Lessons learned from PISA (1999 – 2024)

Therese N. Hopfenbeck
Professor of Educational Assessment
Director of Assessment and Evaluation Research Centre (AERC)
Faculty of Education
Adjunct Professor, Norwegian University of Science and Technology
@TNHopfenbeck @AERC_UniMelb

Masterclass, CITO, 31st January 2024

https://education.unimelb.edu.au/aerc#home

All research published open access from AERC @AERC_Unimelb

Home
Explore
20+ Notifications
Messages
Grok
Lists
Bookmarks
Communities
Premium
Profile
More

Post

Assessment and Eval
@AERC_UniMelb

Edit profile

Assessment and Evaluation Research Centre (AERC)

@AERC_UniMelb

The AERC conducts research on assessment, evaluation, psychometrics and measurement. AERC is part of @EduMelb #education

Melbourne education.unimelb.edu.au Joined November 2019

162 Following **554** Followers

Posts Replies Highlights Media Likes

Assessment and Evaluation Research Centre (A @AERC_UniMe · 36s
Hot of the press, open access here: sciencedirect.com/science/articl...

Therese N.Hopfenbeck @TNHopfenbeck · 12m
Important new study out on developing and validating survey on care leavers well-being with @DrJoshMcGrane @AERC_UniMelb @EduMelb @KelloggOx & Julie Selwyn @ReesCentre @OxfordDeptofEd Claire Baker @CoramVoice


Children and Youth Services Review

The development and psychometric validation of a survey to measure the subjective well-being of care leavers

Joshua McGrane, Julie Selwyn, Claire Baker

Overview of talk

HISTORY Since its inception in 2000, the OECD's Programme for International Student Assessment (PISA) has become an influential large-scale assessment measuring students' achievement at the end of their compulsory schooling in science, mathematics and reading literacy.

WHAT IS PISA? From 32 participating countries in 2000, PISA 2022 was conducted in 81 with over 690 000 students responding on their test in **science**, **reading** and **mathematics literacy** and student questionnaires.

THE USE AND ABUSE OF ASSESSMENT DATA There are wide variations in student performance and educational outcomes between countries. Asian countries like Singapore, Japan and South Korea tend to score very highly on PISA, while many Western countries score around average or below.

WHY PISA STILL MATTERS Basic literacy and numeracy skills are still lacking in many developed and developing countries, indicating more needs to be done on core foundations. Based upon research conducted on PISA the last 20 years, I will reflect upon what lessons can be learned, including the risks of overreliance on rankings and the improper policy use of PISA data as well as the need for methodologic enhancements and the responsible contextualization of findings.

Overview of ILSA development

1950: United Nations Educational Scientific and Cultural Organization (UNESCO) and Organisation for Economic Co-Operation and Development (OECD) started working on collecting data across education system world-wide

1958: Research meeting in Hamburg: Bill Watson from National Foundation for Educational Research, England, Arnold Anderson and Benjamin Bloom, University of Chicago, Robert Thorndike, Colombia University and Torstein Husén, Stockholm University.

(Gustafsson, 2012, Hopfenbeck, 2014)

International Association for the Evaluation (IEA) studies

Study	Purpose and participation
ICCS (International Civic and Citizenship Education Study)	Investigates the ways in which young people are prepared to undertake their roles as citizens, builts upon a study of 9 countries in 1971, 28 countries in 1999 (CIVED) and 38 countries in 2009 (ICCS).
PIRLS (Progress in International Reading and Literacy Study)	PIRLS represents the international standard for reading comprehension at the fourth grade. Carried out in 2001, 2006, 2011, 2016, more than 70 countries participating in 2016.
TIMSS (Trends in International Mathematics and Science Study)	Measuring Science and Mathematics in fourth and eight grade, 1995, 1999, 2003, 2007, 2011, 2015. More than 60 countries have participated
(ICILS) International Computer and Information Literacy Study, http://icils2013.acer.edu.au/	Examined the outcomes of student computer and information literacy (CIL) across countries. CIL refers to an individual's ability to use computers to investigate, create, and communicate in order to participate effectively at home, at school, in the workplace, and in the community
TIMSS Advanced http://timssandpirls.bc.edu/timss2015-advanced/frameworks.html	Final year of secondary schooling, (approximately 18 year olds), Mathematics and Physics, algebra, calculus, and geometry in mathematics; and mechanics and thermodynamics, electricity and magnetism, and wave phenomena and atomic/nuclear physics in physics

OECD studies

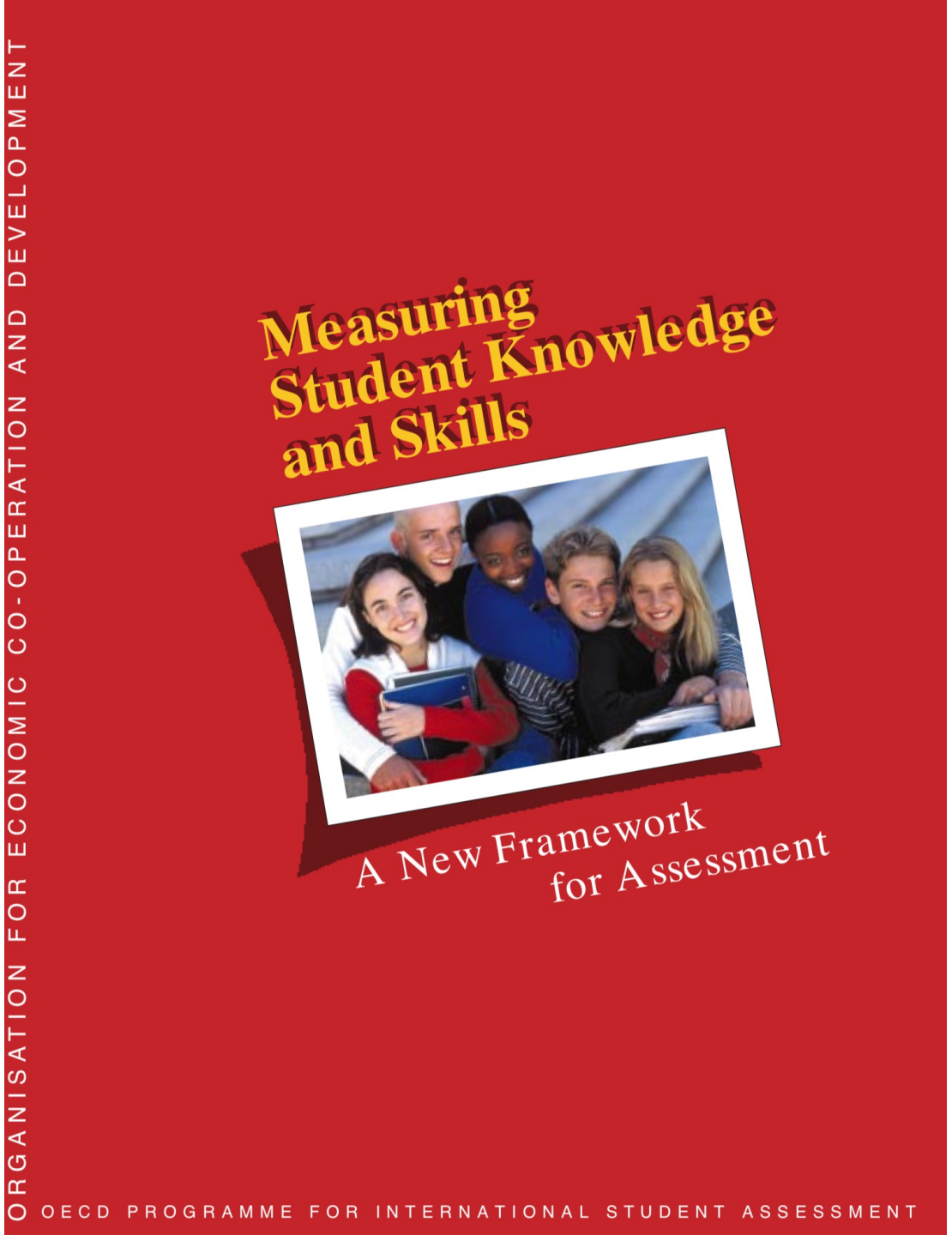
The Organisation for Economic Co-Operation and Development (OECD)

Study	Purpose and participants
AHELO (Assessment of Higher Education Learning Outcomes)	Measuring Learning Outcomes in Higher Education. A feasibiliity study was carried out in 2012 with 17 participating countries. http://www.oecd.org/site/ahelo/ First results presented in 2013.
PIAAC (Programme for the International Assessment of Adult Competencies)	http://www.oecd.org/site/piåc/surveyofadultskills.htm Measures the key cognitive and workplace skills needed for the individual to participate in society. 33 participating countries in 2012. First results presented in 2013.
PISA (Programme for International Student Assessment)	http://www.oecd.org/pisa/home/ Triennial international survey, aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-olds in Science, Mathematics and Reading. Most recent development, global competences, and collaborative problem solving, **PISA for Schools** and **PISA for Development.** From 32 countries in 2000, to more than 98 in 2022.

