www.etn-occam.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

European Training Network OCCAM

White Paper

Outcomes and Causal Inference in International Comparative Assessments

November 2021

Preface

About OCCAM

The acronym OCCAM stands for "Outcomes and Causal Inference in International Comparative Assessments" in educational research. OCCAM is a European Training Network (ETN) which is a sub call in the Marie Skłodowska-Curie Innovative Training Networks (MSCA ITN) of the European Commission's Horizon 2020 framework. ETNs are networks which form a structured, international, intersectoral, and interdisciplinary research and training environment for PhD students.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

About this White Paper

The aims of this white paper are to advance knowledge on a) how educational outcomes and quality are measured throughout the world and on b) the educational and political forces that shape learning environments and affect educational outcomes. We aim to publish research related to international assessments like PISA, TIMSS, PIRLS, TALIS, and ICCS including substantive analyses, technical notes, and critical comments. It aims to establish itself as a reliable source for information on comparative education for non-academic audiences (journalists, teachers, unions, and other non-academic stakeholders).

Contents

Part I International Comparisons and Cross-cultural Research

1 Measurement Counts: International Student Tests and Economic Growth 2
Eric A. Hanushek & Ludger Woessmannn

2 Are PISA Top-Performers Also Good Citizens? 5
Rolf Strietholt, Olesya Gladushyna & Andrés Strello

3 Why is it so Hard to Study School Leadership in Cross-cultural Research? 8
Jelena Veletić

4 From Numbers to Meaning Through Comparisons: An Introduction 10
Rolf Vegar Olsen & Sigrid Blömeke

5 From Numbers to Meaning Through Comparisons Between Countries 12
Rolf Vegar Olsen & Sigrid Blömeke

6 From Numbers to Meaning through Comparisons Within Countries Over Time 14
Rolf Vegar Olsen & Sigrid Blömeke

Part II Measurement and Methodology

7 Is Canada Really an Education "Super-power"? The Evidence Is Not as Clear-cut as You Might Think… 18
John Jerrim, Jake Anders, Silvan Häs, Nikki Shure & Laura Zieger

8 Are the Gender Gaps in PISA Influenced by Its Methodology? 21
Laura Zieger & John Jerrim

9 How Can We Explore Similarities Between Countries' Educational Systems? 24
Edwin Cuellar

Part III Outcomes and Its Determinants

10 It's Not Only Whether Children Read That Matters. It's Also What They Read (at Least According to PISA)... 27
John Jerrim

11 Smarter Teachers Produce Smarter Students. This (Partly) Explains the Country Rankings in PISA. 29
Marc Piopiunik

12 Preparing the Ground: Early Learning Opportunities Can Make the Difference 33
Sabine Meinck & Agnes Stancel-Piątak

13 The More Experienced the More... Boring? 36
Pietro Sancassani

14 The Error Climate and Its Influence on Student Performance 39
Andrés Christiansen

Part IV Equity in Education

15 An Unfair Start: UNICEF's Report Card on Educational Inequalities in Rich Countries 43
Ides Nicaise

16 Equity for All in European Education: What Does PISA Say? 46
Cor Sluijter & Remco Feskens

17 Do Girls and Boys Feel Equally Confident in Their Mathematics Abilities? Evidence from TIMSS 2015 49
Ana María Mejía Rodríguez, Hans Luyten, & Martina Meelissen

18 Early Tracking Reinforces Social Inequality 52
Andrés Strello

Part V Technology in the Classroom

19 Does Home Computer Use Improve or Harm Children's Reading Skills? 56
Monica Rosén & Jan-Eric Gustafsson

20 Are Teenagers in England Addicted to Social Media? (And Does It Matter?) 58
John Jerrim

21 Blackboard vs. Computers: It Depends on the Application 61
Oliver Falck & Ludger Woessmann

22 Where are the "Digital Natives"? Results from ICILS 2018 63
Wolfram Schulz & Julian Fraillon

23 Corona and Education: Are Students Really Ready for Digital Schooling? 65
Evi E. Konstantinidou

Part I
International Comparisons and Cross-cultural Research

Chapter 1
Measurement Counts: International Student Tests and Economic Growth

Eric A. Hanushek & Ludger Woessmannn

While some countries such as Germany are avid consumers of PISA data, others such as the U.S. are much less concerned about PISA scores. In most cases, reactions to PISA scores appear largely based on notions of national pride, in where the country falls in the league tables of results. But the scores tell us much more. The PISA scores are a good index of the future quality of the labor force in each country, and the quality of the labor force in turn has been shown to be a decisive factor in determining the long-run growth rates of nations.

Early studies of school attainment and economic growth

For the last quarter century, economists have focused on why some nations grow faster than others. Economic growth is what determines the future economic well-being of societies, and apparently small differences in growth rates have enormous implications for future income levels.

Until recently, this research focused on school attainment – driven both by the extensive work on individual earnings and by the ready availability of such data – in order to characterize differences in skills of workers around the world. Starting around 1990, economists began studying the determinants of long-run growth (e.g., Romer 1990[1]; Barro 1991[2]; Mankiw, Romer, and Weil 1992[3]). While there were some disagreements about how to do this analysis and how to interpret the results, a key point is that virtually all such studies measured skills by school attainment.

Eric A. Hanushek
Hoover Institution, Stanford University, e-mail: hanushek@stanford.edu

Ludger Woessmann
ifo Institute, University of Munich, e-mail: woessmann@ifo.de

[1] https://doi.org/10.1016/0167-2231(90)90028-J
[2] https://doi.org/10.2307/2937943
[3] https://doi.org/10.2307/2118477

Each of these efforts, however, was subject to on-going criticism for leaving out consideration of variation in skills, for ignoring human capital produced outside of schools such as within families, for ignoring any consideration of school quality, and for concerns about issues of causation. These criticisms have proven to be well-founded.

Measuring skills directly makes all the difference

As better data from the international assessments have become available, the importance of educational quality in determining economic wellbeing has become undeniable. Specifically, differences in the long-run growth of nations are closely linked to international test scores (Hanushek and Woessmann, The Knowledge Capital of Nations, MIT Press, 2015[4]). Figure 1.1 shows the relationship between test scores and average GDP growth rates per capita for a selection of countries between 1960 and 2000.

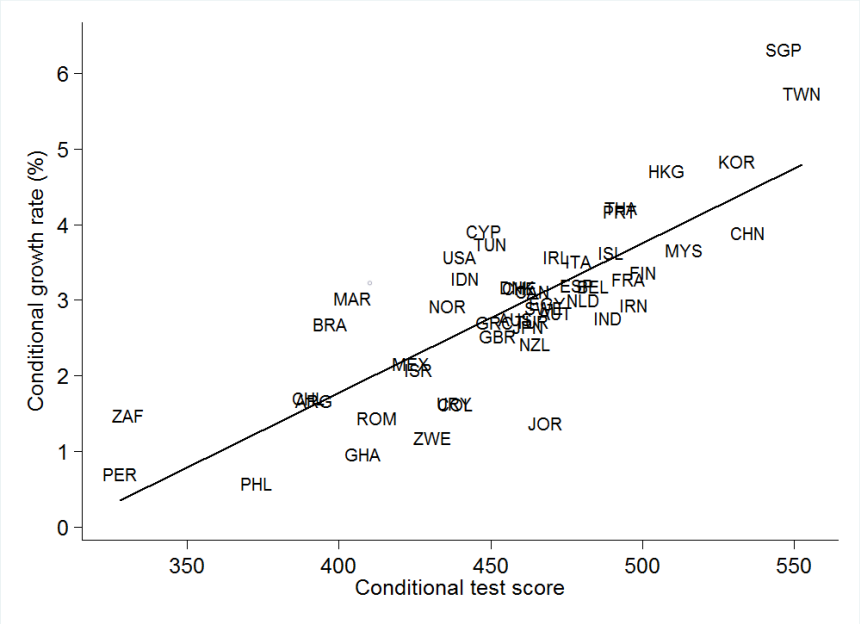


Fig. 1.1 Test scores and economic growth rates across countries
Note. Plot of a regression of average annual rate of growth (in percent) of real GDP per capita between 1960-2000 on average international student achievement test scores, average years of schooling in 1960, and initial level of real GDP per capita in 1960 (mean of unconditional variables added to each axis). Source: Hanushek and Woessmann (2015)

[4] https://mitpress.mit.edu/books/knowledge-capital-nations

In short, growth is tightly tied to the skills of a country's population, and international tests such as PISA and TIMSS do a good job at measuring these important labor-force skills. Moreover, once measures of what students know are put into the growth models, years of school attainment offer no additional explanatory power.

Measurement matters. This shifts the policy focus immediately from just school attainment to the more important issues of school quality.

Chapter 2
Are PISA Top-Performers Also Good Citizens?

Rolf Strietholt, Olesya Gladushyna & Andrés Strello

The education community across the globe is anticipating the release of the latest PISA rankings. These gauge the progress being made by 15-year-olds in reading, science and mathematics across a wide selection of developed and developing countries. One of the widely-acknowledged limitations of PISA as a measure of educational quality is that it covers only a limited range of outputs from an education system. There is hence a threat that, if countries focus efforts upon maximising performance in the PISA tests, then they may divert attention away from important (yet unmeasured) contributions of education to wider society. This includes, for instance, the development of morals and civic engagement. Indeed, in this blog we provide new evidence on the discrepancy between the performance of countries in PISA and children's knowledge of civics.

Scant Rankings Provide a Fragmentized Picture

To understand why PISA provides a too simplistic picture of educational quality, it is worth reconsidering the role of education for society. In democratic societies, there is a general consensus that schools must not only prepare children for the workplace, but also to promote social values and active citizenship.

PISA focuses on how education contributes to economic growth and development. For this purpose, the OECD, organization that administers PISA, assesses math, reading, and science literacy which are considered by OECD experts to be the key areas of human capital. Even though knowledge in these academic domains

Rolf Strietholt
TU Dortmund, e-mail: rolf.strietholt@tu-dortmund.de

Olesya Gladushyna
TU Dortmund, e-mail: olesya.gladushyna@tu-dortmund.de

Andrés Strello
TU Dortmund, e-mail: andres.strello@tu-dortmund.de

is important for other purposes of education, active citizenship requires further understanding about issues such as human rights, freedom of the press, or financial donations to political parties. In the ”post-truth” era when radical populist movements are on the rise worldwide, such civic education is becoming an increasingly important issue. Yet, such knowledge is not well represented by PISA.

Comparing Student Performance in Mathematics and Civic Education

There are, however, alternatives to PISA where the measurement of children's civic knowledge and attitudes is a central goal. The most widely known cross-national study in this area is the International Civic and Citizenship Education Study (ICCS). Hence, by drawing upon data from both PISA and ICCS, one can see how the academic and civic knowledge of young people across different countries compare. We illustrate such a comparison in Figure 2.1, which compares average PISA 2015 mathematics scores (horizontal axis) to average scores on the civic knowledge test of ICCS 2016. This comparison has been done for all European and East Asian that participated in both assessments (Latin American countries were excluded because they frequently perform much lower on international assessments).

This figure clearly illustrates that there is huge variation in civic knowledge amongst countries from a similar geographical, historical and cultural space – and who have children with similar PISA test scores. For instance, Hong Kong, Korea, and Chinese Taipei are all PISA top-performers in mathematics, but their results in civic knowledge are up to 66 test points apart (equivalent to more than one additional year of schooling). Similarly, although Danish, Finnish, Norwegian, and Swedish

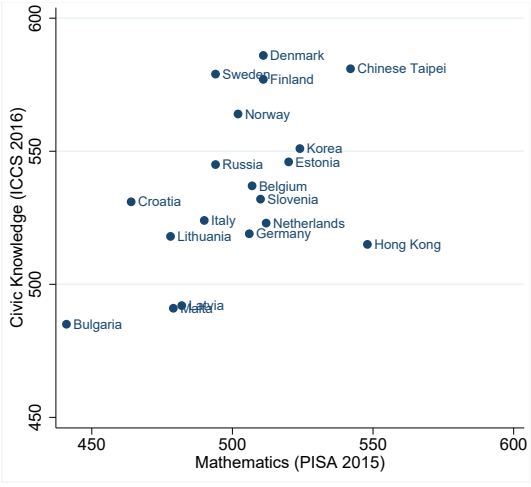


Fig. 2.1 Country mean performance in international assessments on mathematics and civic education (Pearson's r=.51).

children achieve around the international average in mathematics, they have very high levels of civic knowledge. Indeed, in this respect, it is Scandinavia (rather than East Asia) where children lead the world.

The PISA Poster Child: A False Promise

Different assessments lead to different conclusions about educational performance. Consequently, it would simply be naive to believe that the outputs of an education system (or any attempt to measure its ”quality”) could be compressed into a single (or small number of) scores. Nevertheless, this is exactly how the PISA rankings are often interpreted, not least by politicians and media when the results are released. Hence, with the release of the PISA 2018 results just around the corner, we urge that caution needs to be exercise and how it is vital for readers to look behind the headlines. In particular, calls to copy the education systems of the highest PISA performers should be avoided. Instead, it is worth remembering that the outputs of education systems are multifaceted – and that countries who perform highly in one domain may not do so in another.

