

PISA's Inconsistencies

Why policy-makers should be cautious with PISA results

Background

The Organising Bureau of European School Student Unions (OBESSU) is the platform for cooperation between the national school student unions active in general secondary and secondary vocational education in Europe. OBESSU has been advocating for high quality education for all for more than 40 years.

Advocating for an increased focus on qualitative assessment methods has been central to OBESSU's work. Concerning the assessment of national education systems, we have long criticised the reductionist methodology and use of the OECD Programme for International Student Assessment (PISA). With this literature review we want to highlight the main arguments against PISA, repeatedly pointed out by academics and experts all around the world.

PISA claims to provide data on student learning outcomes and allow policy-makers to compare the quality and equity of learning outcomes reached around the world. We call for caution as PISA results rely on a narrow quantitative assessment of few indicators which can by no means provide a holistic picture of quality education. Such a distorted understanding of national education systems shall not be used to implement policy changes in education, as it has been done in the past. It is important that national educational systems state clearly what they value in education. What matters in education must be debated in the public sphere through democratic engagement and not be imposed by the OECD.

Furthermore, the realities of educational systems will never be fully understood without an active dialogue with school students, teachers and their representatives. A greater inclusion of students and teachers in policy-making in education will guarantee more sustainable reforms.

In this context, we demand

- A focus on qualitative assessment. Instead of test-oriented narrow curricula, PISA should measure school student's wellbeing, the impact of socio-economic background and equity in education. These should be considered the main outcomes of the PISA-report and the main grounds on which to focus any education system related changes and developments.
- PISA results to be seen as data to support further research on low-performing schools and areas.

Abstract

Issued by the [Organisation for Economic Co-operation and Development \(OECD\)](#), the Programme for International Student Assessment (PISA) tests the skills and knowledge of 15-year-old students in mathematics, reading, and science every three years. It assesses students' ability to apply these knowledge and skills to real-life situations. It also includes measures of cross-curricular competencies such as collaborative problem solving.

PISA has been controversial since the publication of its first results in 2000 because it is driven by two main concepts, free-market thinking and globalisation. Its results are taken very seriously by policy-makers and lead to neoliberal education policies. Moreover, with its three-year assessment cycle, PISA has caused a shift of attention to short-term policies designed to help a country quickly climb the rankings, despite research shows that sustainable changes in education take decades.

Today PISA is the world's largest empirical study of schools and education. PISA's ambition is to measure life skills needed to function in society, but this claim implies that these skills are the same across countries which is far from the reality and denies the existence of other types of well-functioning societies. Besides, by trying to assess a few measurable aspects of education, PISA ignores other educational fields such as social sciences, ethics, foreign languages, arts, history or geography and therefore it narrows our perspective of what education should be about. Finally, the high importance given to PISA's rankings leads to an increase of the already high level of stress in classrooms and schools endangering the wellbeing of students and teachers.

1. Introduction

The Organisation for European Economic Cooperation (OEEC) was established in 1948 to run the US-financed Marshall Plan for reconstruction of the European continent destroyed by war. The OEEC declined after 1952 due to the unexpected end of the Marshall Plan and a shift in favour of NATO. In September 1961 the OEEC was replaced by the Organisation for Economic Co-operation and Development (OECD) with the purpose of promoting policies that encouraged the growth of member economies and the development of trade and the world economy. The OECD consisted of the European founder countries of the OEEC plus the United States and Canada. The list of member countries has expanded over the years, with 36 countries today.

Since its establishment in 1961, the OECD has been described as a think tank, a geographic entity, a policy-making forum, a network of policy-makers, researchers and consultants, and a sphere of influence (Rizvi and Lingard, 2010). Article 1 of the OECD's convention requires its members to promote policies designed to develop the world economy and trade on a multilateral and non-discriminatory basis. Nowadays, the OECD views itself as *"an international organisation that works to build better policies for better lives"*. Its goal is *"to shape policies that foster prosperity, equality, opportunity and well-being for all"*. (OECD)

Since its creation, the OECD has emphasised the role education must play in economic and social developments. However, there has been, within the organisation, much debate about the ways in which economic and educational policies could be related. In the 1970s and 1980s, the European countries started to push forward their social-democratic agendas, going against the dominant US versions of market liberalism and refusing to view education and social policies as secondary to, or instruments of economic policies.

The place of education work within the OECD has changed significantly since its establishment. Initially, in the OECD's original charter there was no independent space allocated to education. In 1975 it was incorporated into the Directorate for Social Affairs, Manpower and Education, which in 1991 was reconstituted as the Directorate for Education, Employment, Labour and Social Affairs (DEELSA). In 2002, Education became an autonomous directorate and, with the increasing influence of PISA, began to assume an enhanced role within the OECD and beyond.

