Lessons learned from PIRLS and PISA
What every teacher should know about International Large Scale Assessment

Department of Education, University of Oxford
Therese N. Hopfenbeck
Oxford University Centre for Educational Assessment

International Conference on Educational Measurement, Evaluation and Assessment,
Abu Dhabi, November 5th, 2017
Assessment literacy

• What do you need to know as a teacher regarding large scale assessment?
• What is the differences between PISA and PIRLS?
• What is the purpose of the assessment?
• How are they used by policy makers and researchers?
• What can they offer teachers?
What is PISA?

• *Programme for International Student Assessment* (PISA) seen by many as having strategic prominence in international education policy debates (Meyer and Benavot 2013)

• PISA was launched by the Organisation for Economic Co-operation and Development (OECD) in 1999 with the aim to assess “aspects of preparedness for adult life” (OECD 2000:3).
About PISA

The Programme for International Student Assessment (PISA) is a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. To date, students representing more than 70 economies have participated in the assessment.

The most recently published results are from the assessment in 2012.


Consult all PISA 2012 results here.

More than 70 economies have signed up to take part in the assessment in 2015 which will focus on science.

Download the PISA trifold brochure.

What makes PISA different

What the assessment involves
## Timeline: major/minor literacy emphasis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading literacy</strong></td>
<td>Major</td>
<td>Minor</td>
<td>Minor</td>
<td>Major</td>
<td>Minor</td>
<td>Minor</td>
</tr>
<tr>
<td><strong>Mathematical literacy</strong></td>
<td>Minor</td>
<td>Major</td>
<td>Minor</td>
<td>Minor</td>
<td>Major</td>
<td>Minor</td>
</tr>
<tr>
<td><strong>Scientific literacy</strong></td>
<td>Minor</td>
<td>Minor</td>
<td>Major</td>
<td>Minor</td>
<td>Minor</td>
<td>Major</td>
</tr>
</tbody>
</table>
Unlike the International Association for the Evaluation of Educational Achievement (IEA)’s *Progress in International Reading Literacy Study (PIRLS)*, the test content is independent of the participating countries’ school curricula, with a focus upon assessing whether 15-year-olds are able to *apply* what they have learned in school in real life situations by the time they have finished their compulsory schooling (OECD website).
PISA in the world

2000: 32 countries

2012: 65 countries (First time UAE participated)

2015: 71 countries
Germany

Key facts
With a population of 82 million, Germany is the largest country in the European Union. It has a strong literary tradition and is Europe’s strongest economy. It also has a relatively high immigrant population, with more than 7 million people, or nearly 9% of the total population, defined as coming from a non-German immigrant background.

Expanding educational opportunity to improve outcomes for every student
When the first PISA tests in 2000 placed German students well below the average in OECD countries for reading and literacy, the nation was shocked. The revelation sparked a nationwide debate about Germany’s school system and what needed to improve it.

Published in 2001, the PISA 2000 results revealed that the German system was not providing equal opportunities for all. Students from disadvantaged backgrounds were particularly at risk. Rather than basing the choice of secondary school solely on student...
Results from International assessments such as Programme for International Student Assessment (PISA) are used by politicians around the world in the policy discussion on education and assessment.

Pisa results support case for reform, says Gove

3 December 2013 Last updated at 14:03

Figures showing that the UK is falling behind global rivals in international education league tables underlines the need for further reform of the country's education system, Education Secretary Michael Gove has argued.

For the first time that the UK has not been in the top 20 for any of the subjects measured by the international Pisa tests.

"Unless we can provide [children] with a school system that is one of the best in the world, we will not give them the opportunities they need to flourish and succeed.

"That is why it is so important today that we have a unified national commitment to excellence in all our schools for all our pupils," he said in a statement on 3 December 2013.

Mr Gove said the coalition's reforms are improving standards in schools, by drawing on what happens in the best-performing best school systems identified by the OECD.

The Pisa tests, administered by the OECD think-tank and run every three years, measure attainment in literacy, maths and science by 15 year-olds in 65 countries.
... what’s become increasingly clear is that our children’s education has been suffering in relation to their peers over the last decade. The PISA rankings, for example, which I’m sure have already been debated today, show us falling from fourth to sixteenth in science, from seventh to 25th in literacy, and from eighth to 28th in maths.