What is PISA?

OECD (1999) Measuring Student Knowledge and Skills
A New Framework for Assessment, Paris, OECD

Free to download:
https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33693997.pdf



PISA

- Reading, Mathematic, Science literacy
- Problem solving
- Global competencies (first in 2018)
- Student background questionnaire;
- Motivation, reading strategies, interests
- Parents' education, language spoken at home, socio-economic status

Science

Wynne Harlen, Chair
The Scottish Council for Research in Education
Edinburgh, United Kingdom

Peter Fensham
Monash University
Melbourne, Australia

Raul Gagliardi
Geneva, Switzerland

Donghee Shin
Korea Institute of Curriculum and Evaluation
Seoul, Korea

Svein Lie
University of Oslo
Oslo, Norway

Manfred Prenzel
Institut für die Pädagogik der Naturwissenschaften (IPN)
(Institute for Science Education)
Kiel, Germany

Senta Raizen
National Center for Improving Science Education
Washington, DC, United States

Elizabeth Stage
University of California
Oakland, California, United States

OECD test development,
Expert Advisory Groups

Background

PhD student, University of Oslo, supervisor

Prof Svein Lie, (PISA 2006 data, Norway)

Test-developer problem-solving, PISA 2012

Member of QEG PISA 2015 (ETS, US)

Member of QEG, PISA 2018 (ETS, US)

Member of QEG 2021 (ETS, US)

Chair of QEG 2025, (ACER, Australia)

Research Lead for PISA2021 in England, Northern Ireland and Wales (2018 – 2022)

Learning about Students' Learning Strategies

An empirical and theoretical investigation of self-regulation and learning strategy questionnaires in PISA

Therese Nerheim Hopfenbeck

Dissertation submitted to the Faculty of Education

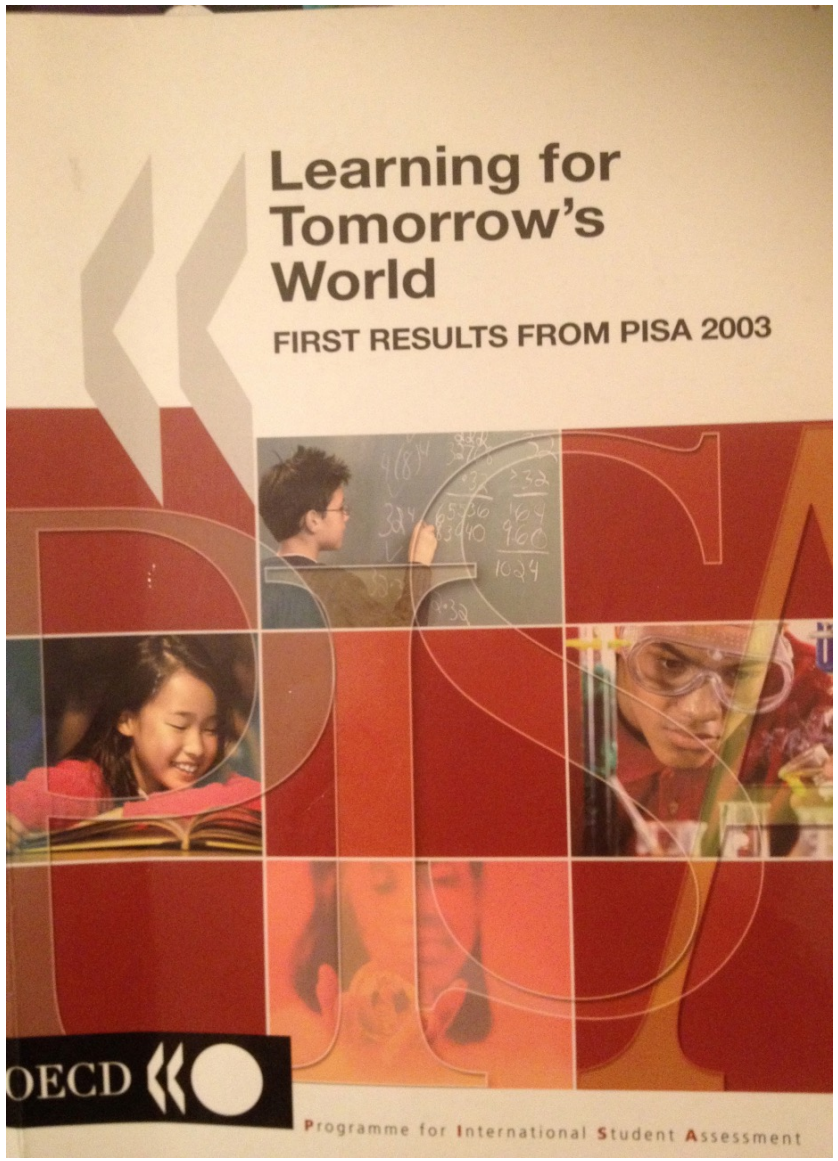
University of Oslo, Norway

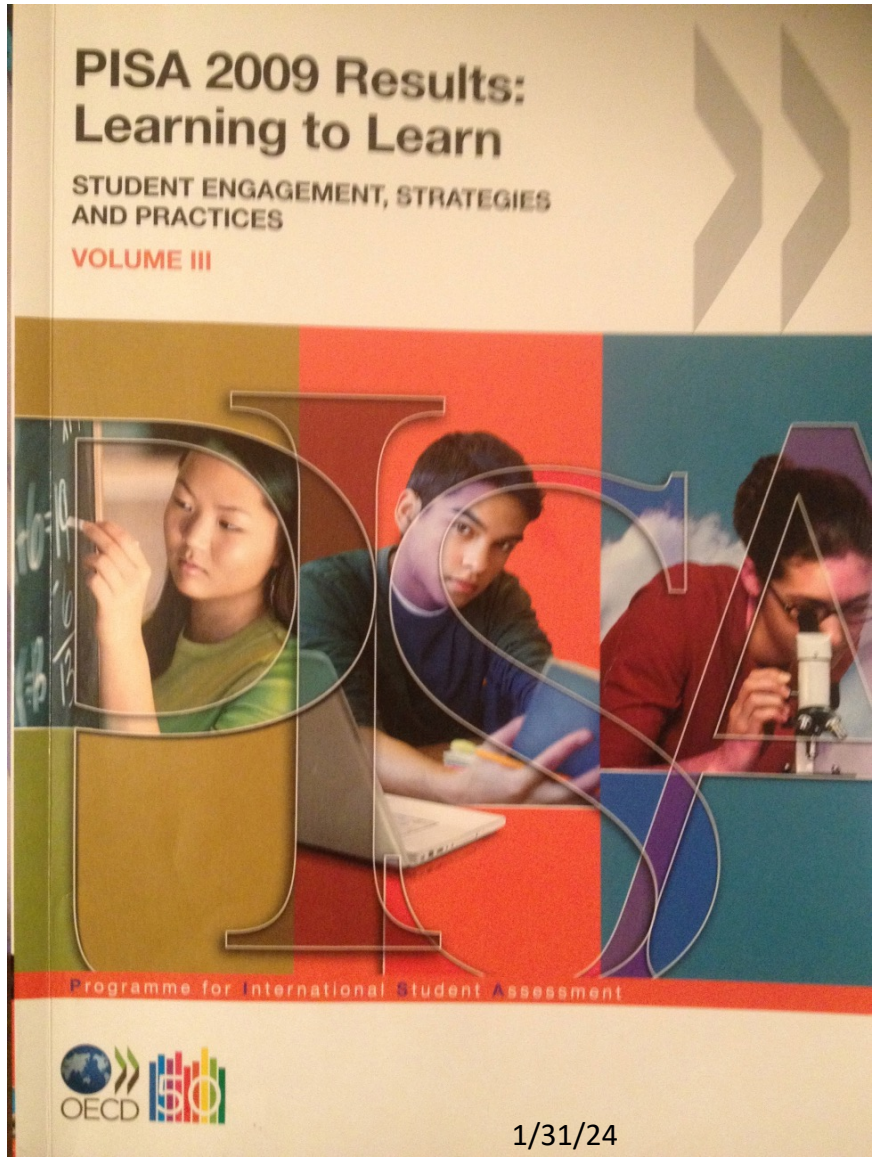
2009

PISA according to OECD

"Are students well prepared for future challenges? Can they analyse, reason and communicate effectively? Do they have the capacity to continue learning throughout life? The OECD Programme for International Student Assessment (PISA) answers these questions and more, through its surveys of 15-year-olds in the principal industrialised countries. Every three years, it assesses to what extent students near the end of compulsory education have acquired some of the knowledge and skills essential for full participation in society".