Chapter 3
Why is it so Hard to Study School Leadership in Cross-cultural Research?

Jelena Veletić

Evidence shows that some of the key aspects of educational policies, like school leadership, are very sensitive and shaped by the culture and context where they are executed. Nevertheless, we frequently want to compare and contrast such concepts across different cultures and contexts. To do so, our instrument (e.g. test, questionnaire) must operate equally across different groups. In more simple words, we must ensure that our instrument measures the same thing in all groups of interest. This is called measurement invariance. You may wonder why it is important to establish measurement invariance. Well, consider some of the challenges we faced in our study of school leadership.

The study

Our study focuses on a specific school leadership conceptualization – instructional leadership. Strong instructional leaders are principals, who emphasise high-quality instruction, give instructional feedback to teachers and support the use of assessment in the classroom. As previously mentioned, leadership is enacted within a context of culturally embedded values and worldviews, and comparisons between different groups may be challenging. Therefore, we use data from the Teaching and Learning International Survey (TALIS) from a set of Nordic and Baltic countries sharing geographical, cultural and linguistic characteristics. Although the selected countries are different in many respects, the selection is relevant because the Baltic countries have been influenced by their Nordic neighbours over the recent decades. However, it turned out that modelling instructional leadership, even in seemingly homogenous countries, was far from being straightforward.

Jelena Veletić
Centre for Educational Measurement, University of Oslo, e-mail: jelena.veletic@cemo.uio.no

What was challenging?

The different interpretation of feedback in diverse countries led to an inability to compare models across them. Giving feedback is an important element in the process of managing instruction in schools, which is an important dimension of instructional leadership. Across countries, does that mean that effective leadership assumes a certain amount of feedback at a certain time? Does it assume that feedback is equally perceived across groups? A simple first look at the TALIS data reveals that teachers in some countries receive significantly more feedback than in others. However, when modelling these responses as a latent variable, it becomes clear that the interpretation of feedback differs across countries. In more technical terms: the measure was not invariant. In our case, this invariance appeared as a cross-cultural difference between the Nordic and Baltic countries. Several cultural dimensions could be relevant to understand why this is so. Some cultural values, such as power distance and collectivism, might be relevant to explain the differences commonly seen between eastern and western countries. Power distance refers to the extent to which the community accepts and endorses authority. For example, in high power distance cultures, feedback and interactions between different hierarchical levels (in the context of school: principals, teachers, students) may be regarded as activities where power and authority are established and maintained. As such, feedback received from principals might connote control or monitoring. Conversely, in low power distance countries, feedback might be welcomed and represent support and a mechanism where leaders and teachers collaborate and engage in joint meaning-making. Furthermore, in individualism-oriented countries, in contrast to more collectively oriented countries, feedback appears to be necessary to ensure that each individual's efforts work in unison with the larger aims and goals of the organisation (i.e., the school). Thus, feedback as one of the instruments of instructional leadership practice appears to differ across these cultures in its sign or value direction (positive-negative), perceived reliability and accuracy as well as the form.

Sensitivity to such differences is essential in making conclusions and giving adequate policy recommendations. Otherwise, any differences we observe between groups may be due to the instrument used and not due to meaningful differences in the constructs of interest.

Chapter 4
From Numbers to Meaning Through Comparisons: An Introduction

Rolf Vegar Olsen & Sigrid Blömeke

Large-scale international assessments studies (ILSA) are regarded as important sources for monitoring educational quality in many countries. The results from the studies are frequently cited in policy documents and they are regularly used as warrants or rebuttals in political debates. Over the last 20 years or so, ILSAs have thus established themselves as powerful knowledge sources. In this series of four blog-posts we will present and discuss two of the main reasons for why these studies have gained this position: they support interpretations of findings from two comparative perspectives – comparisons between educational systems and comparison within one system over time. In the fourth and final post we will identify and discuss some of the challenges for these two comparative perspectives for future ILSAs.

Quantifying education

Measuring qualities in education is hard because the phenomena we are trying to capture are not directly observable. Usually we are interested in making inferences about psychological attributes of persons – how motivated are students to learn in school, how satisfied are teachers with their work environment, or how proficient are students in reading, to mention a few. These attributes cannot be observed or measured directly. We do not have a "reading-proficiency-ruler" or some electronic sensor, which immediately can report a value on a scale. Instead, we have to rely on indirect procedures where one or (usually) several observations of a person are used to establish numbers on a scale. Fortunately, substantive theory in collaboration with test theory (or "psychometrics") provides us with powerful tools to develop reliable

Rolf Vegar Olsen
Centre for Educational Measurement, University of Oslo, e-mail: rolfvo@cemo.uio.no

Sigrid Blömeke
Centre for Educational Measurement, University of Oslo, e-mail: sigrid.blomeke@cemo.uio.no

and valid measures or indicators of such phenomena in the social and psychological realm.

Nevertheless, even if measures have documented good anchoring in substantive theory and psychometric quality, the numerical values themselves are usually hard to interpret, e.g. as exemplified through questions like:

- What is an acceptable level of achievement for a system?
- What value on a bullying scale represents a level of concern?
- At which values should we conclude that a regression coefficient, correlation coefficient, difference between two groups etc. is substantively meaningful?

International comparisons provide a frame of reference

Answers to questions like these are largely normative or political because we are not only lacking substantive theory for deriving answers of this kind, we also may disagree normatively: Someone can therefore just stipulate thresholds and try to add meaning to them. However, politically decided or stipulated thresholds also need backing, and ILSAs provide several ways of creating a more rational basis for such backings by allowing for comparisons.

The choice of a comparison, a criterion or norm, is not a neutral activity either. Rather, it forms the argumentative core or warrant for the interpretations to be made of the data. The following blog posts will discuss how comparisons with other countries aid interpretation of data from ILSAs, but also why such comparisons can be deceiving and methodologically flawed.

Chapter 5
From Numbers to Meaning Through Comparisons Between Countries

Rolf Vegar Olsen & Sigrid Blömeke

In an earlier blog[1], we established that international large-scale assessments can be regarded as powerful and influential knowledge sources for making claims about the quality of educational systems. Comparisons between countries have over time been one of the most dominant ways of interpreting the results coming out of the studies, and this blog post continues by going into further detail about such between country analyses.

The World is an Educational Laboratory

"We, the researchers who… decided to cooperate in developing internationally valid evaluation instruments, conceived of the world as one big educational laboratory where a great variety of practices in terms of school structure and curriculum were tried out. We simply wanted to take advantage of the international variability with regard both to the outcomes of the educational systems and the factors which caused differences in those outcomes.", Torsten Husén, 1973, in the report from FISS, one of the first international large-scale studies in education. This quote captures the essence of why comparisons with others is perceived to be useful:

- Comparisons with other systems helps by providing cases of what is typical across countries. This provides a normative rhetorical framing for statements like "Norwegian 4th graders read more newspapers than 4th graders do on average in other countries".

Rolf Vegar Olsen
Centre for Educational Measurement, University of Oslo, e-mail: rolfvo@cemo.uio.no

Sigrid Blömeke
Centre for Educational Measurement, University of Oslo, e-mail: sigrid.blomeke@cemo.uio.no

[1] https://international-education.blog/en/from-numbers-to-meaning-through-comparisons-an-introduction/

- They provide exemplary cases of what is possible. That is a rhetorical frame for benchmarking, e.g. in statements like "Compared to the highest performing systems in the world, Norwegian 15-year olds lag on average several years of schooling behind".
- International comparisons provide information about phenomena which are somewhat invisible or unobservable within one country alone because there is no or very little variation within one educational system. However, the variation between countries can be substantial. One example is school starting age, which is relatively fixed within many countries but varies across countries.

Using a metaphor from photography: The international variation creates contrast and a background allowing the main object to be highlighted. Initially arbitrary numbers are given meaning through relative comparisons.

Standardization and Quality Management

This "relative" framing has obvious shortcomings, mostly related to the range of countries participating and the limitation that the framing is based on the assumption that the comparisons are relevant and accurate. Or as so often stated: we have to make sure that we are "comparing oranges with oranges, and not with apples". International assessments put a great deal of effort to ensure that comparisons between countries can be made. They have strict rules for sampling of schools and students to ensure comparability across countries, the tests and questionnaires are piloted in all countries, translations of the instruments are verified in several steps, and the test and questionnaire items are empirically checked to ensure that they function equally across countries - to mention just a few of the quality checks being done. Nevertheless, the quality with which we can make valid comparisons between countries is never perfect, and can be potentially quite misleading. As the number and heterogeneity of countries in these studies increases, this challenge of ensuring fair comparisons becomes ever harder. We will return to this and other threats and challenges to interpretations of comparisons across countries in another blog.

Chapter 6
From Numbers to Meaning through Comparisons Within Countries Over Time

Rolf Vegar Olsen & Sigrid Blömeke

Most of the international large-scale assessments are repeated in regular intervals. PISA is conducted every three-years, TIMSS every fourth year, and PIRLS is conducted with five-year intervals. Accordingly, this allows for comparisons within countries over time, with the objective to uncover patterns or trends and to predict future development. The achievement scores are linked over time by having a relatively large number of test questions that are repeated. This makes it possible to anchor subsequent test scores with the previous ones. In addition, sections of the same background questionnaires are repeated over time to also capture changes in the learning context, demographics etc.

The possibility to compare education systems over time overcomes one of the challenges of comparisons between countries (see our previous blog[1]): Comparisons over time are made within the same system, and hidden cultural or other non-observed differences can be regarded as controlled for.

Figure 6.1 provides an example where all average achievement results for Norway in PIRLS, PISA and TIMSS over 20 years from 1995 to 2015 are collected into one picture.

The three studies included cover three broadly defined domains. The figure illustrates the time series in reading with red lines, mathematics with blue lines and science with green lines. The figure also captures three different populations. The trend lines for the 15-year-olds are shown by solid lines, 8th graders with dashed lines and 4th graders with dotted lines. All the three studies report the achievement results on a scale where the international average in the first year of the series is set to 500 (and one standard deviation is set to 100). And here is the obvious weakness of the figure: Even if all the studies make use of what appears to be the same scale,

Rolf Vegar Olsen
Centre for Educational Measurement, University of Oslo, e-mail: rolfvo@cemo.uio.no

Sigrid Blömeke
Centre for Educational Measurement, University of Oslo, e-mail: sigrid.blomeke@cemo.uio.no

[1] https://international-education.blog/en/from-numbers-to-meaning-through-comparisons-between-countries/

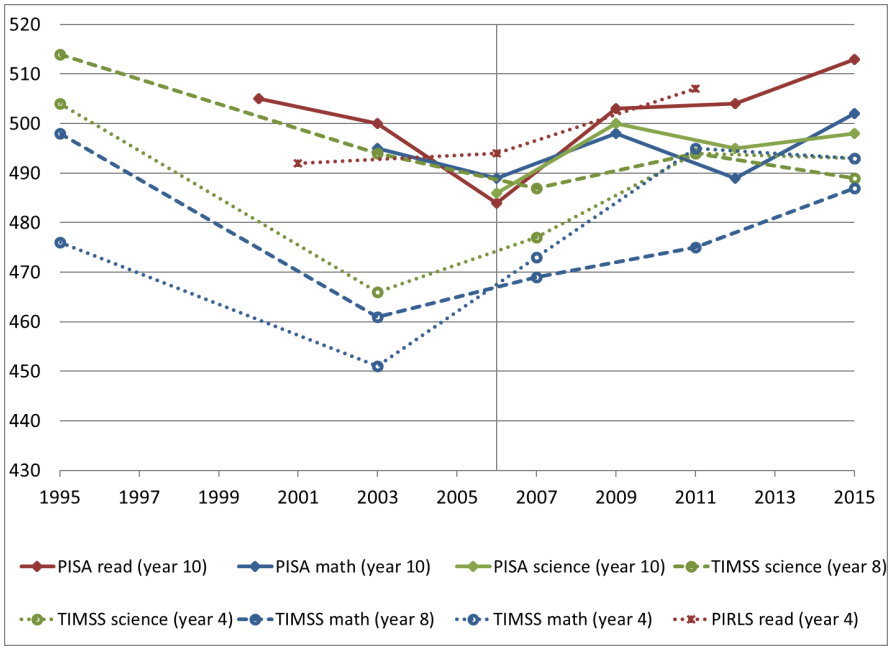


Fig. 6.1 Achievement scores for Norwegian students for three ILSAs, in three domains for three student populations in the period 1995-2015. Take care not to interpret differences between the lines – they are largely meaningless (see text).

direct comparisons cannot be made across the lines in the figure. The studies are not formally linked with each other, and the international average reflects a different composition of countries for each of the studies (e.g. OECD countries participate in PISA but only some of them in TIMSS and PIRLS). Nevertheless, a more holistic interpretation of features in the figure reveals a rather consistent feature of Norwegian students' development in the period:

- There was a huge decline in performance in the first half of the period, no matter which domain or student age we are talking about. For some subjects and studies the decline was close to 40 points on the scale (or 0.4 of a standard deviation). Another way of stating the same finding is that students who started their schooling in the mid- to late-1990ies for some reasons performed much worse than the previous cohorts of students.
- There is an almost equally strong trend of improvement in the last half of the 20-year period–despite a doubling of the number of students with immigrant backgrounds during this period. In particular, the improvement for the 4th grade students largely make up for the decline in the early period, but the 8th grade students are still lagging a little bit behind in 2015 compared to 1995.