Nick Gibb, Minister for Schools in England, Speech to the 100 Group
10 February 2011 [emphasis added]
Results from the Programme for International Student Assessment (Pisa) released on Tuesday also show that UAE students continue to fall below the OECD average in science, reading and maths.

Andreas Schleicher, director of the OECD’s directorate of education and skills, says it is still ‘too early’ to see the results of the UAE’s education reforms. Christopher Pike / The National

ABU DHABI // The UAE has held on to its top ranking for the region in a survey of pupils’ proficiency, but experienced a drop in its global placings for two subjects.

Results from the Programme for International Student Assessment (Pisa) released on Tuesday also show that UAE pupils continue to fall below the Organisation for Economic Co-operation and Development average in the three subjects tested – science, reading and maths.

More than 500,000 pupils aged 15 from 35 OECD countries and 37 partner countries and economies, including 15,000 from UAE public and private schools, took part in the Pisa assessment, which is issued every three years.

The UAE’s National Agenda calls for it to rank among the top 20 in Pisa by 2021.
Pisa tests: UK stagnates as Shanghai tops league table

By Sean Coughlan BBC News education correspondent

Pisa maths scores for selected education systems

<table>
<thead>
<tr>
<th>Country</th>
<th>Rank</th>
<th>Score</th>
<th>Average score: 494</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Shanghai</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Hong Kong</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnam</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UK</strong></td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK breakdown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. Ireland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wales</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*China does not participate as a country, but is represented by cities such as Shanghai and Hong Kong.

The UK is falling behind global rivals in international tests taken by 15-year-olds, failing to make the top 20 in maths, reading and science.
International tests influence learning based upon three processes

1) what counts as valuable learning
2) how national assessment systems are developed around the world
3) how students approach learning since there is evidence that students adopt their learning approaches according to the tests given

(Baird et al 2013)
International and national assessment systems

1) Norway: Introduced national tests in 2004. The reading tests are based upon the PISA reading framework (Frones et al 2012).

2) Denmark: Introduced national tests after low performing in PISA (Egelund, 2008).

3) Japan: changed item format on their national tests to more open-responses like those in PISA (Schleicher, 2009).

4) Korea: PISA like tasks on their University Entrance Exam (Schleicher 2009).

5) Germany: introduction of national educational standards and more focus upon external assessment (Ertl, 2006)
Semmelweis’ Diary Text 1

‘July 1846. Next week I will take up a position as “Herr Doktor” at the First Ward of the maternity clinic of the Vienna General Hospital. I was frightened when I heard about the percentage of patients who die in this clinic. This month not less than 36 of the 208 mothers died there, all from puerperal fever. Giving birth to a child is as dangerous as first-degree pneumonia.’

These lines from the diary of Ignaz Semmelweis (1818-1865) illustrate the devastating effects of puerperal fever, a contagious disease that killed many women after childbirth. Semmelweis collected data about the number of deaths from puerperal fever in both the First and the Second Wards (see diagram).
Physicians, among them Semmelweis, were completely in the dark about the cause of puerperal fever. Semmelweis’ diary again:

‘December 1846. Why do so many women die from this fever after giving birth without any problems? For centuries science has told us that it is an invisible epidemic that kills mothers. Causes may be changes in the air or some extraterrestrial influence or a movement of the earth itself, an earthquake.’

Nowadays not many people would consider extraterrestrial influence or an earthquake as possible causes of fever. We now know it has to do with hygienic conditions. But in the time Semmelweis lived, many people, even scientists, did! However, Semmelweis knew that it was unlikely that fever could be caused by extraterrestrial influence or an earthquake. He pointed at the data he collected (see diagram) and used this to try to persuade his colleagues.