In the 1980s, member countries, in particular the US, were demanding regular and reliable data to compare the performances of their education systems and the OECD's education work evolved in response. (Sellar & Lingard, 2014) The OECD started creating the Programme for International Student Assessment, now well known as PISA in the mid 1990s. PISA was constructed and intended for the 30+ industrialised and wealthy OECD countries, but a similar number of countries have joined at a later stage.

PISA tests the skills and knowledge of 15-year-old students in mathematics, reading, and science every three years. Each domain is given major emphasis every nine years, being allocated two-thirds or more of the available testing space (OECD, 2009). The average age of 15 was chosen because in most OECD countries, young people at this age are nearing the end of compulsory education. PISA assesses students'

ability to apply knowledge and skills to real-life situations. It also includes measures of cross-curricular competencies such as collaborative problem solving. The OECD (2012, p. 10) argues that *"skills have become the global currency of twenty-first century economies"* and develops a narrative that presents education and training as areas of policy intervention to improve the economic strength of nations.

The first PISA survey was conducted in 2000 in 32 countries and economies (including 29 OECD member countries) and assessed reading, mathematics and science, with a primary focus on reading. Another 11 countries and economies completed the same assessment in 2001 and 2002. The last PISA survey, conducted in 2018 in 80 countries, assessed reading, mathematics and science with a focus on reading. Each survey round also assesses additional cross-curriculum competencies such as problem solving (PISA 2003), electronic literacy (PISA 2009) or financial literacy (PISA 2018). However, only a small group of countries decided to participate in these additional assessments.

We believe that PISA is driven by two main concepts, free-market thinking and globalisation. Nowadays, the tendency is to apply concepts and ideas from the market economy to areas of the public sector, such as education, in order to make them more efficient. This phenomenon is known as New Public Management. At the same time, the economy is becoming globalised and the workforce has to be flexible and moveable. In such high globalised and competitive market, there is a need for common standards in education, common systems for exams, degrees, and qualifications. PISA test scores and rankings are ideal for this purpose, whether or not that usage was intended by the PISA researchers. (Sjøberg, 2016)

Today PISA is the world's largest empirical study of schools and education. Its results are taken very seriously by policy-makers and lead to educational reforms. It has become common sense that high scores on PISA reading, mathematics, and science are predictors for the country's future economic competitiveness. Given its size and importance, PISA has to be understood not just as a study of student learning, but also as a social phenomenon in a wider political, social, and cultural context (Lundgren, 2011). PISA success has been made possible, at least in part, by demands from member countries: countries opt into PISA and pay to participate. According to Sellar & Lingard (2014) *"the influence of PISA as a tool for comparison increases as the number of participating countries and economies expands, and the participation of new economic powers such as China and India has made possible new comparisons that have drawn the attention of politicians in countries, where PISA has not had significant effects on policy debate previously"*.

Recent figures estimated that the annual cost, after adding the expense of involving half a million students and tens of thousands of schools and their teachers, is approximately 80 million Euros (Maslen, 2013). It is exclusively financed through direct contributions from the participating countries, usually from the education ministries.

The success of PISA allows the OECD to expand its global testing approaches to the domain of adult skills through the PIAAC, to school-level performance through the new PISA-based Tests for Schools programme, and potentially to higher education through the Assessment of Higher Education Learning Outcomes. The OECD is seeking to expand the scope of the assessment to measure a broader set of skills and competencies, increase the scale of the assessment to cover more countries, economies and schools; and enhance the explanatory power of the assessment for policy-makers. (Sellar & Lingard, 2014)

2. PISA's Ideological Implications and Methodological limitations

PISA's controversy lies on two main elements: ideology and methodology. The first concerns the PISA's conception about education and the second concerns the statistical models and techniques applied in the PISA survey.

2.1. What does PISA claim to measure?

The overall aims of PISA had already been stated in 1999, before the first PISA testing took place in 2000: *"How well are young adults prepared to meet the challenges of the future? Are they able to analyse, reason and communicate their ideas effectively? Do they have the capacity to continue learning throughout life? Parents, students, the public and those who run education systems need to know"*. (OECD, 1999, p. 7)

Such an ambition, of measuring so-called "life skills" and how well the young generation is prepared to meet the challenges of tomorrow's world, implies that these challenges are already known and are similar for young people from different countries and cultures. According to Sjøberg (2016, p.111), *"it assumes one universal conception of success and a universal set of necessary skills" which is "detrimental to the expression of educational values within countries and to cultural identity and diversity in general."* Although life in many countries does have some similarities, one can hardly assume that 15-year-olds in Italy, Japan, Mexico and Turkey are getting prepared for the same challenges or that they need the same skills and competencies to succeed in countries with different cultures and levels of economic development. As noted by Araujo, Saltelli and Schnepf (2017), country comparison in PISA is achieved by *"ignoring the great diversity of curricula across the participating countries – diversity which might in fact be a source of country-specific creativity and well-being"*.