This figure, and the complexities of results they represent, were used in an evaluation of an educational reform in Norway. This reform was labeled as the "Knowl-

edge Promotion" to reflect the fact that improving students' performance across all ages and many domains by strengthening basic knowledge acquisition and clearly defining learning outcomes was one of the major ambitions of the new policy. A large research-based evaluation was organized, but none of these studies could address the issue of whether or not this major ambition of the reform had been realized. Fortunately, data from the international studies were available and could be used to describe how students learning outcomes had changed in the past 20 years. A more detailed analysis of one of these changes (TIMSS 8th grade mathematics from 2003 to 2015) showed that the most important factors related to the positive change was an improved learning environment and school climate.

The above example illustrates the potential usefulness of the time series reported by the international large-scale studies. Not surprisingly, the trend series are increasingly emphasized in the national and international reporting from the studies. Furthermore, the time series design opens up for other ways of analyzing data than merely describing trends or patterns within countries. Features at the system level can be studied by so called differences-in-differences analyses, where changes in a predictor (e.g. a change in classroom size) are related to changes in an outcome (e.g. mathematics achievement) at the country level. Such analyses benefit from the same methodological advantages like panel analyses where individuals are observed repeatedly.

Part II
Measurement and Methodology

Chapter 7
Is Canada Really an Education "Super-power"? The Evidence Is Not as Clear-cut as You Might Think...

John Jerrim, Jake Anders, Silvan Häs, Nikki Shure & Laura Zieger

When the Programme for International Student Assessment (PISA) results are released every three years, it is now little surprise that a set of East Asian nations (e.g. Singapore, Taiwan, Japan, South Korea) dominate the top spots in these rankings. These nations typically substantially outperform most English-speaking Western nations, with one important exception – Canada. This has not gone unnoticed by policymakers and the education media. Indeed, after the release of the PISA 2015 results, Canada was described as an "education superpower"[1] with various theories (from the strong academic performance of immigrants through to high levels of student motivation) put forward to explain this result. Indeed Andreas Schleicher – the man who has led the OECD's PISA programme – suggesting that the strong commitment to equity in Canada is the key.

John Jerrim
University College London, e-mail: j.jerrim@ucl.ac.uk

Jake Anders
University College London, e-mail: jake.anders@ucl.ac.uk

Silvan Häs
University College London, e-mail: s.has@ucl.ac.uk

Nikki Shure
University College London, e-mail: nikki.shure@ucl.ac.uk

Laura Zieger
University College London, e-mail: l.zieger@ucl.ac.uk

[1] https://www.bbc.co.uk/news/business-40708421

But how much confidence can we really have in the Canadian PISA results?

One of the key pillars of PISA is meant to be that it is representative of each country's 15-year-old population. If this is not achieved, then we are not comparing like-with-like. For instance, if country A were to disproportionately exclude some groups of students, then it cannot be fairly compared to country B where a representative cross-section of young people did actually take part. This situation could emerge if, for instance, children with Special Educational Needs (SEN) are identified and treated differently across countries. Alternatively, in some countries, a significant number of students and schools may refuse to participate in the study.

The reality is that this is what happens in PISA – and we believe could substantially undermine the Canadian results.

This point is illustrated in Figure 7.1, which draws upon figures reported in the PISA 2015 technical report. Clearly, the figures for Canada are striking. Only around half (53%) of 15-year-olds in Canada were covered within the PISA 2015 assessment. This compares to more than 90% of 15-year-olds in Japan and South Korea.

Why do these figures for Canada look so bad? There is a mix of reasons.

First, schools in Canada were more likely to refuse to take part than schools in other countries, with the Canadian national report flagging particular issues within Quebec (where less than half of those schools approached agreeing to take part, see

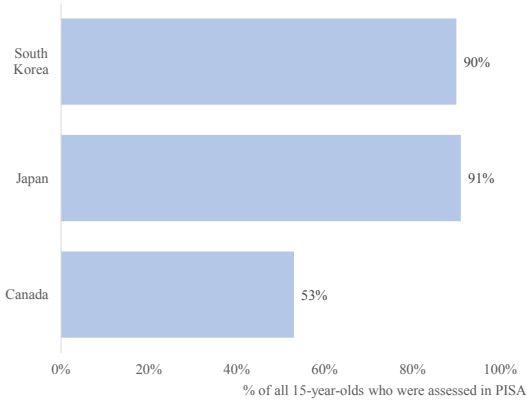


Fig. 7.1 The number of 15-year-olds in Canada, Japan and South Korea and the (weighted) number covered within the PISA assessment[3].

[3] Source: http://www.oecd.org/pisa/sitedocument/PISA-2015-Technical-Report-Chapter-11-Sampling%20Outcomes.pdf

Table A2)[4]. If it certain types of school (e.g. those with lower performing students) are less likely to take part than others (e.g. higher-performing schools), then this could lead to an upward bias in the Canadian PISA results. Second, Canada was much more likely to exclude pupils from taking the PISA assessment due to issues such as Special Educational Needs (7.5% of 15-year-olds were excluded in Canada compared 2.4% in Japan and less than 1% in South Korea) – a group who are likely to be very low achievers.

Finally, students in Canada were less likely to actually sit the PISA assessment - even within schools that agreed to take part. Specifically, the official figures show that almost 20% of Canadian teenagers were counted as absent on the day of the PISA test compared to less than 3% of those in Japan and South Korea. It is well-known that certain types of student (e.g. lower achievers from lower socio-economic backgrounds) have higher absence rates from school and these characteristics are likely to be associated with performance on the PISA test. It hence seems likely that this would lead to an upward bias in the results.

Together, this adds up to a significant problem, which we believe significantly undermines our confidence in the PISA 2015 data for Canada. We believe that there are particular problems in drawing comparisons to other "high-performing" countries – Japan and South Korea in our example – where a genuinely representative cross-section of children took part.

Indeed, after scratching below the surface, evidence of Canada being an "education superpower" does not seem to be particularly strong at all.

[4] https://www.cmec.ca/publications/lists/publications/attachments/365/pisa2015-cdnreport-en.pdf

Chapter 8
Are the Gender Gaps in PISA Influenced by Its Methodology?

Laura Zieger & John Jerrim

PISA measures 15-year-olds achievement scores in mathematics, reading and science and has evolved into a powerful tool in politics, as the scores can be compared across countries and over time. Apart from the scores themselves, there is also significant interest in the achievement differences between boys and girls.
What most people do not know about PISA is that not every child is directly assessed in all three subjects. Moreover, roughly 60% of the students only answer questions in two of the three subjects. Yet, everyone is assigned an achievement score in mathematics, reading and science.
In order to allocate test scores to all students, predictions are made from the subjects where children actually answered questions and their background characteristics – including gender. This is known as 'conditioning' in the academic literature and is considered vital to correctly estimating gender gaps in achievement. Indeed, psychometricians argue that gender differences will tend to be underestimated unless this 'conditioning' takes place.

How does conditioning affect the PISA gender gaps in reality?

In theory, as soon as gender is included in the conditioning model, we should correctly estimate the difference in achievement between boys and girls. But theory and reality can be different things. For this reason, we have investigated the conditioning model[1] used in PISA 2012. As part of this project, we computed three alternative versions of the students' scores in each of the PISA subjects. In the first model, the scores for each student are inferred just from the test responses in the different

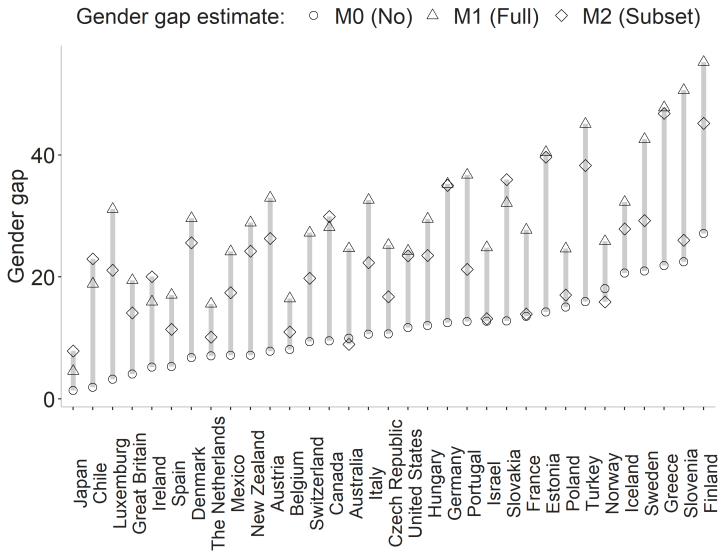
Laura Zieger
University College London, e-mail: l.zieger@ucl.ac.uk

John Jerrim
University College London, e-mail: j.jerrim@ucl.ac.uk

[1] https://repec-cepeo.ucl.ac.uk/cepeow/cepeowp20-09.pdf

domains (no conditioning; M0). In the second model, we used the responses to the PISA test questions and all background variables (full conditioning; M1). In the last model, test responses were combined with just a subset of the background variables (gender, grade, mothers and fathers socio-economic index and booklet IDs) to test the sensitivity of the results (conditioning on subset; M2).

The figure below shows the gender gap in reading using model M0 (no conditioning - circle), M1 (full conditioning - triangle) and M2 (conditioning on subset - diamond). We expect a big difference between no conditioning and both versions of conditioning, while M1 and M2 should be very similar. And indeed, for most countries, the triangle (M1) and the diamond (M2) are pointing in the same direction, and for about a third of countries, they even sit on top of each other. This suggests that, in most countries, the gender gap is not sensitive to the exact specification of the conditioning model (once gender has been included) with a potential small increase or decrease when more variables are included. There are, nevertheless, some important changes to the results for some individual countries (that are somewhat difficult to explain). For instance, in Australia, Israel, France, Poland, Slovenia, and Norway the estimated gender gap from M0 and M2 are similar. Yet there is a large jump in the magnitude of the gender gap in M1.



What does this mean?

Our research led to two conclusions for gender gaps: First, the gender gaps amplify to a less biased estimate as soon as gender is included in the conditioning model. Second, theoretically, the exact specification of the included background variables should not matter as long as gender is included. While this holds true for the majority of countries, our research shows that some countries experience delicate impacts on their gender gap which are also reflected in changes in the ranking of those countries. Summarising we can say that regarding gender gaps, it matters if and which background variables are used in the computation of the students' scores.

Chapter 9
How Can We Explore Similarities Between Countries' Educational Systems?

Edwin Cuellar

Beyond averages: response patterns in achievement tests

PISA, TIMSS, ICCS, and other international educational assessments have become more and more popular over the last 20 years as a means of comparing educational achievement worldwide. These assessments include dozens of questions that aim to measure students' level of proficiency in certain skills and competences. These studies then use mathematical models to transform students' responses into scores. The logic behind the scoring is simple: If a student correctly answers a more difficult question, his/her score increases more than when the student correctly answers an easier question. The difficulty of each question is, of course, relative to the other questions and to the students' abilities.

A topic of great relevance is whether these patterns of difficulties are similar across countries or they change substantially, and what those differences and similarities mean in terms of education. For instance, if we focus on mathematics, we may find that geometry questions are easier for students in some countries and more difficult for others, signaling a substantive difference in terms of curriculum, instruction, learning processes, etc. Note that this is a different outcome of educational assessments: Two countries may have the same score in mathematics, but their patterns of difficulties could be completely different.