(http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html)

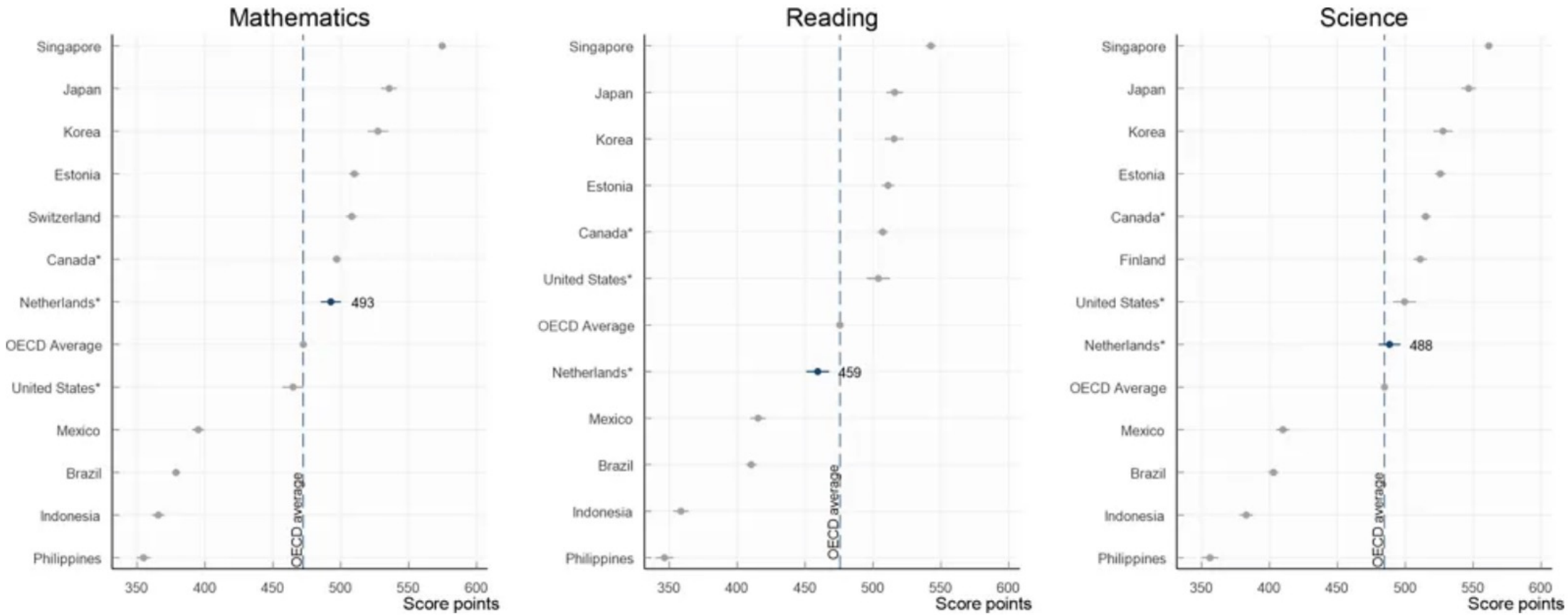

Learning for Tomorrow's World
FIRST RESULTS FROM PISA 2003
OECD
Programme for International Student Assessment


PISA 2009 Results: Learning to Learn
STUDENT ENGAGEMENT, STRATEGIES AND PRACTICES
VOLUME III
Programme for International Student Assessment
OECD

How do the Netherlands compare?

Figure 2. **Mean performance in mathematics, reading and science in PISA 2022**

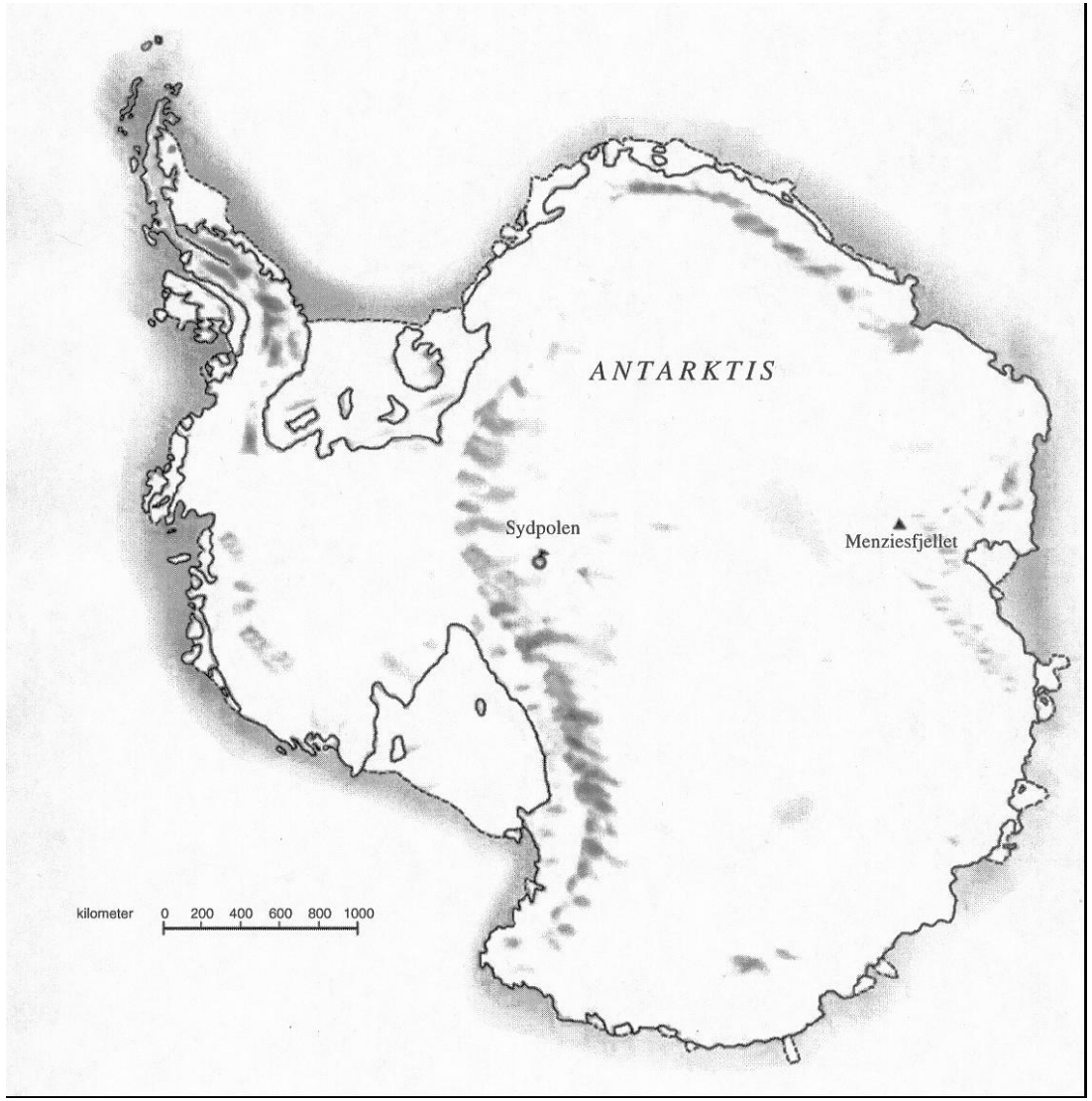
Netherlands, OECD average and selected comparison countries



Notes: Comparison countries include the six highest-performing countries in each subject and the five

From: TAKE THE TEST: SAMPLE QUESTIONS FROM OECD´S PISA ASSESSMENTS (OECD 2009):

Estimate the area of Antarctica using the map scale.
Show your working out and explain how you made your estimate. (You can draw over the map if it helps you with your estimation).