**QUESTION 1.1**

Suppose you were Semmelweis. Give a reason (based on the data Semmelweis collected) why puerperal fever is unlikely to be caused by earthquakes.
Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment

Therese N. Hopfenbeck, Jenny Lenkeit, Yasmine El Masri, Kate Cantrell, Jeanne Ryan and Jo-Anne Baird

Department of Education, Oxford University Centre for Educational Assessment, University of Oxford, Oxford, UK

Abstract

International large-scale assessments are on the rise, with the Programme for International Student Assessment (PISA) seen by many as having strategic prominence in education policy debates. The present article reviews PISA-related English-language peer-reviewed articles from the programme’s first cycle in 2000 to its most current in 2015. Five literature bases were searched, and results were analysed with SPSS. Results map the frequency of publications according to journal, country, and scientific discipline. They also summarise major themes within three identified categories: secondary analysis, policy impact, and critiques. Findings indicated that studies based on the PISA dataset has led to progress in educational research while simultaneously pointing to the need for caution when using this research to inform educational policy.
Reviews PISA-related English-language peer-reviewed articles from the programme’s first cycle in 2000 to its most current in 2015.

They also summarise major themes within three identified categories; secondary analysis, policy impact and critiques.

Findings indicated that studies based on the PISA dataset has led to progress in educational research while simultaneously pointing to the need for caution when using this research to inform educational policy.
Offer an assessment of how PISA

(i) has affected the research literature so far,
(ii) advanced research around educational topics in national and cross-national contexts,
(iii) initiated intense discourse around technical aspects of measurements in international comparisons and
(iv) forged scholarly discussion around the impact of global actors on educational policies.
Step one:
• electronic search was conducted in five scientific databases: ERIC, PsycINFO, Scopus, Web of Science and Zetoc,
• The keywords employed in the search were ‘PISA’ and ‘Programme for International Student Assessment’. The search terms were set to all fields (title, abstract and keywords etc.) and the date range spanned a period of 15 years, from January 1999 to September 2015
• Software EndNote X7

Step two:
• initial search was repeated but with two digital search engines, Google Scholar and SOLO (Oxford University’s library catalogue).
Initial count

• an initial count of database items revealed a total of 1,001 publications:

  910 journal articles
  46 reports
  24 conference papers
  14 book chapters
  3 book reviews
  3 newspaper articles
  1 editorial

Removed articles with standard phases such as: In such a paper, a standard phrase might include a popular maxim, such as ‘In recent years, international studies of educational achievements, such as PISA, have become important benchmarks for policy makers.’
Final analytical sample of articles consists of 654 original peer-reviewed articles published in peer-reviewed journals reporting PISA related research.
Results: Rise in publications

While only one article per year was published from 1999 to 2001, seven articles were published in 2002 following the release of PISA’s first cycle results in December 2001.

Highest number in 2014 with a total of 103 articles.
Figure 1. Publications per year.
The thematic focus includes:

1. **Secondary data analysis**: research making use of the PISA database to address a specific research question.

2. **Critique of PISA**: research that either conceptually evaluates the developed constructs and those elaborated in the background questionnaire or that focus on statistical approaches and procedures applied to scales and constructs.

3. **Impact or policy study**: research discussing the role of PISA in monitoring a country’s performance in comparison to others or evaluates the impact of PISA national and international educational policies.
Themes in secondary data analysis

- socioeconomic characteristics
- teacher characteristics and instruction practices
- immigration/language
- bullying
- systemic characteristics
- Age
- ICT use
- affective characteristics
- leadership
- learning strategies
- gender
- ability grouping
- class size
- environment
- health
- a mixture of variables
Themes in critique of PISA

1. construct
2. data processing
3. questionnaires
4. scaling procedures/IRT
5. sample
6. translation/language
7. interpretation of results
8. item features
Specific themes in impact or policy study

1. country performance
2. curriculum
3. national assessment
4. policy
5. governance and media
Top 5 journals by number of publications

1. Educational Research and Evaluation (21)
2. The International Journal of Science Education (20)
4. The International Journal of Science and Mathematics Education (18)
5. Teachers College Record (17)
PISA articles across countries and disciplines

United States (114)
Australia (72)
Germany (69)
United Kingdom (52)
Ireland (31)
A substantial number (109) of the articles classified as secondary data analysis (430 articles in this category in total) explored *educational inequalities* related to *socioeconomic status* (Marks, 2006), *migration or language use* (Shapira, 2012), *gender* (Lietz, 2006) or *ability grouping* (Chmielewski, Dumont, and Trautwein, 2013). Of those, 52 specifically investigated inequalities related to SES disparities at student, school, community, or system level.
1. Caro et al. (2013) raised concerns about the differential meaning of the cultural, social and economic constructs across the countries making cross-cultural comparisons difficult to establish and to some degree invalid.