International similarities in response patterns

In a recent paper[1], we discussed a wide variety of alternative ways of exploring and visualizing these patterns across countries, using PISA 2012 for illustrative pur-

Edwin Cuellar
Research and Innovation Team, Cito, e-mail: edwin.cuellar@cito.nl

[1] https://rdcu.be/ce0Br

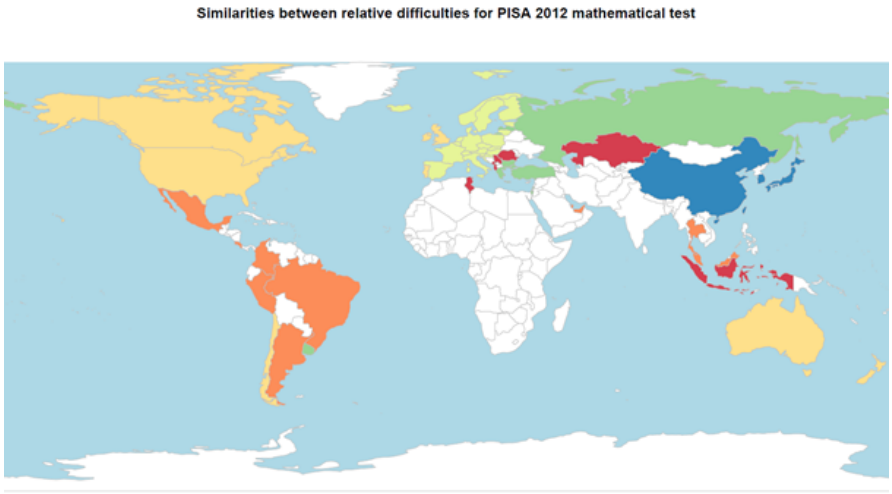


Fig. 9.1 Relationship between error climate, attitudes towards mathematics, and achievement in mathematics (coefficients are standardized).

poses. Our approach is simple: We estimate the difficulties for each country separately and then compare them, taking some technical considerations into account. In particular, we take special care to analyze the difficulties as relative values. All of the analyses are based on the idea of minimizing the distances between these relative difficulties across groups. We discuss how this approach connects to a widely known phenomenon in educational testing called differential item functioning (DIF), which is especially challenging when there are many groups. We present some multivariate techniques that can be easily replicated and enable cheap and fast computation.

For instance, we discuss the use of clustering techniques, which allows us to explore clusters of countries whose patterns of difficulties are internally similar and heterogeneous with other clusters. The figure below shows what the cluster of countries looks like for the mathematics test in PISA 2012, with each color corresponding to a cluster of countries:

In the paper, we also show how this clustering analysis can be performed simultaneously for questions and countries, so we can study subsets of questions whose difficulties are invariant across clusters of countries. With this approach, we unveil substantive similarities between countries at the item level that merit more attention. At the same time, we take a different approach to investigate measurement non-invariance (i.e., DIF) in large-scale assessments. We believe this insight may be useful in deepening the understanding of test scores and exploring the differences and similarities between participating countries.

Part III
Outcomes and Its Determinants

Chapter 10
It's Not Only Whether Children Read That Matters. It's Also What They Read (at Least According to PISA)...

John Jerrim

It is widely considered important that young people read regularly. A wide range of previous research has linked reading during childhood to improved language skills and higher levels of academic achievement more generally[1].

But does it matter what they choose to read? Does flicking through a magazine or reading a daily newspaper have the same benefits for young people as becoming engrossed in a novel? A lot less evidence currently exists about this issue.

In my recent research paper [2] – published earlier this year – my co-author Gemma Moss [3] and I decided to explore this topic in detail.

The data we used was drawn from the OECD's Programme for International Student Assessment (PISA) study. This asked 15-year-olds how frequently they read the following different text types:

- Fiction books
- Non-fiction books
- Newspapers
- Magazines
- Comic books

Within our paper, we examine how the frequency young people read each of these different types of materials is linked to young people's PISA scores. Critically, our analysis controls for a wide array of other factors (such as gender, socio-economic status and school attended) to try and rule out alternative explanations for our results.

A summary of our key findings can be found in Figure 10.1. This illustrates how PISA scores differ between teenagers who read each text type regularly (i.e. almost every day) versus those who almost never read that type of text.

John Jerrim
University College London, e-mail: j.jerrim@ucl.ac.uk

[1] https://onlinelibrary.wiley.com/doi/10.1002/berj.3180

[2] https://onlinelibrary.wiley.com/doi/abs/10.1002/berj.3498?af=R

[3] https://iris.ucl.ac.uk/iris/browse/profile?upi=PJGMO52

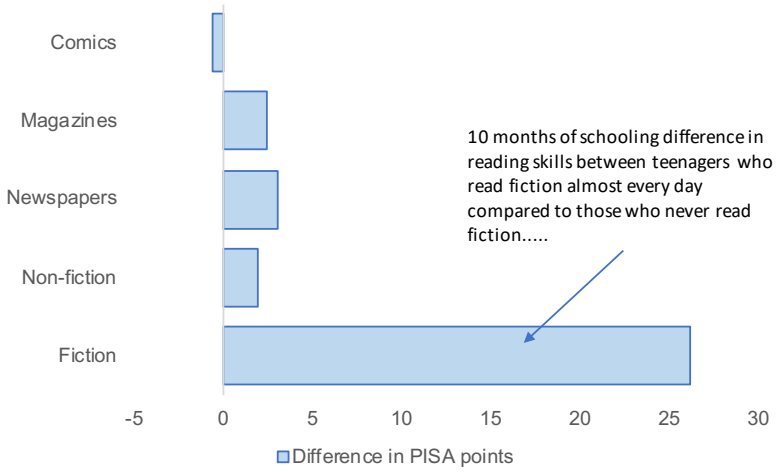


Fig. 10.1 The advantage of reading different text types for teenagers reading skills.
Note: Ten PISA test points is roughly equivalent to an effect size of 0.1 or around four months of additional schooling. Results refer to difference between teenagers who read the text type almost every day compared to those who read the text type almost never.

As this graph illustrates, teenagers who frequently read newspapers, magazines, comics and non-fiction books do not achieve higher PISA reading scores than those who do not. Yet the same is not true for young people who read fiction books or novels. Specifically, teenagers who read fiction almost every day score around 26 points higher on the PISA reading test than those who never read such books. This difference in achievement is large – the equivalent of around 10 months of additional schooling according to the OECD.

Is this result simply due to young people who read fiction books just reading for a greater amount of time in total? After all, dipping in and out of a magazine, comic or newspaper takes a lot less time than trying to slog through a novel like War and Peace. Interestingly, our analysis provided no evidence that this was the case. We continued to find just as strong "fiction effect" even after we controlled for young people's total weekly reading time.

This finding has important implications. Parents and teachers should not encourage teenagers to "just read something", no matter what this is. Rather, they should focus their efforts on encouraging young people to engage more with novels and other lengthy fictional texts that encourages deep reading for sustained periods of time. This is likely to be particularly important for boys from lower socio-economic backgrounds, the group we find to be reading this type of text the least, and who also have comparatively poor reading skills.

Chapter 11
Smarter Teachers Produce Smarter Students. This (Partly) Explains the Country Rankings in PISA.

Marc Piopiunik

The OECD's Programme for International Student Assessment (PISA) continually illustrates how the reading, science and mathematics performance of 15-year-olds differs across countries. Differences in teacher quality are commonly cited as a key determinant of these huge differences. In a recent study, my co-authors Eric A. Hanushek (Stanford University), Simon Wiederhold (KU Eichstätt-Ingolstadt), and I use internationally comparable data on teacher cognitive skills – one important dimension of teacher quality – to investigate this claim.

Teacher skills differ strongly across developed countries

We use unique data from the OECD's Programme for International Assessment of Adult Competencies (PIAAC) to compute - for the first time - consistent measures of teachers' skills in numeracy and literacy across 31 countries. Figure 11.1 shows that teachers' skills differ widely. For example, average numeracy and literacy skills of teachers in countries with the lowest measured skills in our sample (Chile and Turkey) are well below the skills of employed adults with just vocational education in Canada. In contrast, the skills of teachers in countries with the highest measured skills (Japan and Finland) exceed the skills of adults with a master's or PhD degree in Canada.

Marc Piopiunik
ifo Institute, e-mail: piopiunik@ifo.de

Teacher skills partly explain differences in student performance across countries

We relate these country-level teacher skill measures to individual-level data on student performance in math and reading from PISA. Figure 11.2 shows that the cross-country differences in teacher cognitive skills partly explain international differences in student performance. For instance, students in Italy and Russia would experience an increase in math performance equivalent to the learning progress of an entire school year if their teachers were brought up to the skill level of teachers in the highest-performing country (Finland); students in Spain, the United Kingdom, and the United States would experience an improvement of about three-quarters of a school year. Bringing teachers in each country to the Finnish level would reduce the international gap in PISA scores by about one quarter.

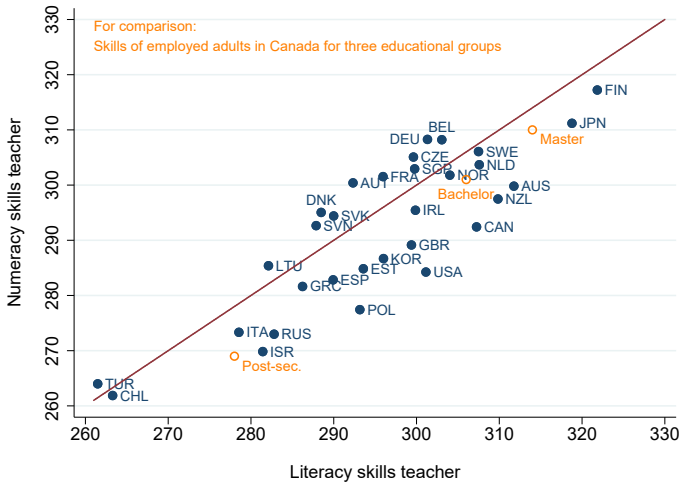


Fig. 11.1 Teacher Cognitive Skills Across Countries (Compared to Canadian Workers)
Note. Solid dots indicate country-specific median teacher skills in numeracy and literacy. Hollow orange circles indicate the median cognitive skills for three educational groups of employed adults aged 25–65 years in Canada: Post-sec. includes individuals with vocational education (postsecondary, nontertiary) as their highest qualification, Bachelor includes individuals with a bachelor degree, while Master includes individuals with a master or doctoral degree. Data sources: PIAAC (2012, 2015).

Can other country factors explain the relationship between student performance and teacher skills?

Critically, our analysis controls for a wide array of other factors to rule out alternative explanations for our results. Most importantly, we account for all differences between countries that are constant across subjects (e.g., educational spending, average cognitive skills of the population, and general cultural background such as importance of education) by relating differences in teachers' cognitive skills between numeracy and literacy to differences in students' performance between the same two subjects. This method yields similar results as when relating teachers' numeracy skills to students' math performance, or teachers' literacy skills to students' reading performance separately.

Additional analyses strongly support the interpretation that our findings indeed reflect the impact of teachers' skills. When we relate student performance to the cognitive skill levels in other broad occupations assessed in PIAAC (e.g. managers, scientists and engineers, health professionals, and business professionals), we find no systematic relationship. Only the skills of teachers are found to be consistently related to students' performance in PISA.

What can policymakers do to improve the cognitive skills of teachers?

Our international data also allow us to investigate how policy choices affect the skills of the teacher workforce and ultimately, student outcomes. Importantly, we find that international differences in wage premiums paid to teachers (given their gender, work experience, and cognitive skills) are directly related to teachers' cognitive skills. These results speak to the potential value of increasing teacher pay.

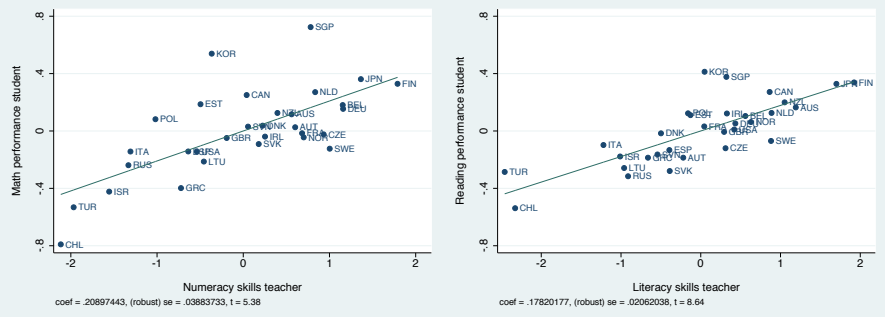


Fig. 11.2 Student Performance and Teacher Cognitive Skills
Note. Each circle indicates a country, and the solid line indicates the best linear fit. Data sources: OECD, PIAAC (2012, 2015), and PISA (2009, 2012)

However, policymakers will need to do more than raise the pay of teachers across the board to ensure positive results. They must ensure that higher salaries go to the most effective teachers.

Chapter 12
Preparing the Ground: Early Learning Opportunities Can Make the Difference

Sabine Meinck & Agnes Stancel-Piątak

For decades, educational researchers have found that children from disadvantaged families often lack access to stimulating learning activities[1] [2]. Back in 1995, a famous study called "The early catastrophe. The 30 million words gap" by Hart & Risley[3] uncovered a large gap in language exposure of toddlers between high and low-income families. At the same time, a stimulating learning environment at home or in childcare settings is important for child development and later schooling outcomes.

Capturing information on early learning opportunities

TIMSS and PIRLS[4], two large-scale educational assessments of IEA[5], include an "Early Learning Survey"[6] [7]. Parents of fourth-grade students are asked how often[8] they engaged with their children in early learning activities such as reading books, telling stories, counting things, or playing with construction toys, before they entered school.