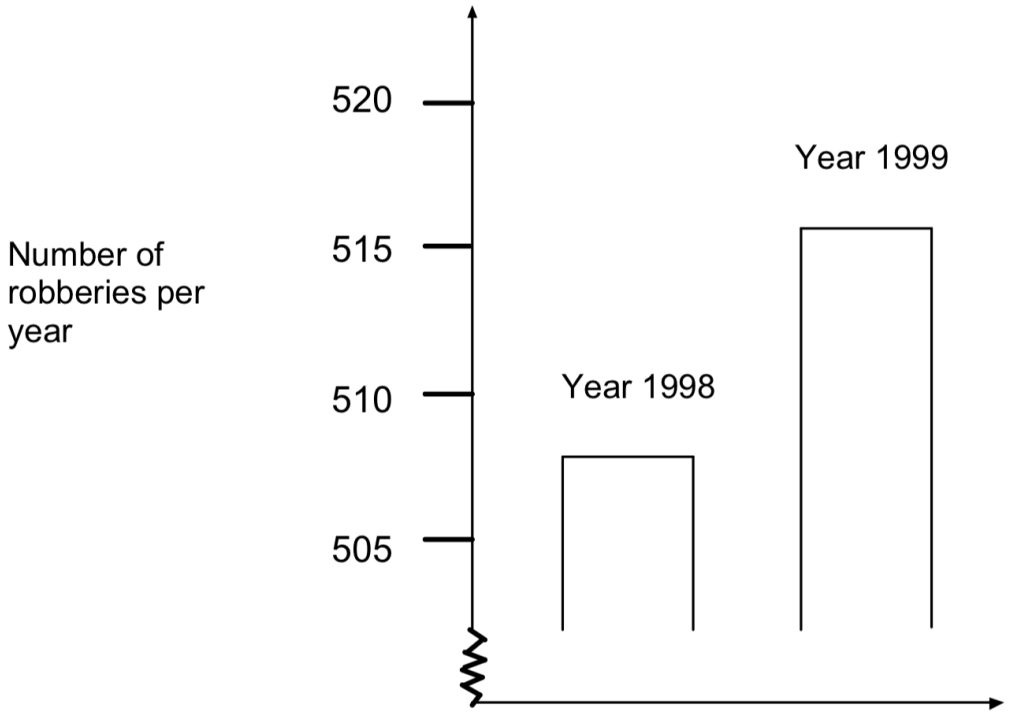
M179: Robberies

Question 1: ROBBERIES

M179Q01- 01 02 03 04 11 12 21 22 23 99

A TV reporter showed this graph and said:

"The graph shows that there is a huge increase in the number of robberies from 1998 to 1999."



Do you consider the reporter's statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.

THE CURRICULA APPROACH – TIMSS, Science, grade 8

Most underground caves are formed by the action of water on

a) Granite

b) Limestone

c) Sandstone

d) Shale

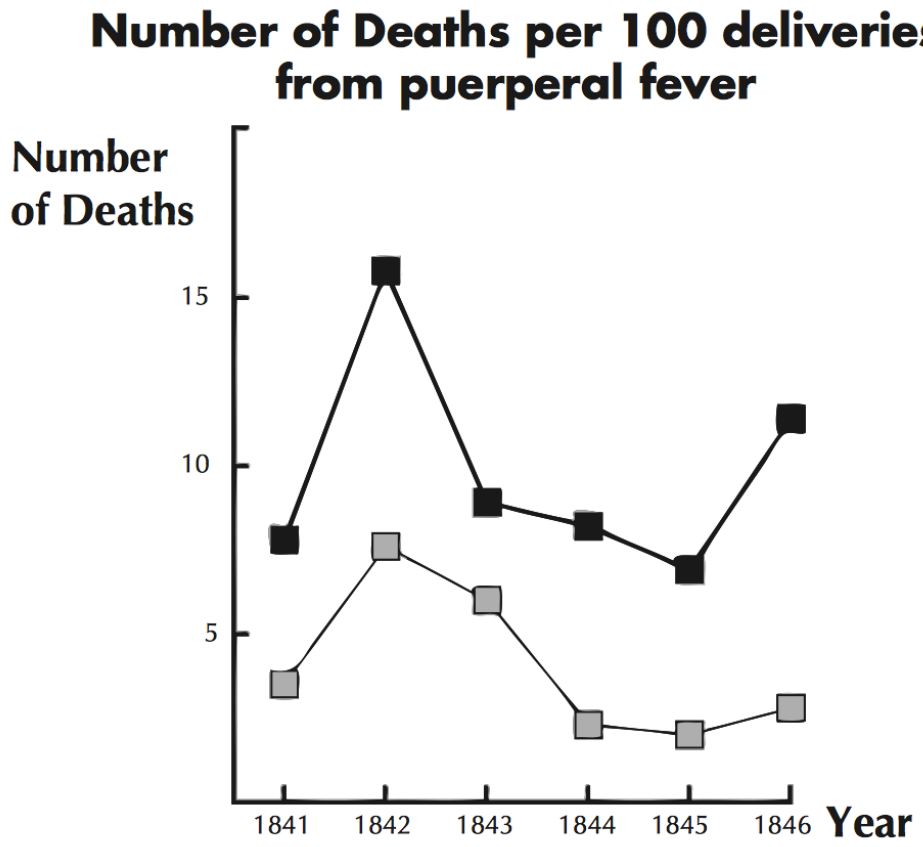
Cognitive domain: Factual Knowledge, main topic Earths structure and physical features.

THE LITERACY APPROACH SCIENCE - PISA

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).



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.

.....

.....

.....

International Journal of Testing, 11: 95–121, 2011

ISSN: 1530-5058 print / 1532-7574 online
DOI: 10.1080/15305058.2010.529977

Routledge
Taylor & Francis Group

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 strategy scales, and implications for future research on the measurement of learning strategies are discussed.

Keywords: cognitive interviews, learning strategies, PISA, person fit, response processes, validity

THE PISA LEARNING STRATEGY SCALES

The classic definition of validity: that the scales measure the variables they claim to measure, (Garrett, 1937, p. 324), a definition also endorsed by modern scholars such as Borsboom, Mellenberg, and Van Heerden (2004).

How do we know?

In the case of the PISA learning strategy scales, this can be restated as students respond to test items in a manner consistent with the predictions of theory, and their responses can be taken as indicative of their real-world learning behaviours .

Instrument validation I

“ Modern views of instrument validation (e.g., Kane, 2006; Messick, 1995) note

that it involves “an evaluation of whether the proposed interpretations and uses [of

a test] are plausible and appropriate” (Kane, 2006, p. 17); seen this way, validity

itself is defined by the Standards for Educational and Psychological Testing as

“the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests”

and is “the most fundamental consideration in developing and evaluating tests” (American Educational Research Association, American Psychological Association, & National Council for Measurement in

Education [AERA, APA, & NCME], 1999, p. 9).

Instrument validation II

Thus, test validity requires a unitary, logical argument linking item responses to the latent object of measurement,forwhich multiple kinds of evidence can and should be examined; in particular, the Standards refer to five “strands” of evidence relevant to the evaluation of validity: **evidence based on test content, response processes, internal structure, relations to other variables, and the consequences of testing**”

(Hopfenbeck and Maul 2011)

Imagine that you are preparing for a test in science. To what degree do you agree with the following statements?

Elaboration strategies

1. I concentrate in order to understand the relationship between what I am learning and other subjects.

2. I try to relate what I have just read to what I already know.

3. I try to find out how the subject can be used in everyday life.

4. I try to understand how the new subject concepts are interrelated.

Memorization strategies

5. I usually go through every step in the procedure for solving a typical science task

until I know the procedure by heart.

6. I practice again and again answering some typical science questions.

7. I try to memorize as much as possible.

8. I try to remember all the details.

9. I go through procedures over and over again in order to answer common science questions.

Control strategies

10. I start by working out exactly what I need to learn.

11. I try to find out which concepts I do not understand properly.

12. I try to control that I have understood the main subject properly.

13. I try to find out what is most important to learn.

Four point Likert scale:

(1) Strongly disagree (2) disagree (3) agree (4) strongly agree.

