, for example involving mathematical properties such as length.
Exhibit 22: Bales of Cotton

This is one bale of cotton.  
Some bales are stacked on a truck.

The item *Bales of Cotton* was one of the easiest items used in the Class 6 assessment. It presents a visual image of a collection of stacked cuboids. Bales of cotton may not be familiar to all students, but certainly stacks of regular-shaped cuboids do occur very frequently in different contexts (for example, stacks of bricks, or children’s blocks, or apartment buildings). Solving this problem involves spatial reasoning to interpret the graphic image, and devising a counting strategy to calculate $3 \times 2 \times 2$ (or its equivalent).

How many bales are there on the truck?

A. 10  
B. 11  
C. 12  
D. 16  

Key: C  
Difficulty: 181 (Level 6)
Suggestions for teaching

*Bales of Cotton* (Exhibit 22) provides an opportunity to explore different counting and calculation strategies in a simple practical context, which could be explored and extended to different numbers and types of regularly shaped objects. Start by asking students how they solved the problem. Possibilities could include direct counting (for example counting 6 bales on the top layer, and adding another 6 because there are two layers, or counting 4 bales at the back, and adding another 2 lots for the other layers, or counting 6 bales seen on the side, and doubling this), or seeing the three-dimensional multiplicative array involved.

This task involves recognising one of the symbolic mathematical expressions that could be used to represent these numbers. It could be a starting point for discussing different possible counting methods, and from there to explore different but equivalent symbolic expressions (for example $6 + 6$ is equivalent to $4 + 4 + 4$; $3 \times 2 \times 2$ is equivalent to $2 \times 3 \times 2$; $6 \times 2$ is equivalent to $4 \times 3$). In a further step, the task could be used to encourage students to move from a strategy that involves additive reasoning to a strategy that involves multiplicative thinking.
Ways of reporting trends

• Changes in averages and distributions
• Changes in percentage of proficiency levels, of students and above or below benchmarks
• Changes in the achievement of subgroups of students (including relative growth of subgroups)

Forster, 2001
Contexts
Why collect contextual data?

- To investigate and describe associations with achievement (differences between subgroups, relationships and effects on achievement)
- To describe and monitor important outcomes of schooling, in addition to achievement
- To monitor processes and practices
- To monitor resources
Contextual factors

- **Individual and family factors**, e.g. gender, grade, socioeconomic factors, migration background, etc.

- **Process factors**, e.g. engagement and motivation, attitudes and beliefs, learning strategies, quality of instruction, etc.

- **School factors**, e.g. school location, school authority, school size, resources, management and leadership, school climate, etc.
Associations between achievement and contexts
Differences between subgroups

**STEP 1:** ASSEMBLE PART A TO PART B. **STEP 2:** GLUE THESE PIECES SECURELY. **STEP 3:** FIND PART C AND CONNECT TO PART D...
Mean differences

(PIRLS2011)

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th></th>
<th>Male</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>WA</td>
<td>533</td>
<td>6.0</td>
<td>500</td>
<td>5.8</td>
</tr>
<tr>
<td>NT</td>
<td>522</td>
<td>11.3</td>
<td>498</td>
<td>11.2</td>
</tr>
<tr>
<td>TAS</td>
<td>536</td>
<td>9.6</td>
<td>514</td>
<td>8.6</td>
</tr>
<tr>
<td>QLD</td>
<td>521</td>
<td>4.3</td>
<td>502</td>
<td>7.5</td>
</tr>
<tr>
<td>VIC</td>
<td>548</td>
<td>6.0</td>
<td>531</td>
<td>3.6</td>
</tr>
<tr>
<td>NSW</td>
<td>542</td>
<td>6.8</td>
<td>528</td>
<td>4.8</td>
</tr>
<tr>
<td>ACT</td>
<td>561</td>
<td>5.6</td>
<td>556</td>
<td>6.0</td>
</tr>
<tr>
<td>SA</td>
<td>520</td>
<td>5.9</td>
<td>516</td>
<td>5.4</td>
</tr>
<tr>
<td>International average</td>
<td>520</td>
<td>0.5</td>
<td>514</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Figure 2.7** Gender differences in reading achievement, by state

Thomson et al., 2012
Differences in performance distribution

*Figure 2.4* Distribution of reading achievement within Australia, by gender

(PIRLS2011)
Differences in performance distribution

<table>
<thead>
<tr>
<th></th>
<th>Below Low</th>
<th>Low</th>
<th>Intermediate</th>
<th>High</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>5</td>
<td>16</td>
<td>34</td>
<td>33</td>
<td>12</td>
</tr>
<tr>
<td>Males</td>
<td>9</td>
<td>19</td>
<td>34</td>
<td>31</td>
<td>8</td>
</tr>
</tbody>
</table>

*Figure 2.5* Percentages of Australian students at the international benchmarks for reading, by gender (PIRLS2011)
Differences between subgroups

A greater proportion of girls than boys performed in the highest writing proficiency levels. The highest level, Level 10, comprises more than three times as many girls as boys: 14% of girls but only 4% of boys.