Sabine Meinck
IEA Hamburg, e-mail: sabine.meinck@iea-hamburg.de

Agnes Stancel-Piątak
IEA Hamburg, e-mail: agnes.stancel-piatak@iea-hamburg.de

[1] https://www.sciencedirect.com/science/article/abs/pii/S0049089X03000772
[2] https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8624.2006.00911.x
[3] https://eric.ed.gov/?q=The+Early+Catastrophe.+The+30+Million+Word+Gap.&id=EJ672461
[4] https://timss.bc.edu/
[5] https://www.iea.nl/
[6] https://timssandpirls.bc.edu/timss2015/questionnaires/downloads/T15_HQ_4.pdf
[7] https://timssandpirls.bc.edu/pirls2016/questionnaires/downloads/P16_HQ.pdf
[8] Response options: often, sometimes, never or almost never

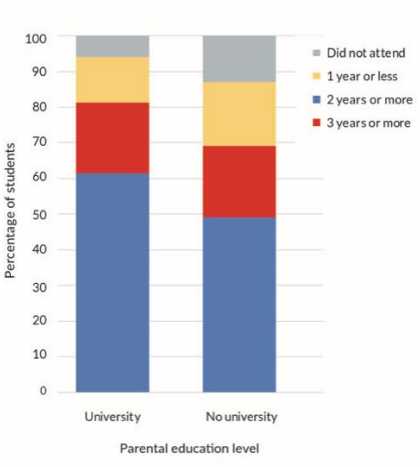


Fig. 12.1 Percentage of children who attended pre-primary education, by parental education level (international average).

This survey has inspired us to examine whether the importance of early learning opportunities for later schooling outcomes hold in different countries and cultures around the world. Our analysis included data from more than 30 education systems that participated in TIMSS 2011, TIMSS 2015 and PIRLS 2011. The results were published in a policy brief of IEA's *Compass: Briefs in Education series*[9].

High-educated parents engage more frequently into activities stimulating early learning

The analysis shows that highly educated parents engaged more often in learning activities with their children during early childhood than parents with lower levels of education. This finding holds across all countries.

Early learning activities are related to higher school performance

Our analysis also shows that students who – according to their parents – were more frequently engaged in early learning activities achieve higher scores in the reading, mathematics, and science tests. The positive association between the frequency of early learning activities and achievement holds irrespectively of the education level of parents.

[9] https://www.iea.nl/publications/series-journals/iea-compass-briefs-education-series/september-2018-preparing-ground

Children of highly-educated parents more frequently attend pre-primary education, and for a longer period of time

Moreover, compared to children with parents with lower levels of education, those with highly-educated parents attended pre-primary education for a longer period of time (Figure 12.1). For example, 82% of children with at least one parent holding a university degree attended institutionalized pre-primary education for 2 years or longer, while only 69% of other children did.

Why this matters and what we can learn from it

How frequently parents engage in early learning activities is related to students' performance in primary school around the world. Further, children who attend pre-primary education also have higher levels of achievement later in school.[10] [11] [12] [13]

It therefore seems that early learning opportunities at home or pre-school can help mitigate the learning gaps between socially disadvantaged and advantaged children. Therefore, we recommend efforts be made to increase participation in early childhood education programs, especially for disadvantaged children. Further, awareness should be raised amongst parents and caregivers regarding the lifelong impact of early learning, and opportunities should be offered to learn how to be more responsive to children's needs.

[10] https://timssandpirls.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf

[11] https://timssandpirls.bc.edu/pirls2011/downloads/P11_IR_FullBook.pdf

[12] http://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/fullpdfs/T15-International-Results-in-Mathematics.pdf

[13] https://www.iea.nl/publications/series-journals/policy-brief/november-2014-participation-arab-countries-early

Chapter 13
The More Experienced the More... Boring?

Pietro Sancassani

Teachers are arguably one of the most important ingredient of education systems. If countries want to improve their students' education and skills, it is therefore natural to analyse what characteristics make teachers more effective, as this can provide valuable information for policymakers.

Does teacher experience matter?

An aspect that has been widely investigated in the teacher-related literature is the impact of teacher experience on student test scores. While there is a long-held belief that teachers will become more effective as their experience grows, this has not always been confirmed in empirical studies. This might come as a surprise, as we expect that the more experienced a teacher is, the better she or he will be able to manage a classroom, master the curriculum and so forth. However, other aspects related to teaching methods and enthusiasm may deteriorate as teachers get more experienced.

It does! But not necessarily in a good way...

In my latest paper,[1] I analyse the impact of science teacher characteristics, including experience, on students' science test scores as well as on the engagement and enjoyment of students in learning science. I use data from TIMSS 2015, an international large-scale assessment, and in this blog post I focus on the impact of teacher ex-

Pietro Sancassani
ifo Institut, Munich, e-mail: sancassani@ifo.de

[1] https://www.ifo.de/publikationen/2021/working-paper/effect-teacher-characteristics-students-science-achievement

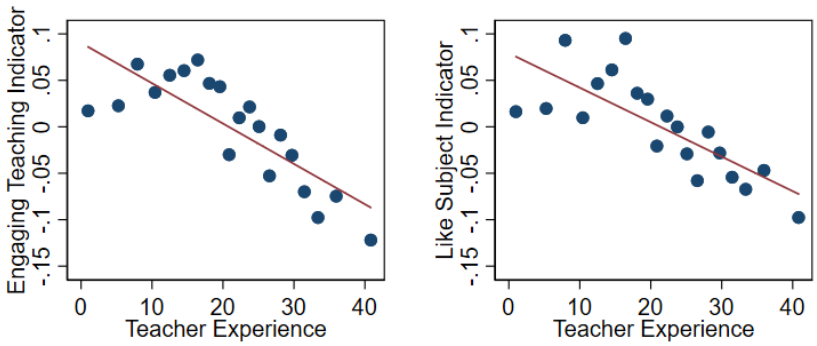


Fig. 13.1 The Impact of Teacher Experience on the *Students Find the Teaching Engaging* and the *Like Learning the Subject Indicators*
Source: own elaboration of TIMSS 2015 data. The figures show the impact of years of teacher experience on standardized students' indicators after controlling for students' and teachers' background characteristics (like student SES, gender etc.) for 39,827 students in 10 countries (Armenia, England, Georgia, Hungary, Kazakhstan, Lithuania, Malta, Russia, Slovenia and Sweden)

perience on whether students find the teaching engaging and whether students like learning a subject using 4 four science subjects: Physics, Chemistry, Biology, and Earth science.

I restrict the analysis to countries where these four science subjects are taught by at least two different teachers, which gives me a sample of 10 countries and roughly 40,000 students. In this setting, the deviation of outcomes in one subject from the average science outcome of each student is associated with the deviation of teacher characteristics in the same subject from the average science teacher characteristics of each student. This procedure has several advantages. First, it controls for important factors, such as student socio-economic status or average science knowledge, that are likely to affect the outcomes of interest. Second, the estimated coefficients are not affected by the fact that students from more affluent background tend to be matched with more prepared teachers. In this scenario, we can credibly attribute the impact on the outcomes of interests to the teacher characteristics.

My results consistently show that experience has a negative and statistically significant impact on both the engaging teaching and enjoying learning indicators. The results are consistent across all subjects and are also confirmed when using standard regression models.

This is clearly shown in Figure 13.1, where I report the impact of teacher experience on the students' indicators net of many students' and teachers' background variables. In both cases, the slope of the line capturing this relationship is clearly negative.

So what?

Keeping students engaged should be the objective of every teacher. Similarly, teachers should strive to make their students enjoy learning the subjects they teach. These aspects, in turn, also contribute to students' success. While the relation between teacher experience and students' test scores has been widely investigated, we know relatively little about the relation between teacher experience and students' engagement or students' enjoyment of a subject. There is therefore need for further research on what aspects can maintain teachers engaging throughout their careers. For example, teachers (and, consequently, students) could benefit from professional development programs aimed at the deployment of digital technologies in class. This has the potential of keeping teaching methods in line with students' interests and preferences, an aspect that could play an important role in maintaining a high level of pedagogical efficacy for teachers.

Chapter 14
The Error Climate and Its Influence on Student Performance

Andrés Christiansen

Is it wrong to do something wrong?

Learning is an active process that requires students to explore and discuss their own opinions and doubts (Tulis, 2013[1]). Thus, during this process, it is natural to make mistakes. Indeed, making mistakes usually generates learning opportunities and gives the teacher clues about students' cognitive process.

However, it is usual for students to feel negative emotions when making a mistake in a task or an exam because they may perceive this situation as embarrassing and threatening for their self-concept. How a student perceives the treatment of their mistakes in the classroom is called the error climate (Steuer, Rosentritt-Brunn, & Dresel, 2013[2]). Hence, the error climate is constructed through the interactions between teachers and students, and it has repercussions for their behavior and attitudes.

Where there is a positive error climate, students are allowed to make mistakes without severe consequences; as a result, they are motivated to correct the error and understand why they made a mistake, using it as an integral element of their learning process. On the contrary, in a classroom where there is punitive treatment of mistakes—where students are criticized or humiliated—it inhibits them from participating in challenging academic situations (Steuer, & Dresel, 2015[3]).

Andrés Christiansen
Centre of Educational Effectiveness and Evaluation, KU Leuven, e-mail: andres.christiansen@kuleuven.be

[1] https://doi.org/10.1016/j.tate.2013.02.003

[2] https://psycnet.apa.org/doi/10.1016/j.cedpsych.2013.03.002

[3] https://www.psychologie-aktuell.com/fileadmin/download/ptam/2-2015_20150624/08_Steuer.pdf

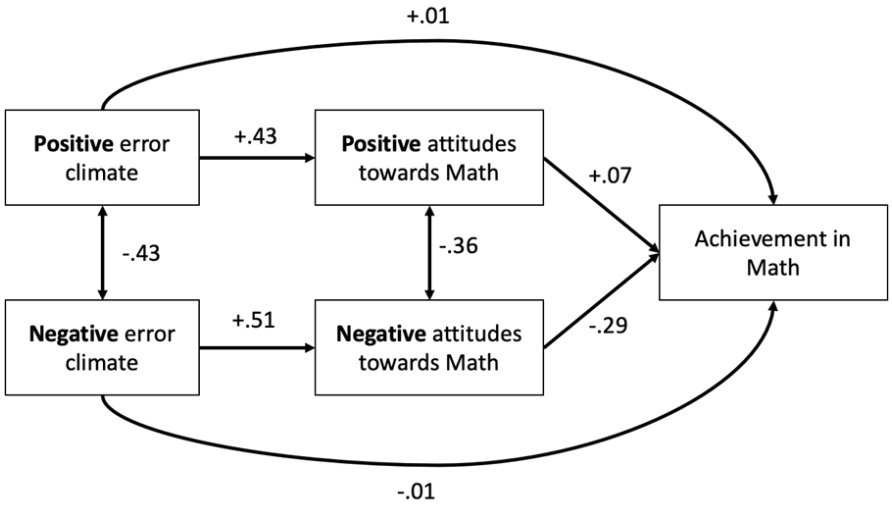


Fig. 14.1 Relationship between error climate, attitudes towards mathematics, and achievement in mathematics (coefficients are standardized).

Allowing mistakes can create positive attitudes toward learning and boost achievement

A recent study published by the Peruvian Ministry of Education (2018)[4] evaluated the relationship between the error climate and attitudes towards mathematics and how this relationship ultimately affected student performance in mathematics.

First, the relationship between positive error climate and positive attitudes was evaluated separately, as well as the relationship between negative error climate and negative attitudes. The study found that a positive error climate was able to explain 18.5% of the variability of positive attitudes towards mathematics. By contrast, a negative error climate explained 30.3% of the variability of negative attitudes towards mathematics.

Then, in a comprehensive model (Figure 14.1), the study evaluated the relationship between these constructs and their impact on performance. This model explained 14.2% of the variability of achievement in mathematics between students.

As Figure 14.1 shows, there is a significant influence of error climate on performance. This is mediated by attitudes toward mathematics, predominantly by negative attitudes toward mathematics. Thus, it is clear that how the teacher handles mistakes within the classroom shapes students' attitudes, which contributes to

[4] http://umc.minedu.gob.pe/la-influencia-del-clima-del-error-sobre-las-actitudes-hacia-la-matematica-en-estudiantes-de-secundaria/

the formation of students' self-concept and ultimately impacts performance (Cueto, 2003[5]; Marsh et al., 2012[6]).

Finally, it is important to note that these effects do not change when other covariates such as socioeconomic status, language, or the school geographical area are considered. Therefore, the relationship between error climate, attitudes, and achievement remains constant regardless of students' social conditions. This finding is thus relevant for any student or school.