Questioning the learning strategies construct

TABLE 5
Student Response Patterns by Achievement Level

	Varied Response Pattern	"Strongly Agree" with All	"Agree" with All	"Disagree" with All	"Strongly Disagree" with All	Total Suspect Response Patterns	Total Number of Students
Group 1 (0–19th percentile)	598	49	83	15	37	184	782
	76.5%	6.3%	10.6%	1.9%	4.7%	23.5%	
Group 2 (20th–39th percentile)	729	24	64	11	3	102	831
	87.7%	2.9%	7.7%	1.3%	.4%	12.3%	
Group 3 (40th–59th percentile)	768	20	67	5	7	99	867
	88.6%	2.3%	7.7%	.6%	.8%	11.4%	
Group 4 (60th–79th percentile)	817	14	44	1	5	64	881
	92.7%	1.6%	5.0%	.1%	.6%	7.3%	
Group 5 (80th–99th percentile)	856	26	33	1	2	62	918
	93.2%	2.8%	3.6%	.1%	.2%	6.8%	
Total	3768	133	291	33	54	511	4279
	88.1%	3.1%	6.8%	.8%	1.3%	11.9%	

Note. Standard errors on percentages are approximately 2 points.

Do surveys provide reliable and valid data?

We tend to answer in a socially desirable way (Burstein et al 1995).

State of fatigue, lack of motivation (Entwhistle & McCune, 2004)

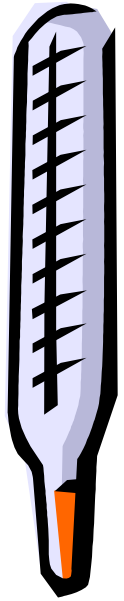
Agreeableness (Allan, 1997; Nunally, 1967).

Check off for the same response category (Hopfenbeck and Maul 2011)

EFFORT SURVEY

How much effort did you invest?

Please try to imagine an actual situation (at school or in some other context) that is highly important to you personally, so that you would try your very best and put in as much effort as you could to do well.

In this situation you would mark the highest value on the "effort thermometer," as shown below:	Compared to the situation you have just imagined, how much effort did you put into doing this test?	How much effort would you have invested if your marks from the test were going to be counted in your school marks?
		
☒ 10	☐ 10	☐ 10
☐ 9	☐ 9	☐ 9
☐ 8	☐ 8	☐ 8
☐ 7	☐ 7	☐ 7
☐ 6	☐ 6	☐ 6
☐ 5	☐ 5	☐ 5
☐ 4	☐ 4	☐ 4
☐ 3	☐ 3	☐ 3
☐ 2	☐ 2	☐ 2
☐ 1	☐ 1	☐ 1

Thank you

Media and Research

But what is the evidence behind it?

The Guardian
A decade of making a difference

News | Opinion | Sport | Culture | Lifestyle

Australia | World | AU politics | Environment | Climate crisis | Indigenous Australia | Immigration

Opinion Australian education

Australian students aren't trying in the Pisa exams. They should be congratulated for their disdain

Van Badham

The most meaningful information to come out of the Pisa assessment is the revelation that Australian kids don't give a shit about it

Fri 12 Jan 2024 13.00 AEDT

352

'I think there's cause for national pride in our true Pisa results after the Save Our Schools survey's revelations.' Photograph: Tim Macpherson/Getty Images

The use and abuse of assessment data

"In an area where the media produces narratives on schools and education systems based upon rankings in PISA, researchers in the field of large-scale assessment studies have a particularly important role in stepping up and advising on how to interpret and understand these studies, while warning against potential misuse".

Hopfenbeck (2016) The power of PISA – limitations and possibilities for educational research, Assessment in Education: Principles, Policy & Practice, 23:4, 423-426, DOI: 10.1080/0969594X.2016.1247518

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE, 2016
VOL. 23, NO. 4, 423–426
http://dx.doi.org/10.1080/0969594X.2016.1247518

Routledge
Taylor & Francis Group

EDITORIAL

The power of PISA – limitations and possibilities for educational research

On 6 December 2016, the Programme for International Student Assessment (PISA) releases its report on the achievements of 15-year-olds from 72 countries and economies around the world. This triennial international survey aims to evaluate education systems across 72 contexts by testing skills in Mathematics, Science and Reading Literacy. This is the sixth cycle of PISA and the OECD suggests countries and economies now have the capability to compare the results over time to 'assess the impact of education policy decisions'[1]. Compared to other education studies, the media coverage of PISA must be described as massive (Baird et al., 2016; Meyer & Benavot, 2013) and, as with previous years, it is expected that PISA will attract considerable discussion among policy-makers, educators and researchers (Wiseman, 2014). It is therefore timely to present a thematic issue of *Assessment in Education*, where we publish four articles that have analysed previous data-sets from the PISA studies each commenting upon the challenges, limitations and potential future assessment research on the PISA data.

The articles touch upon issues regarding sampling, language, item difficulty and demands, as well as the secondary analyses of students' reported experiences of formative assessment in the classroom. One important message from the authors in this thematic Special Issue is the need for a more complex discussion around the use and misuse of PISA data, and the importance of pointing to the limitations of how the results are presented to policy-makers and the public. In an area where the media produces narratives on schools and education systems based upon rankings in PISA, researchers in the field of large-scale assessment studies have a particularly

The PISA shocker

Dev Lahiri Oct 9, 2012, 12.00AM IST

Tags: Pisa | International Curriculum Development

India's poor performance in the Programme for International Student Assessment (PISA) tests (India ranked 71 out of 73 nations) sparked a raging controversy about the adequacy of our education system. The fact is that in India we are still, by and large, trapped in the paradigm of teaching "content". All over the world (including China), the shift has been towards teaching "skills", that are necessary for making a truly global 21st century citizen.

What are these skills? They are, first and foremost, the ability to ask questions. Unless a child develops her innate curiosity, she will be a misfit in the new-age work environment. In India, except for the most progressive schools (and how many of those exist?), questioning is seriously discouraged. This is partly the result of cultural conditioning (asking questions of elders is considered disrespectful), partly because of the extraordinarily large number of students in classes, and largely because the teachers don't know the answers!

Strong Performers and Successful Reformers in Education

A video series profiling policies and practices of education systems that demonstrate high or improving performance in the PISA tests

FILMS: Select a Country

Germany

VIDEO FEATURE: Germany + Embed + Download + Transcripts

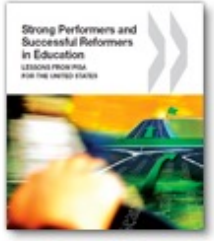
Key facts

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 was 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



Read the full report

DEMOCRACY LIVE

Search, find and watch the politics that affects you

Search Democracy

Find a representative

Home | House of Commons | House of Lords | Northern Ireland | Scotland | Wales | Cymru | Europe

Previous schedule | Your representatives | Institution guides | Historic moments | BBC Parliament

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 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.

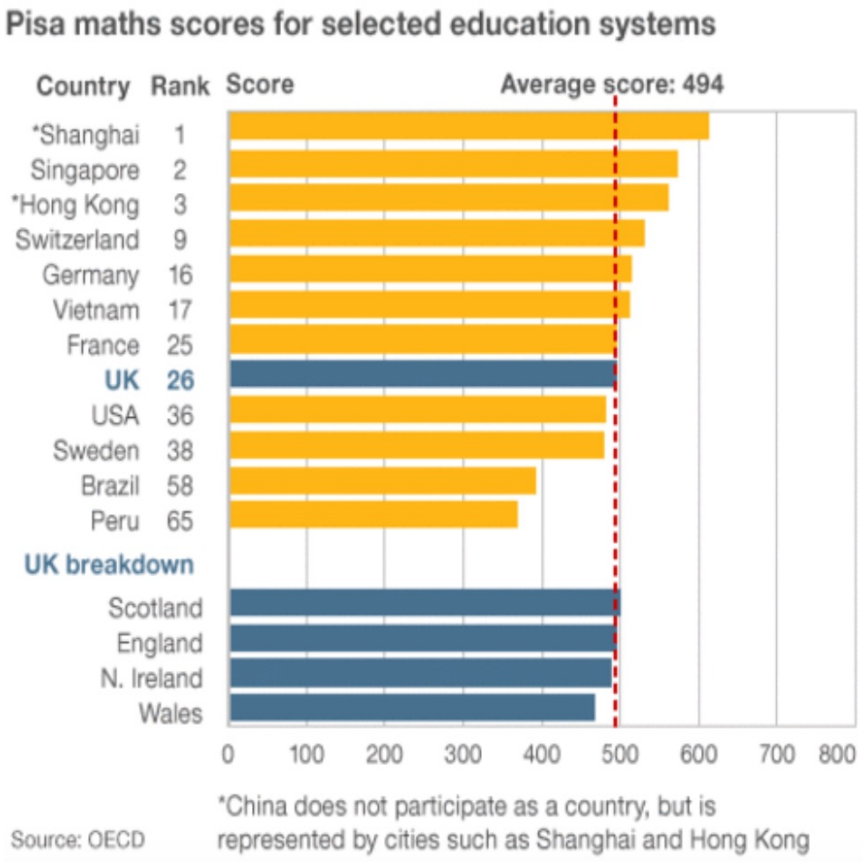
Results from International assessments such as Programme for International Student Assessment (PISA) are used by politicians around the world in the policy discussion around education and assessment.