Writing: Highest proficiency level (Level 10)

14%  4%
Differences between subgroups

PISA 2012 Mathematics

Austria total

Low performers

High performers

Schwantner, Toferer & Schreiner, 2013
Relationships and effects

• *Correlation:* extent of linear relationship between two variables

• *Regression analysis:* predictive modelling technique to predict an outcome variable, using one or more explanatory variables

• *Multilevel analysis* (hierarchical linear model): potential effects of clustering in a complex sample design (e.g. students within classes, within schools)
Correlation vs regression

- Correlation is *not* directional. The degree of association goes both ways.
- Correlation is *not* appropriate to explain one variable with the other.

-> Use regression instead.
Regression analysis

• Describes a relationship, but assumes a direction:
  – Relationship between a dependent variable, and one or more independent variables

• Explains one variable with one (or more) other variable(s)
  – How well does SES predict performance?
Regression analysis

• Two main statistics:
  – Size of the effect or slope
  – Strength of the effect or explained variance
Regression analysis

Schwantner & Schreiner, 2010

<table>
<thead>
<tr>
<th>Country</th>
<th>R2</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUN</td>
<td>19,2%</td>
<td>25,0</td>
</tr>
<tr>
<td>BEL</td>
<td>18,9%</td>
<td>24,9</td>
</tr>
<tr>
<td>LUX</td>
<td>17,4%</td>
<td>24,9</td>
</tr>
<tr>
<td>AUT</td>
<td>14,2%</td>
<td>22,5</td>
</tr>
<tr>
<td>CZE</td>
<td>14,0%</td>
<td>25,5</td>
</tr>
<tr>
<td>SVN</td>
<td>13,2%</td>
<td>19,5</td>
</tr>
<tr>
<td>DEU</td>
<td>12,5%</td>
<td>21,2</td>
</tr>
<tr>
<td>SVK</td>
<td>12,0%</td>
<td>21,8</td>
</tr>
<tr>
<td>CHE</td>
<td>12,0%</td>
<td>20,0</td>
</tr>
<tr>
<td>GBR</td>
<td>11,5%</td>
<td>20,0</td>
</tr>
<tr>
<td>NLD</td>
<td>11,1%</td>
<td>18,9</td>
</tr>
<tr>
<td>ITA</td>
<td>10,7%</td>
<td>18,9</td>
</tr>
<tr>
<td>DNK</td>
<td>10,6%</td>
<td>16,2</td>
</tr>
<tr>
<td>IRL</td>
<td>10,0%</td>
<td>18,0</td>
</tr>
<tr>
<td>SWE</td>
<td>9,9%</td>
<td>18,5</td>
</tr>
<tr>
<td>LIE</td>
<td>7,4%</td>
<td>13,6</td>
</tr>
<tr>
<td>FIN</td>
<td>5,2%</td>
<td>12,3</td>
</tr>
</tbody>
</table>
Multilevel analysis

Relationship between school performance and schools’ socio-economic profile
Interpretation
Provide background information

WHY Aim, purpose and context of the assessment
WHO is responsible for what; organisation
WHEN Assessment cycle, timeframe
WHAT Assessment framework and instruments (cognitive and contextual), proficiency scales and levels
HOW Methodology and procedures
Specific information about:

• Sample and groups of comparison (national, regional or international)

• Scales, scores and proficiency levels

• Contextual indices and scales
• Methodological-statistical explanations, e.g.
  – Rounding of figures
  – Standard errors and confidence intervals
  – Statistical significance
  – Comparisons over time

See for example Thomson et al., 2012; Lumley et al., 2015
Remember to provide

• Standard error (and confidence interval) for a population estimate (e.g. mean performance)

• Tests (indication) of statistical significance

• Consider practical significance: effect size
Challenges of evidence-based reporting

Source: google.com
Beware of spurious correlations

• Just because two variables correlate highly does not mean there is a valid relationship between them.

• Correlation is *not* causation.

• To investigate directions of relationships, the effect of an independent variable, or to control for confounding factors, use regression analysis.
Complex models and analysis

• Developing and analysing complex models requires careful preparation in order to be meaningful.
  – Move more complex analysis to an in-depth thematic report.
Acknowledge limitations

- Be careful about over-interpreting the results.
- What can the assessment tell us? What can’t it tell us?
- Conclusions drawn from the assessment should be based on the reported findings.
References


Activity: Interpretation

• Form groups

• Choose one of the three topics:
  – Who are the low-performing students?
  – What lies behind gender inequality in education?
  – Do teacher-student relations affect students’ well-being at school?

• Activity sheet
Activity: Interpretation

• Choose one of the three topics:
  – Who are the low-performing students?
  – What lies behind gender inequality in education?
  – Do teacher-student relations affect students’ well-being at school?

• Activity sheet
Activity reflection

• Which article and graphs did your group choose?
• Which analysis techniques did your group identify?
• What are the main findings your group derived from the interpretation of the graphs?
• Are the main messages in the articles clearly backed up by evidence? What additional analyses could be included to strengthen the evidence provided?