[5] https://www.grade.org.pe/publicaciones/531-las-actitudes-de-los-estudiantes-peruanos-hacia-la-lectura-la-escritura-la-matematica-y-las-lenguas-indigenas/

[6] https://doi.org/10.1080/00461520.2012.670488

Part IV
Equity in Education

Chapter 15
An Unfair Start: UNICEF's Report Card on Educational Inequalities in Rich Countries

Ides Nicaise

Two OCCAM contributors (John Jerrim and myself) participated in the Advisory Board of UNICEF's latest Report Card on educational inequalities in rich countries. It was a great experience, not least because of the diversity of perspectives on the topic. There were many underlying conceptual dilemmas, data problems and difficult choices to be made in selecting the indicators and the ensuing ranking of countries. The UNICEF research team finally decided on all these methodological choices, which was fortunate in view of the lack of consensus within the Advisory Board.

Inequality versus unequal opportunities

A key issue in the philosophical and social literature on inequality relates to the level of 'acceptability' or 'desirability' of inequality. Whereas many may want to see inequalities decline, some would argue that inequalities are 'natural' and - to some extent - desirable. The meritocratic view claims that some level of inequality will always exist due to 'innate' differences between individuals; it is therefore futile and inefficient for policymakers to try to eradicate inequality completely. Moreover, according to meritocrats, inequalities in educational attainment that emerge through effort should be nurtured, not countered. The only source of inequality that most parties would agree should be tackled is inequality of opportunity by social background. In other words, each generation of children should have an equal opportunity to 'succeed' (e.g. in education, in the labour market), irrespective of their background - including the wealth of their parents.

UNICEF decided to adopt a neutral position by ranking countries according to their overall inequality in educational outcomes (measured, for example, as the performance gap between the top and bottom 10% of children in reading). One could

Ides Nicaise
Catholic University of Leuven (KU Leuven), e-mail: ides.nicaise@kuleuven.be

obviously argue that there is a strong correlation (across countries) between overall equality and equality of (social) opportunity – and hence, praise countries at the top of the ranking for having 'equitable' educational systems. However, there are many other differences across countries (e.g. income distributions, share of migrants) that may explain some countries strong results; it may not be due to the education system per se.

The 'Unfair Start'[1] report also provides information about the impact of socio-economic background on reading performance. The measure used for this purpose is the determination coefficient (R^2) of a regression of performance on socio-economic background. The latter parameter is also used as the key indicator of inequality in the OECD's PISA research: rather than the overall degree of educational inequality in a country, it reflects the proportion of that inequality that is explained by social origin. Whether a country has 10% or 30% working class children in its school population does not affect the determination coefficient.

'An Unfair Start' shows a rather different picture of inequality in a country depending upon which measure is used. Take Hungary, for example. It is a country with 'average' levels of inequality is terms of the difference between the top and bottom 10% of readers, but is one of the worse performing countries in terms of the social determination coefficient.

Hidden inequalities

The international large-scale assessments in education such as PISA, TIMSS, PIRLS and ICCS have been criticised for using biased samples which exclude truants and school dropouts or where students with special educational needs are under-represented. Without ignoring these issues, I would like to stress a major difference between PISA and other datasets, namely age sampling (as opposed to grade sampling).

Grade-based samples compare the performance of students of the same grade across countries, as their main aim is to assess curricula. Figure 28 in 'An Unfair Start' displays the huge cross-country variation in school arrears due to grade repetition among 15-year olds: from 1.1% in Iceland to 34% in Belgium. Moreover, the risk of grade repetition, just like under-performance, is very unequally distributed across social groups within countries.

It is no surprise, therefore, that country rankings by social inequality in education differ strongly between PISA and other international large-scale assessment studies. PISA correctly reflects the part of inequalities that are shaped through grade repetition, whereas other datasets tend to dissimulate this part of the picture. The position of Belgium and France in PIRLS and PISA provides a nice illustration of the resulting bias: both countries belong to the top 15 in equality according to PIRLS (figure 8 in 'An Unfair Start', and to the bottom 15 in PISA (figure 18 in 'An Unfair

[1] https://www.unicef.org/publications/index_103355.html

Start'; Admittedly, the different rankings may also reflect shifts in social inequality between primary and secondary education.).

All in all, 'An Unfair Start' sketches a very comprehensive picture of educational inequalities from early childhood to tertiary education. It also synthesises a lot of scientific research that explains, qualifies and enriches the picture. It is important for users to take all these qualifications on board in debates about the ranking of countries.

Chapter 16
Equity for All in European Education: What Does PISA Say?

Cor Sluijter & Remco Feskens

Equity and fairness

Equity in education is a hotly debated topic both nationally and internationally. This has led to a view that educational equity is a problem where action needs to be taken. Our forthcoming article in the Journal of Supranational Policies of Education[1] investigates whether this opinion is correct. Using data from the OECD's Programme for International Student Assessment (PISA) we investigate how "fair" education is across the European Union. We measure "fairness" in this context by considering how well countries perform in PISA independent of the background characteristics of students.

How fair are the European education systems?

We used gender, immigration background, home language, age, and the socio-economic status of students together to predict students' PISA 2015 scores in science, mathematics and reading. We assume that the fairer an education system is, the smaller the influence that these background characteristics have upon PISA scores.

Cor Sluijter
Cito, e-mail: cor.sluijter@cito.nl

Remco Feskens
Cito, e-mail: remco.feskens@cito.nl

[1] https://revistas.uam.es/index.php/jospoe

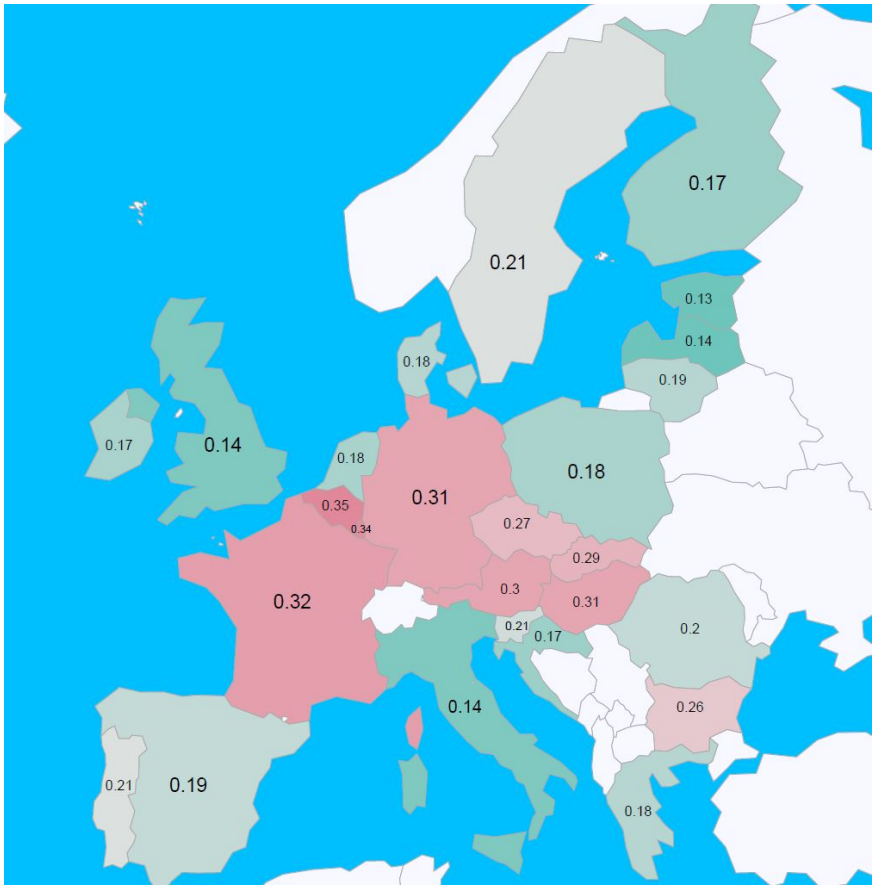


Fig. 16.1 The combined influence (Cohen's f^2) of gender, immigration background, home language, age, and the socio-economic status of students on PISA scores on science.
Note: The numbers on the map shows the effect sizes for each country (expressed in Cohen's f^2, a measure for effect size based on the squared multiple correlation between the background characteristics and the PISA scores).
Cohen's f^2 values of 0.02, 0.15, and 0.35 are considered small, moderate, and large effects, respectively. Countries with effect sizes larger than the EU average are coloured red. Countries in which the effect size is smaller than the EU average are coloured green.

So, what about equity?

The average influence of all background characteristics together on science scores is 0.22 in the EU, which is a moderate effect size. In Belgium, the effect size is the largest (0.35), followed by France (0.32) and Germany (0.31). Countries with relatively small effect sizes include Estonia, Latvia, the United Kingdom, and Italy.

For reading and mathematics the average influence of all background characteristics together on the PISA scores is comparable with respectively a moderate effect size of 0.26 and 0.23.

Based on these results we can conclude that there is still a long way to go before education systems in Europe can be called fair. Moderate effect sizes for science, reading and mathematics shows that there is work to be done. But countries scoring below the EU average would be foolish to rest on their laurels. The effect sizes in these countries are far from insignificant. The outcomes of this study show unequivocally that equal chances to all are still far from a reality.

For detailed information and results, please read our upcoming article (DOI: 10.15366/jospoe) mentioned above.

Chapter 17
Do Girls and Boys Feel Equally Confident in Their Mathematics Abilities? Evidence from TIMSS 2015

Ana María Mejía Rodríguez, Hans Luyten, & Martina Meelissen

Despite the importance of science, technology, engineering, and mathematics (STEM), not enough students are interested in pursuing a STEM career. Moreover, among the few students who do show an interest – and do pursue a STEM career – the proportion of women is small. For example, in OECD countries[1], only 23% of tertiary education graduates belong to the field of science and engineering and only 31% of these students are women. This underrepresentation is also seen in academia, where only 22% of scientific authors are women.

Spearman and Watt (2013)[2] proposed three explanations for the STEM gender gap: (1) differences in ability between boys and girls, (2) differences in motivation and attitudes towards STEM, and (3) differences in socialization. In terms of ability, results from several cycles of TIMSS and PISA show that, in most countries, there are no significant gender differences in mathematics achievement and, if they exist (usually in favour of boys) they have narrowed over the years. Studies have shown that the desire to pursue a STEM career seems more closely related to the second explanation, especially to attitudes such as mathematics self-concept (how confident students feel of their own mathematics abilities). These attitudes are shaped by the environment surrounding girls and boys, which refers to the explanation of socialization. For example, girls and boys might be influenced by gender stereotypes that identify mathematics and mathematics-intensive fields as typically male domains. These gender stereotypes can be conveyed by significant adults – such as parents and teachers – as well as the media[3].

Ana María Mejía Rodríguez
University of Cyprus, e-mail: mejia-rodriguez.ana-maria@ucy.ac.cy

Hans Luyten
University of Twente, e-mail: j.w.luyten@utwente.nl

Martina Meelissen
University of Twente, e-mail: m.r.m.meelissen@utwente.nl

[1] https://doi.org/10.1787/9789264268821-en

[2] https://link.springer.com/chapter/10.1007/978-94-6209-209-9_10

[3] https://seejane.org/wp-content/uploads/portray-her-infographic.pdf

Despite small or non-significant gender differences in mathematics achievement, research has shown considerable male advantages in mathematics self-concept. This research has been mainly conducted in Western countries and with secondary school students. The aim of our study was to explore gender differences among younger students and in a wider set of countries, and to assess the role that socialization has in shaping students' self-concept.

Assessing the gender gap

Using data from TIMSS 2015, we explored gender differences in mathematics self-concept among fourth-grade students in 32 countries[4], taking into account student achievement and socialization variables. For the latter, we used parents' attitudes and characteristics (e.g., attitude towards mathematics and science, the frequency to which they did early numeracy activities with their children) as a representation of socialization at home.

Our findings

Our results indicate that gender gaps in mathematics self-concept already exist among grade 4 students and, therefore, develop before the age of ten. Although the association between gender and self-concept varies across countries, girls have lower self-concept than boys in most of the countries (see Figure 17). A further look shows that girls tend to perceive their mathematics ability more negatively compared to boys not only in countries where girls (on average) performed less well than boys, but also in countries where girls performed equal to or even better than boys.

Why should we care about girls' lower mathematics self-concept?

Previous research shows that confidence in mathematics ability is an important predictor of whether girls choose to study more advanced mathematics courses and, in some cases, these choices can determine access to STEM fields of study in university. Although students in fourth grade are far from making such choices, if girls

[4] Bahrain, Belgium (Flemish), Bulgaria, Chinese Taipei, Croatia, Cyprus, Czech Republic, Denmark, Finland, France, Georgia, Hong Kong SAR, Hungary, Indonesia, Islamic Republic of Iran, Ireland, Italy, Japan, Kazakhstan, Republic of Korea, Lithuania, Morocco, Oman, Poland, Portugal, Russian Federation, Saudi Arabia, Serbia, Singapore, Slovak Republic, Turkey, and United Arab Emirates.