Pisa tests: UK stagnates as Shanghai tops league table

By Sean Coughlan

BBC News education correspondent

3 December 2013 | Education & Family




PISA夺冠后的教育思考

——从PISA看中国上海学生阅读素养

上海PISA2009阅读素养测评主要结果

多种角度看PISA引发的阅读关注

Shanghai China

152 schools and a total of 5115 students participated in PISA 2009, representing 10 million students in Shanghai.

Angel Gurria, Secretary General of OECD, preface in PISA 2009:

"..the stunning success of Shanghai China which tops every league table in this assessment by a clear margin, shows what can be achieved with moderate economic resources in a diverse social context. Overall PISA shows that an image of a world divided neatly into rich and well-educated countries and poor and badly –educated countries is out of date" (OECD, 2010, p.5)".

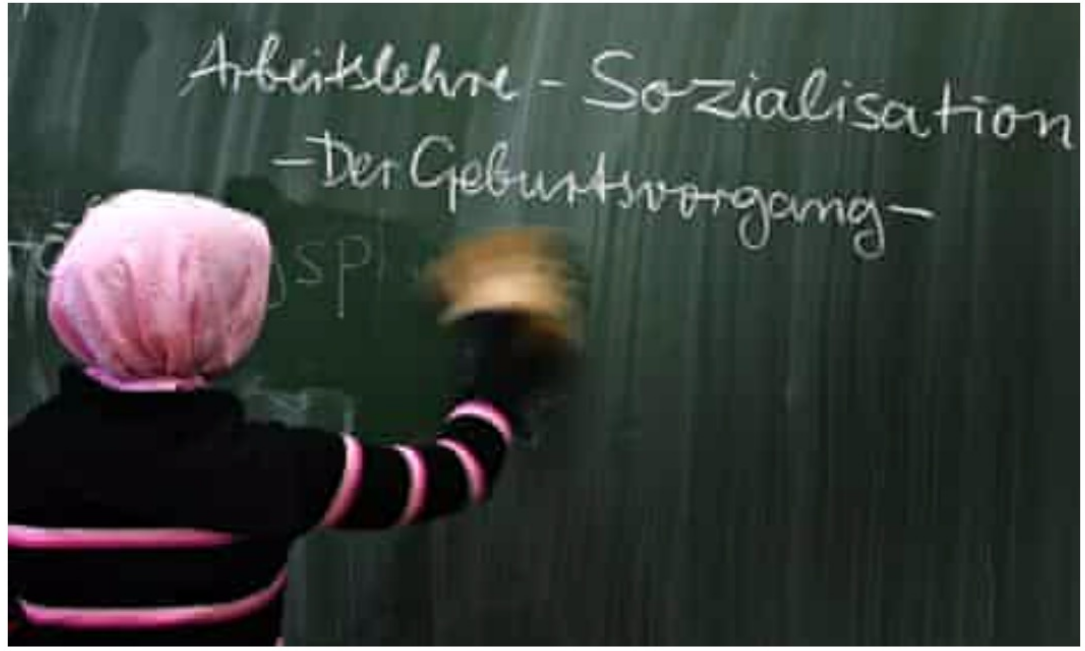
Similarities? Differences?

Although it is difficult to research effectively, the portion of students in China who receive private tutoring in math was 28.8% and in English, 29.3% (Zhang, 2011 (Heyneman, 2014: 294)

Harvey Goldstein: do we measure the school system or variables outside the school?

Global school tests under attack as OECD accused of killing 'joy of learning'

Leading academics from 12 countries including UK call for next round of OECD Pisa tests on 15-year-olds to be scrapped



▲ The signatories say the results of the Pisa tests are 'known to be imperfect' because they focus narrowly on the economic goals of education. Photograph: Tobias Schwarz/Reuters

Leading academics have accused the Organisation for Economic Co-operation and Development (OECD) of acting as an unaccountable super-ministry of education which kills the "joy of learning" and turns schooling into "drudgery".

A letter signed by 120 leading academics and teachers from 12 countries - including Britain, the US and Germany - argues the OECD's Programme for International Student Assessment (Pisa) tests on 15-year-olds distort the curriculum, reduce teachers' autonomy and increase children's stress levels.

The results of the Pisa tests, which the signatories say are "widely known to be imperfect" because they focus narrowly on the economic goals of education,

Letter in the Guardian, 6th May 2014

https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics

https://www.globalpolicyjournal.com/blog/05/05/2014/open-letter-andreas-schleicher-oecd-paris

OECD and Pisa tests are damaging education worldwide - academics

In this letter to Dr Andreas Schleicher, director of the OECD's Programme for International Student Assessment, academics from around the world express deep concern about the impact of Pisa tests and call for a halt to the next round of testing

▲ School children in Sichuan province in China. Academics say the OECD should develop alternatives to league tables and find more meaningful ways of reporting assessment, taking account of different cultures. Photograph: James Zeng Huang/Corbis Sygma

Dear Dr Schleicher,

We write to you in your capacity as OECD's (Organisation for Economic Co-operation and Development) director of the Programme of International Student Assessment (Pisa). Now in its 13th year, Pisa is known around the world as an instrument to rank OECD and non-OECD countries (60-plus at last count) according to a measure of academic achievement of 15-year-old students in mathematics, science, and reading. Administered every three years, Pisa results are anxiously awaited by governments, education ministers, and the editorial boards of newspapers, and are cited authoritatively in countless policy reports. They have begun to deeply influence educational practices in many countries. As a result of Pisa, countries are overhauling their education systems in the hopes of improving their rankings. Lack of progress on Pisa has led to declarations of crisis and "Pisa shock" in many countries, followed by calls for resignations, and far-reaching reforms according to Pisa precepts.

We are frankly concerned about the negative consequences of the Pisa rankings. These are some of our concerns:

Psychometrika
April 2014, Volume 79, Issue 2, pp 210–231 | Cite as

Analyses of Model Fit and Robustness. A New Look at the PISA Scaling Model Underlying Ranking of Countries According to Reading Literacy

Authors | Authors and affiliations

Svend Kreiner, Karl Bang Christensen

Article
First Online: 14 June 2013

Abstract

This paper addresses methodological issues that concern the scaling model used in the international comparison of student attainment in the Programme for International Student Attainment (PISA), specifically with reference to whether PISA's ranking of countries is confounded by model misfit and differential item functioning (DIF). To determine this, we reanalyzed the publicly accessible data on reading skills from the 2006 PISA survey. We also examined whether the ranking of countries is robust in relation to the errors of the scaling model. This was done by studying invariance across subscales, and by comparing ranks based on the scaling model and ranks based on models where some of the flaws of PISA's scaling model are taken into account. Our analyses provide strong evidence of misfit of the PISA scaling model and very strong evidence of DIF. These findings do not support the claims that the country rankings reported by PISA are robust.

Comments on *Kreiner 2011: Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment*

Ray Adams, April 19 2011[i]

This paper is concerned with two issues:

- Do the outcomes of PISA – more particularly the rankings – depend on the items chosen?
- Could alternative scaling procedures, produce different outcomes – more particularly the rankings.

A careful examination of these two issues is clearly an important activity that should be the focus of rigorous scholarly investigation. Unfortunately this article presents neither a rigorous nor a scholarly investigation.

In making comments on the article I shall organise them into two sections. First, I shall make some general points about the work that cover the implications for PISA. Second, I shall make some more specific points that relate to the paper's acceptability as an academic piece of work.

Some of the critiques…

“We cannot wander at pleasure among the educational systems of the world, like a child strolling through a garden, and pick off a flower from one bush and some leaves from another and then expect that if we stick what we have gathered into the soil at home, we shall have a living plant (Sadler, 1900, in Higginson, 1979, side 49).