2. Hopfenbeck and Maul (2011) used cognitive interviews and statistical methods to find that the scales of self-report questionnaires in Norway led to invalid responses largely due to poor questionnaire design and language ambiguity.

3. Rutkowski and Rutkowski (2010) argued that background questionnaires, namely the ones administered in non-OECD countries, suffered a number of weaknesses such as missing data, possible respondent misinterpretation and low reliability on some of the scales. Poor questionnaire design has great implications on the validity of results given that background questionnaires are an integral part of development of achievement scores of the assessed samples. Policy makers should be aware of the limited interpretation that can be made based on flawed data resulting from weak questionnaires (Rutkowski and Rutkowski 2010; Ercikan et al. 2015).
Comparable across Countries?

PISA instruments are more comparable across Western countries than they are across Middle Eastern or Asian countries (Grisay and Monseur 2007; Grisay et al. 2007; Grisay et al. 2009; Kankaraš and Moors 2014). Empirical studies suggest that as target languages become less related to the source languages, the probability of bias in the adapted versions increases (Grisay et al., 2009). Hence, the German version would result in less bias than would the Arabic or Chinese version because of the proximity of the German language to the English language.
PIRLS 2016 ASSESSMENT FRAMEWORK, 2ND EDITION


The PIRLS 2016 Assessment Framework, 2nd Edition provides the foundation for the three international assessments planned as part of IEA’s PIRLS 2016: PIRLS, PIRLS Literacy, and ePIRLS. PIRLS represents the international standard for reading comprehension at the fourth grade. PIRLS Literacy, a less difficult version of PIRLS, is designed to assess foundational reading skills that are prerequisites for success on PIRLS. New in 2016, ePIRLS is an innovative assessment of online reading that makes it possible for countries to understand how successful the are in preparing fourth grade students to read, comprehend, and interpret online information.

Chapter 1 presents the PIRLS 2016 Reading Framework, and describes in some detail the major purposes of reading and processes of comprehension to be tested at the fourth grade. PIRLS assesses two purposes of reading that fourth grade students typically engage in: reading for literary experience, as well as to acquire and use information. PIRLS also assesses four broad processes of comprehension predominantly used by fourth grade readers: focus on and retrieve explicitly stated information, make straightforward inferences, interpret and integrate ideas and information, and evaluate and critique content and textual elements. Chapter 2 contains the PIRLS 2016 Contextual Framework describing the types of schooling situations and factors associated with students’ development in reading literacy that will be investigated via the PIRLS 2016 Encyclopedia in addition to the questionnaires completed by students, their teachers, and schools. Finally, Chapter 3 provides an overview of the PIRLS and PIRLS Literacy test booklet and ePIRLS task designs, including general parameters for item development.

Publications are available in Portable Document Format (PDF). You may need to download a copy of Adobe Reader, to enable you to read and print the report.
PIRLS

- Progress for International Reading Literacy Study
- Conducted by the IEA (International Association for the Evaluation of Educational Achievement)
- Assessment of reading, but also information about home and school environments through questionnaires and language spoken at home.
PIRLS

• **Background information**
  • Pupil characteristics, e.g. motivation to read, parents’ interest in school work
  • Teacher and classroom characteristics, e.g. formal education, career satisfaction
  • School and head teacher characteristics, e.g. school policies and resources
  • Curriculum features for international comparisons, e.g. policies on ability grouping, integration of reading in the curriculum
Reading comprehension processes

- Questions are based on texts presented to pupils
- Each question deals with one of the processes

Interpret and integrate ideas and information 30%
Focus on and retrieve explicitly stated information 20%
Evaluate and critique content and textual elements 20%
Make straightforward inferences 30%

Reading for literary experience 50%
Reading to acquire and use information 50%
Enemy Pie
by Derek Munson

illustrated by Tara Calahan

It was a perfect summer until Jeremy Ross moved in right next door to my best friend Stanley. I did not like Jeremy. He had a party and I wasn’t even invited. But my best friend Stanley was. I never had an enemy until Jeremy moved into the neighborhood. Dad told me that when he was my age, he had enemies, too. But he knew of a way to get rid of them.