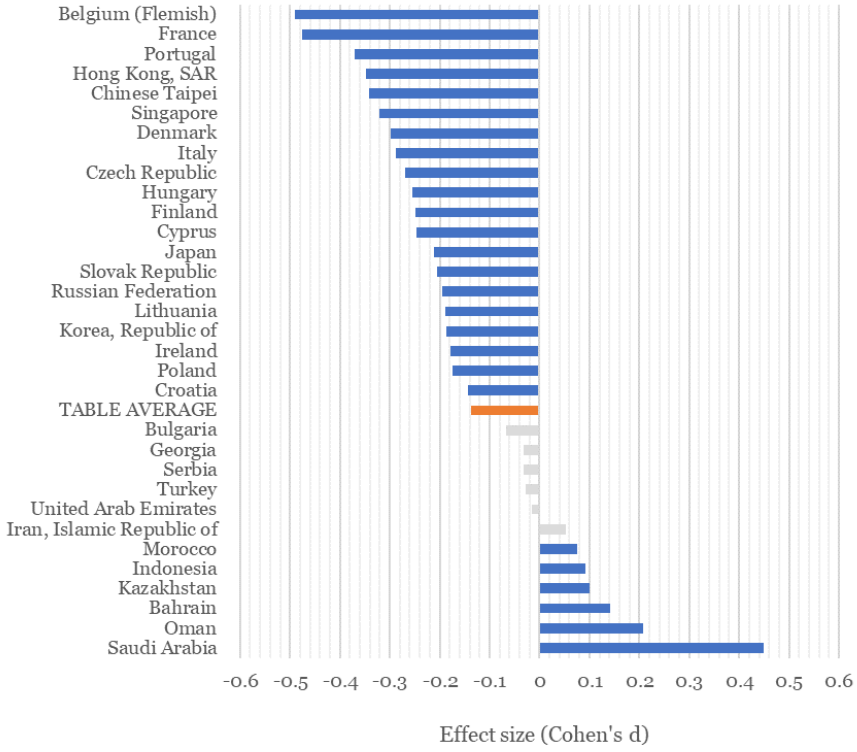


Fig. 17.1 Effect of gender on mathematics self-concept, by country.
Note. Significant effect sizes ($\alpha < .05$; two-tailed) printed in blue; positive effect sizes indicate advantages for girls and negative effect sizes indicate advantages for boys.

already have a low mathematics self-concept by then, it might be too difficult to encourage their STEM participation in the future.

Women participation in STEM is a matter of concern. If women are as able as men to pursue a STEM career but do not do so because of a low mathematics self-concept, we are losing an important source of talented STEM workforce, which could impact productivity and innovations. Furthermore, STEM participation can bring better economic prospects for women, as STEM jobs are among the fastest growing and lucrative careers.

We must work together to ensure girls have the same opportunities as boys when it comes to STEM education, fighting against gender stereotypes and encouraging girls' achievement and confidence in subjects like math.

Chapter 18
Early Tracking Reinforces Social Inequality

Andrés Strello

All educational systems eventually divide their students between different types of educational institutions. However, some countries separate students into different schools at particularly early stages. For example, Germany and Austria sort their students as early as after fourth grade into the academic and vocational school "tracks" . This is what we call (early) between-school tracking. As summarized by Hanushek and Woessmann[1], the arguments behind this system rest on a trade-off between efficiency and equity. This last point, the equity, is where we want to focus on.

Early tracking and social inequality

Tracking, in other words, means that students receive a different education and are taught with different curricula depending on their school track. This sorting leads to a differentiation in terms of results and attainment. The students from lower socioeconomic backgrounds are overrepresented in "lower" tracks (even after controlling for actual achievement), which means that there is a social selection bias to the different tracks. Especially at a very early age, such a selection seems problematic.

In a recent article co-authored by Rolf Strietholt, Isa Steinmann, and Charlotte Siepmann, we used all available cycles of the international large-scale assessments of PISA, TIMSS, and PIRLS to test the effects of early tracking on social inequality, among other achievement indicators. While these tests sample different students in each cycle, each sample is representative of its respective population, enabling us to construct a country-level longitudinal dataset. We used this data to compare social inequality in achievement before and after the tracking takes place to estimate

Andrés Strello
Center for Research on Education and School Development, TU Dortmund University, e-mail: andres.strello@tu-dortmund.de

[1] http://doi.org/10.1111/j.1468-0297.2006.01076.x

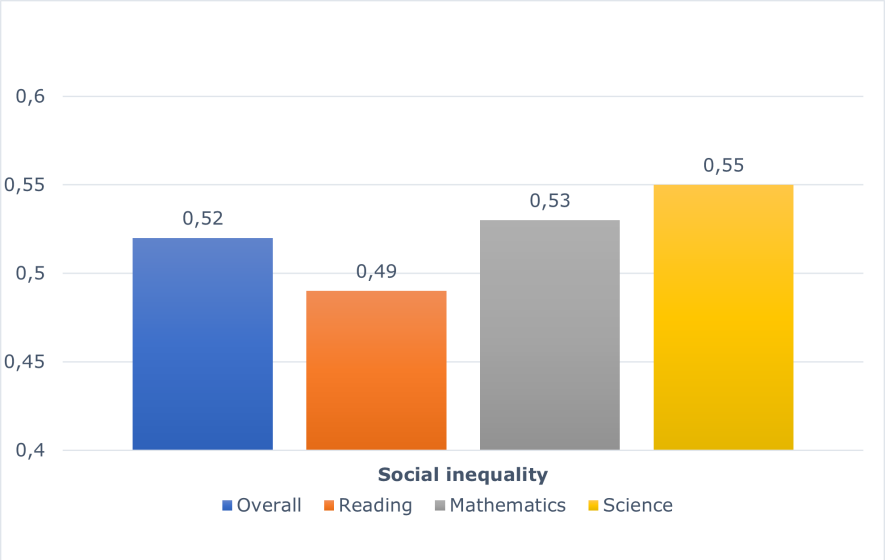


Fig. 18.1 Effect of between-school tracking on social inequality
Note. Effects expressed as Cohen's D. n.s. = not statistically significant

the effect of tracking. We replicated our analyses using multiple combinations of datasets; below, we show the overall combined effect.

What did we find?

The analyses suggest that tracking greatly increases the achievement gap between socioeconomically advantaged and disadvantaged students. Figure 18 shows that the overall effect corresponds to an effect size of around 0.5, which is a very large effect. Interestingly, the effect is surprisingly stable across different outcomes as measured in the achievement tests in reading, math, and science. This stability in effects underscores the far-reaching importance of school structure in creating and perpetuating social inequality.

What do these results mean?

Our results indicate that early tracking widens the difference in outcomes between students from different social backgrounds. Whatever your point of view, this gap widening cannot be overlooked when evaluating tracking policies. It should also be mentioned that, at present, there is no evidence that tracking increases average per-

formance levels, something that our full study also confirmed. Overall, our findings call into question the usefulness of tracking.

Part V
Technology in the Classroom

Chapter 19
Does Home Computer Use Improve or Harm Children's Reading Skills?

Monica Rosén & Jan-Eric Gustafsson

Currently, there is much discussion about the effects of "screen time" and computer use on children's learning and development. Opinions are divided, not only between children and their parents, but also between researchers who report contradictory findings. While some[1] report positive effects on children's development of knowledge and skills, others[2] find computer use to harm achievement, and to increase social inequity.

Correlation does not imply causation

Fuchs and Wößmann[3] found a positive relation between computer access at home and student performance in analyses of PISA data. However, the relationship disappeared when they took student home background into account. When they also took school resources into account, the effect of computer access at home on student performance turned strongly negative. The Fuchs and Wößmann study showed that students who had access to computers at home tended to have a higher social background and attend more resourceful schools than students who did not have access to computers at home. Unless such differences are controlled for, conclusions about effects of computer use at home will be incorrect.

Monica Rosén
University of Gothenburg, e-mail: Monica.Rosen@ped.gu.se

Jan-Eric Gustafsson
University of Gothenburg, e-mail: Jan-Eric.Gustafsson@ped.gu.se

[1] https://journals.sagepub.com/doi/abs/10.1177/1069397105277602

[2] https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecin.12089

[3] https://www.econstor.eu/bitstream/10419/18686/1/cesifo1_wp1321.pdf

Longitudinal analyses provide more robust evidence

In order to ensure that groups with and without access to computers at home are comparable, we could conduct an experiment where we randomly assigned a computer to half the group, and withdrew any existing home computers from all students. Of course, such an experiment can never be conducted in practice. A question is then if existing non-experimental data can be analysed in such a way that they allow equally strong conclusions about causal effects as the hypothetical experiment?

One way to do that is to use a longitudinal approach and investigate change over time. The international studies are repeated at fixed intervals and many countries participate regularly. Thus, at the country level the international studies have a longitudinal design. If changes in different factors, such as home computer use, relate to changes in achievement levels, this suggests a causal relationship, because the countries are their own controls.

In our study[4] we used data from two IEA trend studies of reading in grade four; the Reading Literacy Study 1991 which was repeated in 2001 with nine countries, and the IEA PIRLS studies from 2001and 2006 with sixteen OECD countries.

More screen time decreases reading performance

The results from both studies showed negative effects of home computer use on reading literacy, the negative effect being stronger in the Reading Literacy data than in the PIRLS data. The negative effect also was stronger for reading continuous text, than for finding factual information in non-continuous texts.

In previous research a displacement theory has been proposed to account for negative effects of home computer use. This theory argues that home computer use leads to a diminishing amount of time and interest being spent on reading. The theory also implies that the strongest negative effects would be found on reading of continuous text, which involves skills that are developed during extensive practice of reading. Our results are well in line with the displacement theory. One practical implication of our findings is that measures should be taken to prevent home computer use from reducing the amount of time and effort students spend on reading.

[4] https://link.springer.com/article/10.1186/s40536-016-0020-8

Chapter 20
Are Teenagers in England Addicted to Social Media? (And Does It Matter?)

John Jerrim

There have been lots of concerns raised recently about social media use among young people. This includes links found between time spent on social media and declines in mental health. Similarly, some are now likening constant internet use to an addiction, with teenagers suffering withdrawal symptoms if their smartphone is taken away.

But how frequently do young people in England access social media? Are they any more or less 'addicted' to the internet than young people in other countries? And does this supposed addiction to the internet really do their mental health any harm?

Evidence from the latest PISA data available may well hold some clues.

Social media use and internet addiction as measured by PISA

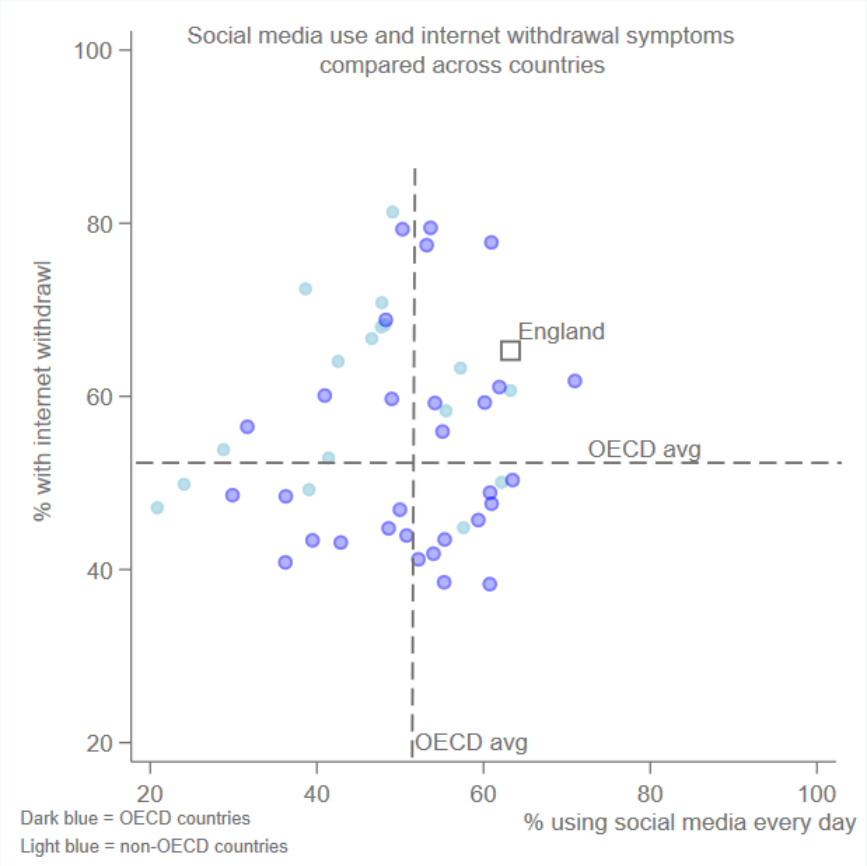
In PISA 2015, 15-year-olds from more than 40 countries including England completed a questionnaire about their use of technology.

This included a question on the frequency with which they used social networks on a five-point scale (from almost never through to every day). They were also asked the following question about whether they suffered withdrawal-like symptoms if they were not able to connect to the internet:

'I really feel bad if no internet connection is possible'. (Strongly disagree to strongly agree).

The chart below compares responses to these two questions from teenagers in England to those of teenagers in other countries.