PISA's influence on education policy

"Today's PISA report underlines the urgent need to reform our school system. We need to learn from the best-performing countries" (Department of Education, 2010). Secretary for State of Education, Michael Gove.

President Obama, in a speech to a college audience in North Carolina, recalled the Soviet Union's 1957 launching of Sputnik and how it provoked the United States to increase investment in math and science education. Commenting on the PISA results he concluded: "Fifty years later, our generation's Sputnik moment is back… America is in danger of falling behind"

Oxford Review of Education
Vol. 32, No. 5, November 2006, pp. 619–634

Routledge Taylor & Francis Group

Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany

Hubert Ertl*
University of Oxford, UK

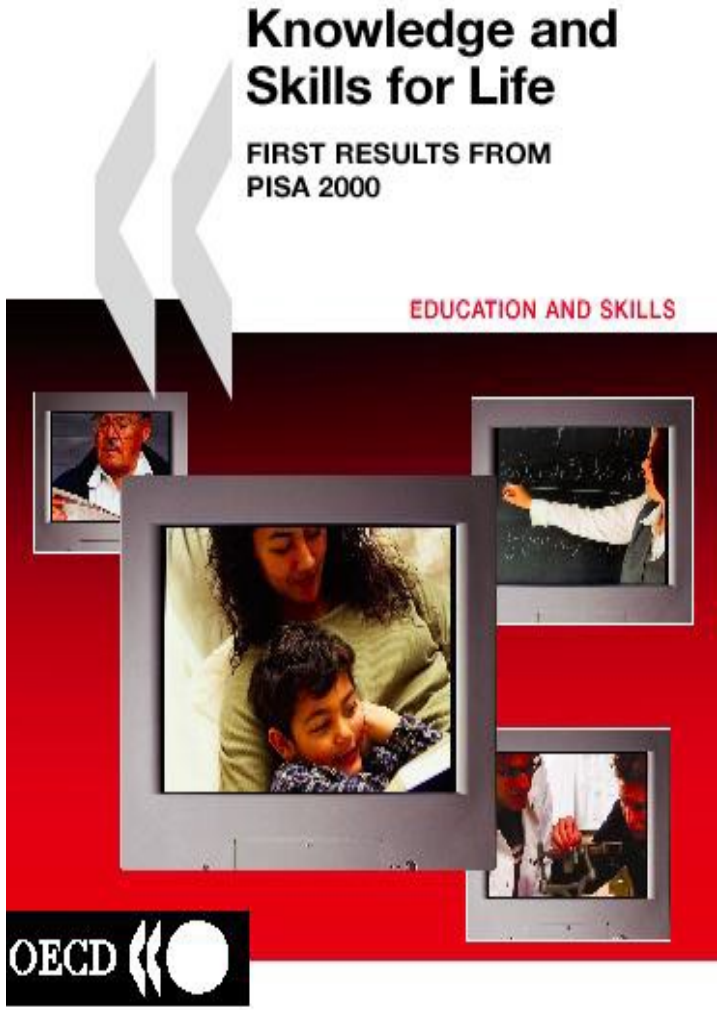
This paper examines the ways in which the results of the OECD Programme of International Student Assessment (PISA) have influenced educational discourse in Germany. It argues that the results caused shockwaves in the educational landscape and led to a re-evaluation of other international comparisons which had presented an unfavourable picture of education in Germany. The paper outlines three main areas affected by the reception of the PISA study in Germany: political discourse, curriculum development processes and academic discourse on education. Changes in the political discourse resulted in a wide-ranging reform agenda, with the introduction of national educational standards being the most significant of a number of initiatives and programmes. In terms of curriculum development processes, the paper argues that PISA has led to the growing importance of principles such as outcome control, competence orientation and external assessment. The post-PISA academic discourse in Germany can be characterised by the re-orientation of educational studies towards a greater emphasis on the empirical research of pedagogic practice (*empirische Unterrichtsforschung*). All three areas of change are informed by more or less systematic comparisons

The PISA results in 2001

came as a shock to the Norwegian public:

We thought we were good!

It turned out we were average.



The OECD perspective: The case of Norway

"…it is a country where the Ministry of Education has really very limited control over what happens in the classrooms" Andreas Schleicher

I think that is why really PISA in the first time that there has been some kind of objective reflection on results for which the perceptions of teachers is very different. Three times, I was involved in the discussions on national tests"

(Interview in 2012, From Hopfenbeck et al. 2013).

DOI: 10.1111/ejed.12219

ORIGINAL ARTICLE

WILEY

The politics of PISA: The media, policy and public responses in Norway and England

Therese N. Hopfenbeck[1] | **Kristine Görgen**[2]

[1] Department of Education, University of Oxford, 15 Norham Gardens, OX2 6PY, Oxford, UK

[2] Department of Education, University of Oxford, Green Templeton College, Oxford, UK

Correspondence
Therese N. Hopfenbeck, Department of Education, University of Oxford,
15 Norham Gardens OX2 6PY Oxford, UK
Email: therese.hopfenbeck@education.ox.ac.uk

Abstract

Using the PISA 2015 releases in Norway and England, this article explores how PISA has been presented in the media and how the policy level has responded to the results. England will be used as an example for comparison. The article presents early media responses from the 20 most circulated daily newspapers in the two countries and discusses them in relation both to the national PISA reports in Norway and England, as well as the international report of the OECD. The media responses are further interpreted in light of previous research in both countries, with a particular focus upon Norway, where previous Ministers of Education have been interviewed about assessment policy and education reforms.

KEYWORDS
education policy responses to PISA results, England, Media reception of PISA, Norway

PISA 2018: did Sweden exclude students according to the rules?

Christian Andersson [a,b] and Sofia Sandgren Massih [a]

[a]Department of Performance Audit, The Swedish National Audit Office, Stockholm, Sweden; [b]Department of Industrial Economics, Blekinge Institute of Technology, Karlskrona, Sweden

ABSTRACT

This study assesses whether student exclusions from PISA 2018 in Sweden followed the criteria set by the OECD. We do this using both qualitative and quantitative methods. Our conclusion is that the exclusions made in PISA 2018 in Sweden did not follow OECD criteria and were much too high. Furthermore, interviews with school coordinators indicate that many of them misunderstood the OECD criteria. We also conclude that the National Agency for Education did not sufficiently follow up on exclusions. A review of the Swedish exclusion rate made by the OECD did not present credible results but accepted the results. A recalculation of PISA 2018 scores for Sweden where we assume non-participating students to be low performers show that results are significantly affected.

ARTICLE HISTORY

Received 20 January 2022
Accepted 27 February 2023

KEYWORDS

PISA; Sweden; OECD; participation rates; exclusion rate

"... the authors suggest that a review of the Swedish exclusion rate made by the OECD did not present credible results but accepted the results. The authors have conducted a recalculation of PISA 2018 scores for Sweden, demonstrating that the results are significantly affected. It is in the public interest to debate this article and the analysis and results presented. We also need a conversation on how countries interpret and practice the exclusion criteria across participating countries. Large-scale studies depend upon the public trust, transparency around sampling procedures and interpretation of exclusion criteria is hence key for these studies moving forward"

(Hopfenbeck, 2023)

SCANDINAVIAN JOURNAL OF EDUCATIONAL RESEARCH, 2018
VOL. 62, NO. 3, 333–353
https://doi.org/10.1080/00313831.2016.1258726

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.