Dad pulled a worn-out scrap of paper from a recipe book.

"Enemy Pie," he said, satisfied.

You may be wondering what exactly is in Enemy Pie.

Dad said "the recipe was so secret," he couldn't even tell me.

I begged him to tell me something—anything.

"I will tell you this, Tom," he said to me. "Enemy
1. Who is telling the story?
   A. Jeremy
   B. Dad
   C. Stanley
   D. Tom

2. At the beginning of the story, why did Tom think Jeremy was his enemy?
   
   [Blank answer space]
PIRLS for Teachers project: Milestones

Workshop 1

On 1 July 2015, the first workshop of the PIRLS for Teachers project took place. Eight teachers from different schools in the south of England visited the Department of Education to work with the OUCEA team on how we can make PIRLS and its findings more accessible and useful for practitioners.
How using a high variety of reading resources and genres – 7 to 8 different ones – benefits motivation towards reading for different groups of pupils

Results
While an important goal in primary school years is to teach pupils how to comprehend different text forms and genres (see English programmes of study: key stages 1 and 2), 1 in 3 teachers in England use a low or medium variety of reading resources in their lessons. Our research suggests that using a high variety of reading resources can:

- enhance motivation towards reading for pupils of all reading abilities. It is especially important for lower ability pupils, as it fosters their motivation, engagement and confidence to read.
- benefit girls’ reading confidence and boys’ motivation to read.

Contact: Jenny Lenkeit (Research Fellow) jenny.lenkeit@education.ox.ac.uk; Therese N. Hopfenbeck (Associate Professor) therese.hopfenbeck@education.ox.ac.uk
PIRLS for Teachers website: http://ouce.aeducation.ox.ac.uk/research/recent-research-projects/pirls-for-teachers/
Questions or feedback?

therese.hopfenbeck@education.ox.ac.uk

@TNHopfenbeeck
References


Baird, J.-A, Caro, D.H. & Hopfenbeck, T.N. (2016) Student Perceptions of Predictability of Examination Requirements and Relationship with Outcomes in High-Stakes Tests in Ireland, Irish Educational Studies (online). Click here to download – the first 50 downloads are free.


Elwood, J., Hopfenbeck, T.N. & Baird, J. (2015 online) Predictability in high-stakes examinations: students’ perspectives on a perennial assessment dilemma, Research Papers in Education. Click here to download – the first 50 downloads are free.
The Norwegian case

• Strong claims in the media that the Norwegian students would not be motivated to do their best on PISA -

• Are Norwegian students’ less motivated?
Examining Evidence for the Validity of PISA Learning Strategy Scales Based on Student Response Processes

Therese N. Hopfenbeck and Andrew Maul
Unit for Quantitative Analysis in Education, University of Oslo, Norway

The aim of this study was to investigate response-process based evidence for the validity of the Programme for International Student Assessment’s (PISA) self-report questionnaire scales as measures of specific psychological constructs, with a focus on scales meant to measure inclination toward specific learning strategies. Cognitive interviews (N = 22) and descriptive and statistical analysis of Norwegian questionnaire data (N = 4279) provide converging evidence that (a) a non-trivial portion of students may not be providing responses to items that can be considered valid, for a variety of reasons; (b) students who perform more poorly on PISA’s science scales are less likely to provide valid responses to the questionnaire items; and (c) quantitative examinations of response patterns provide limited ways of identifying students who may not be providing valid responses, although changes in the questionnaire design could strengthen the utility of these methods. These findings are discussed in the context of the larger interpretive argument surrounding the PISA learning
The Secret of Self-Control

Children who are able to pass the marshmallow test enjoy greater success as adult. Read more [here](http://www.newyorker.com/reporting/2009/05/18/090518fa_fact_lehrer#ixzz1j67UWWUE)