John Jerrim
University College London, e-mail: j.jerrim@ucl.ac.uk



England is in the top-right quadrant – illustrating that it is above the OECD average on both measures. That is, young people in England are more likely to use social media every day and more likely to express withdrawal symptoms if internet access is taken away than young people in most other countries across the world.

The important question

But does this matter?

The table below illustrates a correlation between daily social media use and internet 'addiction'.

In total, 71% of daily social media users in England experienced internet withdrawal symptoms. This compares to 53% of teenagers who use social media less frequently. A similar difference held after differences in the gender, socio-economic

Social media use and internet withdrawal symptoms of teenagers in England

	Social media use	
	Not every day	Every day
Experiencing symptoms of internet withdrawal	53 %	71 %

status, academic achievement and school attended of social media users were taken into account.

Teenagers who experience the strongest withdrawal symptoms also reported the lowest life satisfaction scores on a scale of 0 to 10, as the next table shows.

Internet withdrawal symptoms and the life satisfaction of teenagers in England

	I really feel bad if no internet connection is possible		
Life satisfaction score	Disagree	Agree	Strongly agree
4 and below	13%	13%	22%
5	7%	9%	10%
6	11%	10%	10%
7	17%	18%	17%
8	22%	23%	17%
9	19%	16%	12%
10	12%	12%	12%
Average score	*7.31*	*7.23*	*6.8*

In particular, the average life satisfaction score for teenagers who strongly agreed that they felt bad when no internet was available was 6.8 out of 10. This is half a point lower than for teenagers who did not report suffering internet withdrawal symptoms, who had an average life satisfaction score of 7.31 out of 10.

In other words, daily social media use among teenagers in England is linked to symptoms of internet withdrawal. This, in turn, is then associated with lower levels of life-satisfaction. Of course, these relationships are correlational rather than causal. They nevertheless highlight how important it is that we develop a better understanding of the impact that social media has upon teenagers' lives.

Chapter 21
Blackboard vs. Computers: It Depends on the Application

Oliver Falck & Ludger Woessmann

Can the use of computers in the classroom take learning to a new level? The emergence of computer-based teaching methods has raised high expectations. Will the next generation of digital natives learn much quicker and better when technology is involved? Most previous studies[1] have arrived at one surprising answer: No. Most scholars find little to no effect of classroom computers on student achievement.

Blackboard vs. Computers?

Does this mean the purchase of expensive computers for the classroom is merely a waste of money and we should stick to blackboard and chalk? Evidence from our own[2] study suggests the answer must again be: No. The devil is in the details here. We find that the lack of effects is the result of using computers in different ways. Some are productive while others are not. The combination of using computers for activities that are more productive than traditional teaching methods and using computers in ways that are less effective than conventional teaching practices leads to overall null effects.

Our empirical analysis uses information from the so-called TIMSS, the Trends in International Mathematics and Science Study. Our sample of the 2011 TIMSS test covers the math and science achievement of over 150,000 students in 30 countries in 8th grade and nearly 250,000 students in 53 countries in 4th grade. In detailed background questionnaires, TIMSS surveys how often teachers in each subject have their students use computers in three distinct activities: look up ideas and infor-

Oliver Falck
ifo Institute, University of Munich, e-mail: falck@ifo.de

Ludger Woessmann
ifo Institute, University of Munich, e-mail: woessmann@ifo.de

[1] https://doi.org/10.1016/B978-0-444-63459-7.00005-1
[2] https://doi.org/10.1111/obes.12192

mation; practice skills and procedures; and (only in 8th grade) process and analyze data.

In line with most of the literature, we do not find a significant effect of computer use on student achievement in math and science on average. However, we find that this null effect is indeed the result of positive and negative effects of specific computer uses: Students' achievement profits from using computers to look up information but suffers from using computers to practice skills (and it is unaffected by using computers to process data).

Blackboard and Computers

The central point in our reasoning is that there are opportunity costs of time, meaning that each classroom minute can only be used in one way or another. Time spent on the computer is time that is unavailable for other modes of teaching. Researching information on the internet does not have a conventional teaching equivalent that is equally effective. In the context of a geography class for instance, rummaging through an encyclopedia is less efficient than using an online search engine to gather detailed and up-to-date information. Thus, using classroom computers to explore information does, in fact, increase student learning.

Practicing skills, however, apparently can be achieved more efficiently applying traditional teaching methods. Solving mathematical equations under the teacher's guidance, for instance, will further mathematical understanding more efficiently than using inflexible drill and practice programs. In this case, the use of computers will decrease student learning.

Policy Implications

Our results also have implications for policy. Recently there has been a big push in many countries to bring computers into classrooms resulting in numerous initiatives that invest large sums of money in classroom technology. According to our findings, the success of these initiatives will depend on the specific uses that the computers are brought to. This not only entails investing in the appropriate hardware but also developing suitable applications.

Chapter 22
Where are the "Digital Natives"? Results from ICILS 2018

Wolfram Schulz & Julian Fraillon

What is ICILS 2018 about?

ICILS is a large-scale, international assessment of grade 8 students' computer and information literacy and computational thinking skills. It addresses a question of critical importance: how well are students prepared for study, work, and life in a digital world? 46,000 students and 26,000 teachers from more than 2,200 schools in 12 countries and two sub-regions took part in ICILS 2018.

ICILS assesses the core knowledge, skills and understanding students need to succeed in our dynamic information environment using authentic tasks in a computer-based environment. ICILS provides countries with reliable, comparable data about young people's development of 21st century computer and information literacy (CIL) skills. In addition, ICILS is unique in offering participating countries the option of directly assessing computational thinking (CT) skills of students. ICILS further collects data from schools, teachers and students about students' use of and opportunities to use ICT.

What does ICILS 2018 say about students' CIL?

ICILS data and test content were used to describe four levels of student CIL proficiciency. The first proficiency level reflects only very basic familiarity and functional working knowledge, while subsequent levels (Level 2 and above) indicate more complete knowledge and skills ranging from basic and explicit information-

Wolfram Schulz
Australian Council for Educational Research, e-mail: wolfram.schulz@acer.org

Julian Fraillon
Australian Council for Educational Research, e-mail: julian.fraillon@acer.org

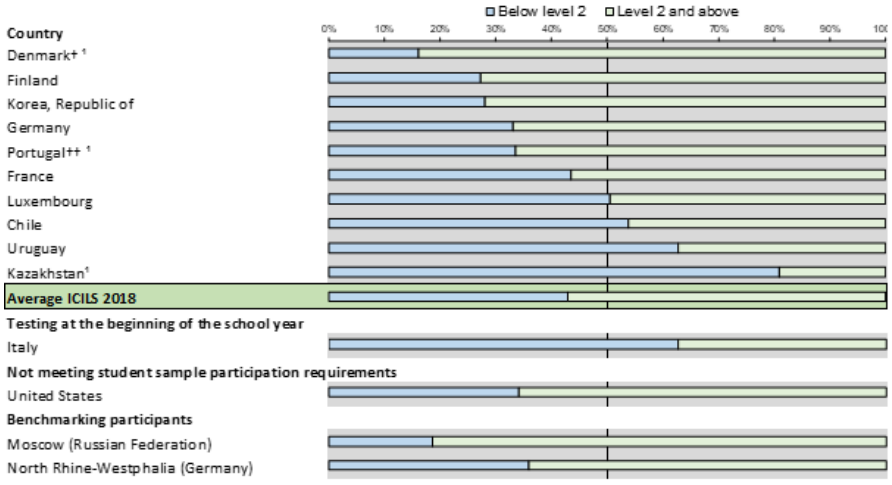


Fig. 22.1 Percentages of students with CIL below Level 2 and at Level 2 and above.
Note. [1] National defined population covers 90% to 95% of the national target population; [†] Met guidelines for sampling participation rates only after replacement schools were included; [††] Nearly met guidelines for sampling participation rates after replacement schools were included.

gathering and management tasks to high-level capacity such as creating sophisticated information products.

Figure 22.1 shows the percentages of grade 8 students in participating countries who demonstrated only basic familiarity with computer and information technology or even less knowledge and understanding (Below Level 2) and those who were able to show more complete skills in this learning area (Level 2 and above). Across participating countries more than 40% had only basic familiarity with computer and information technology and there were large differences between countries with national percentages of grade 8 students with more complete CIL skills ranging from over 80% to only about 20%.

While these figures show large differences between countries, it is also important to highlight the large gaps within countries: The difference between the highest and lowest average CIL scores across countries was 157 scale points. Within countries, the gap separating the average CIL scores of the top and bottom 5% of students ranged from 216 (Denmark) to 347 scale points (Kazakhstan).

The results of this second cycle of the study call into question the generalization that young people are "digital natives" who through exposure to the use of digital devices develop expertise in their use. Data collected by this study demonstrate that providing students and their teachers with information and communications technology (ICT) equipment alone does not automatically result in the development of sophisticated digital literacy skills. Students need to be taught how to use computers effectively, and their teachers need to be supported in their use of ICT in teaching.

Chapter 23
Corona and Education: Are Students Really Ready for Digital Schooling?

Evi E. Konstantinidou

The Rise of Distance Learning

During recent months, mitigating COVID-19 has meant the severe curtailment of normal life, including global closures of schools. This approach to control the spread of the virus has produced an unprecedented wave of disruption in the education of more than 80 per cent of students worldwide, as reported by UNICEF. In light of this, educators, students, learning material providers, policymakers and parents have been forced to engage in distance learning. In order to facilitate distance learning, a vast majority of countries use the power of technology. Presently, there is a global movement towards an 'open access' approach to learning content, as many publishing houses and other material providers open up their platforms and other online resources. One of the critical imperatives under discussion is whether countries and their schools are capable of supporting online learning. Preparedness of schools, including access to educational platforms and the provision of online school resources and learning support are vital. However, while this supply-side debate continues, the elephant in the room is the demand-side question. Namely, are students prepared for online learning? Can they accept, engage and flourish with technology-integrated education for an unspecified period of time?

Are Students Prepared for Online Learning?

To understand possible pitfalls in the move towards online learning it seems useful to refer to Davis' (1986) Technology Acceptance Model (TAM). Davis suggests

Evi E. Konstantinidou
Department of Educational Administration & Evaluation, University of Cyprus, e-mail: konstantinidou.evi@ucy.ac.cy

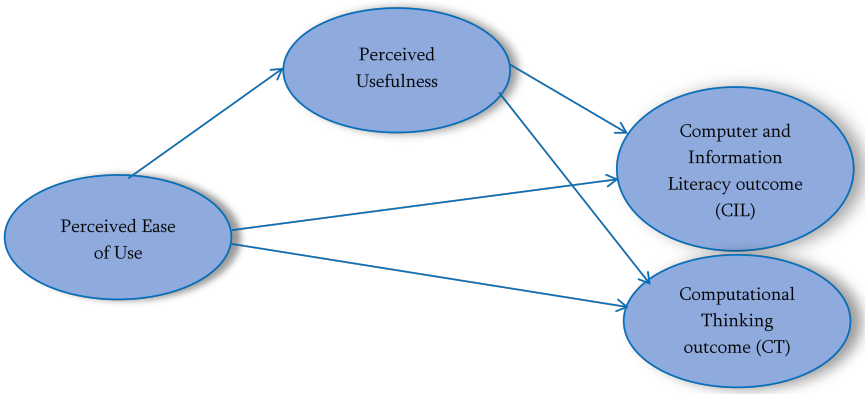


Fig. 23.1 Conceptual model of this study considering the two core aspects from TAM model.

that both the ease of use and usefulness of a technology affects users' intention to engage.

Perceived usefulness means that the students believe that the technology will improve their performance, while perceived ease of use refers to the belief that using the technology will be free of effort. In a recent study, we used data from the International Computer and Information Literacy Study (ICILS)[1] to investigate these issues. The study asked 8th grade students how well they feel prepared to use computers for various tasks, how frequently they use computers, and administered a test on how well students can actually handle computers for learning.

So, What about Students' Technology Acceptance?

For all 12 countries[2] it is found that the effect of students' self-efficacy is positive and significant for both student outcomes while it also positively affects how much students use Information and Communications Technology (ICT). For example, in Finland it is found that if students' self-efficacy increases by one standard deviation, student performance in Computer and Information Literacy (CIL) increased by 0.33 standard deviations, implying a moderate effect.

The effect of students' use of ICT on Computer and Information Literacy (CIL) is found to be significant in 9 out of 12 countries, whereas its association with Computational Thinking (CT) was significant in 4 out of 8 countries. Similarly, the indirect

[1] https://www.iea.nl/studies/iea/icils

[2] Sample of the study: Chile, Denmark, Finland, France, Germany, Italy, Kazakhstan, Republic of Korea, Luxemburg, Portugal, United States, Uruguay

effect of students' self-efficacy via students' use of ICT was more significant for CIL and specifically in 9 out of 12 countries.

These findings imply that the provision of resources by educational systems in order to support digital learning is crucial – but not enough. It is also important to consider whether students find technology useful for their learning and if they feel comfortable using it.