ARTICLE HISTORY
Received 3 January 2016
Accepted 30 October 2016

KEYWORDS
PISA; policy impact; secondary analysis; critiques

PISA Review: articles on the rise

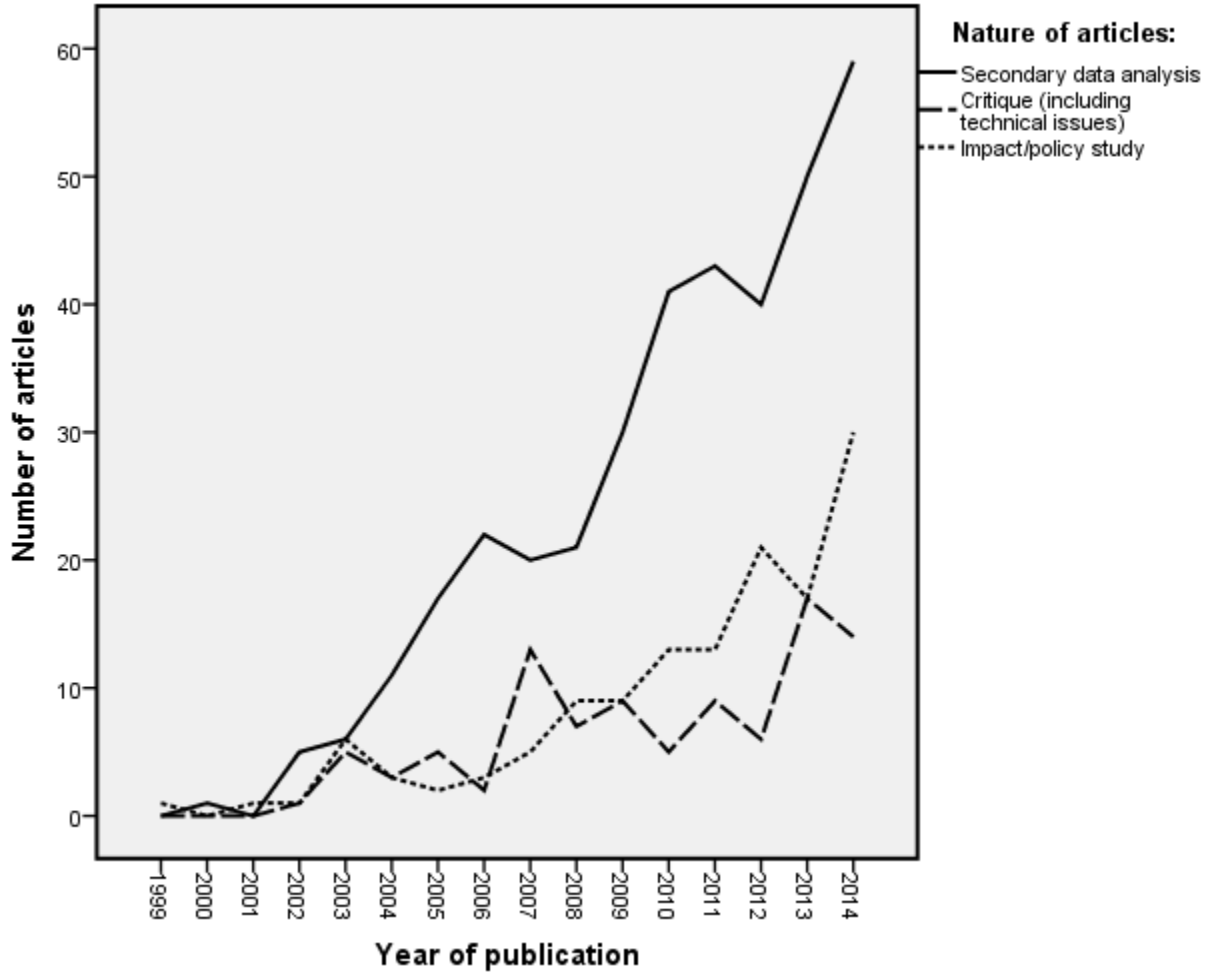
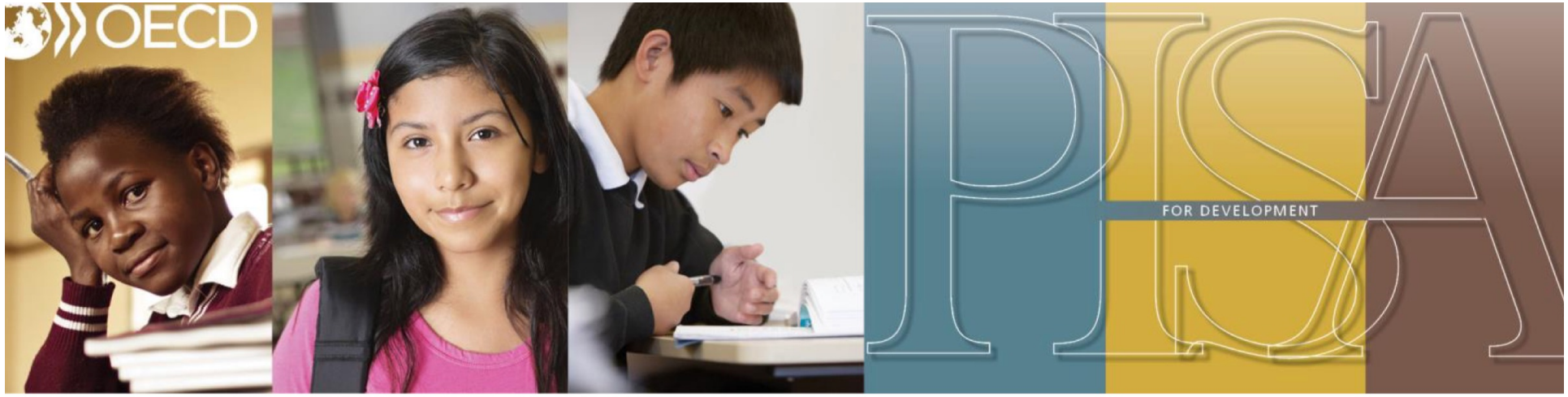


Table 1. Publication frequency: top five countries.

Discipline	USA	Australia	Germany	UK	Ireland
Educational research	29	23	18	18	22
Educational psychology	1	1	3	0	0
Sociology/social sciences	6	4	3	3	2
Comparative research	8	6	2	6	0
Literacy/reading	2	1	3	0	1
Mathematics research	5	2	2	1	1
Science research	10	9	2	4	0
Economics	5	3	11	5	2
Teaching and instruction	14	6	4	1	0
Psychology	6	1	3	2	3
Policy studies	5	1	6	3	0
Assessment and testing	13	10	6	4	0
Total number of articles	114	72	69	52	31

OECD publishes data, reports and videos on the following website: https://www.oecd.org/pisa/



PISA for Development Brief 1

What is PISA for Development?

- The PISA for Development (PISA-D) initiative launched by the OECD and its partners aims to encourage and facilitate PISA participation by interested and motivated low- and middle-income countries.

- The project builds capacity for managing large-scale student learning assessment and using the results to support policy dialogue and decision making in participating countries: Cambodia, Ecuador, Guatemala, Honduras, Panama, Paraguay, Senegal and Zambia.

- PISA-D contributes to the monitoring and achievement of the Education Sustainable Development Goal, which emphasises quality and equity of learning outcomes for children, young people and adults.

So why PISA?

- Trend data
- Possibility to ask questions that matters (science, climate change, self and co-regulation, diversity, well-being, critical thinking, democracy, SES)
- Matched datasets
- BUT – we need more knowledge on data use/and abuse

Selected references

Andersson, C., & Sandgren Massih, S. (2023). PISA 2018: Did Sweden exclude students according to the rules?, *Assessment in Education: Principles*. Policy & Practice. https://doi.org/10.1080/ 0969594X.2023.2189566

Eklöf, H. & Hopfenbeck,T. N. (2018) *Self-Reported Effort and Motivation in the PISA Test*, chapter in book edited by Bryan Maddox, Bloomsbury Publishing Plc.

Goldstein, H. (2004) International comparison of student attainment: Some issues arising from the PISA study. *Assessment in Education: Principle, Policy and Practice*. 11, 319 – 464

Goldstein, H. 2004. "Education for All: The Globalization of Learning Targets." *Comparative Education* 40 (1): 7–14.

Hopfenbeck, T.N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J. & J.A. Baird (2018) Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International student Assessment, *Scandinavian Journal of Educational Research.*

Hopfenbeck, T.N. & Maul, A. (2011) Examining Evidence for the Validity of PISA Learning Strategy Scales Based on Student Response Processes, *International Journal of Testing*, 11 (2), 95-121.

Hopfenbeck, T.N. (2023) Cultivating realistic self-appraisal: examining student overclaiming and fair assessment through PISA and classroom data, *Assessment in Education: Principles, Policy & Practice*, 30:5-6, 347-350, DOI: 10.1080/0969594X.2023.2294428

OECD. (2019a). *PISA 2018 Results (Volume I): What Students Know and Can*. Paris: OECD Publishing. https://doi.org/10.1787/5f07c754-en

OECD. (2019b). PISA 2018W technical report. (Accessed 27 January. 2021). https://www.oecd.org/pisa/data/pisa2018technicalreport/.

OECD. (2014b). *PISA 2012 Results in Focus. What 15-year-olds know and what they can do with what they know.* http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf


images.
Thank you
Therese.hopfenbeck@unimelb.edu.au
@TNhopfenbeck