Since the emergence of PISA, comparative education analysis has changed. The analysis has shifted from studying what students have been taught and how much they have learned, to studying what students can do with what they have been taught. (Hutchison & Schagen, 2007)

We claim that PISA tests do not measure quality according to national school curricula, but based on the perceived universal, culture-free, curricula made by the OECD-appointed PISA experts. This explains why the PISA subjects are reading, mathematics, and science. According to OECD officials, these subjects are key elements for competitiveness in a world economy driven by science and technological development. This also implies that domains such as social sciences, ethics, foreign languages, history, geography, or physical education, to name just a few, are seen as not relevant or not important for the development of young people. Besides, PISA does not address attributes that are central in many countries' educational systems, such as equity, empathy, solidarity, curiosity and engagement or care for the environment. (Sjøberg, 2016)

Of course, one test, even a test like PISA, cannot assess all possible school subjects; but by selecting some subjects and ignoring others, PISA implicitly sends a message to the public as well as politicians about what is important for schools and future life. The actual choice of reading, science, and mathematics also reflects the OECD concern for economic competitiveness in a global, high-tech market economy: *"In a*

global economy, the yardstick for success is no longer improvement by national standards alone, but how education systems perform internationally. The OECD has taken up that challenge by developing PISA, the Programme for International Student Assessment, which evaluates the quality, equity and efficiency of school systems in some 70 countries that, together, make up nine-tenths of the world economy". (OECD, 2010, p. 3) As mentioned, when PISA extended its repertoire in 2012, the added domain was "financial literacy." (OECD, 2013) This addition was undoubtedly a consequence of the free-market economic mandate and priorities of the OECD and was included in the testing by some 40 countries.

It is also interesting to note that the importance of PISA is defined with terms such as the "*nine-tenths of the world economy*", rather than in terms of the fraction of the world's population, which indicated PISA's focus on economic competitiveness, not comprehensive education. (Sjøberg, 2016)

Although PISA states that it does not test knowledge acquired in schools and that it does not test according to national curricula, PISA results are interpreted by OECD officials and policy-makers around the globe as valid measures of the quality of national school systems, and the PISA reports are full of policy recommendations regarding schools. (Loveless, 2009)

2.2. How does PISA measure "life skills"?

Using a contradictory format

PISA survey consists in a two-hour test combining a diversity of test items that range from multiple-choice items to extended-response questions. Students also complete a 30 minute student background questionnaires that gathers information about the student's socioeconomic backgrounds and learning habits and attitudes. A background questionnaire is also filled out by the school to provide information about the teaching and learning environment. Both questionnaires are used to analyse the effects of these student-level and school-level variables on learning outcomes. Surveys are coordinated by participating governments, who meet as the PISA Governing Board. The assessment materials are developed by experts, and the fieldwork is designed and managed by an international study centre. (Baird & al., 2011)

As mentioned, PISA's ambition is to measure how well young people are prepared to meet the challenges of a future world economy driven by science and technological development. This goal is contradicted by the format of the testing, which until 2012 was exclusively a traditional pen-and-paper test, where students sit for 2 hours to answer written questions, alone, and without access to sources of information. Sjøberg (2016, p.118) addresses a few questions: "*how "real life" is that test situation? How does it relate to the challenges that young people may face in their future lives as citizens, as participants in tomorrow's democracy, and as members of a skilled workforce? The fact is that initial PISA test does not resemble any real life, problem-based situations*".

Students in the 32 countries that participated in the PISA 2012 pencil-and-paper assessment were invited to take a test of reading and mathematics on computers. This was supposed to expand the range of situations in which student's ability to apply their knowledge can be measured. The computer-based test was composed of 18 reading questions and 41 specially designed mathematics questions.

With doubtful neutral translations

The original language of the assessment materials is predominantly English, and the materials are translated into more than 40 languages. (Baird & al., 2011)

Arffman (2010) study revealed that meaning and context change in a translation, something is always lost, or meaning is modified in translation. Bad translations also may cause readers to lose interest and motivation to become engaged with the text, and this may have a negative impact on the test results. According to Arffman, this effect is likely to be greater in countries where students are critical, independent, and unwilling to obey the authority of schools and teachers.

In order to diminish the risks of losing meaning and the readers' engagement, the OECD has to address the challenge of creating items that are culturally neutral. The OECD generally runs trials before implementing the final PISA questionnaire, although not in all countries, since some countries choose not to cover the expenses. The purpose is to identify 'differential functioning', that is items for which student performance is unusually high or low in one country or another, or for one gender or another, compared with model expectations. Such items are removed from the set for survey use or revised, with potential consequences for the validity of the literacy domain that is ultimately assessed. (Baird & al. 2011)

The lack of transparency in this approach is an issue that has been often raised. We do not know which items have been removed or revised from the initial questionnaire and why, nor we know in which countries the OECD ran trials, on how many students participated in these trials and who were they. All these questions are legitimate and raise concerns regarding the reliability of the mechanism used by PISA and its consequences on the end results.

Not involving (enough) teachers and parents

A parental questionnaire has been optional since the 2006 test. This questionnaire collects important information on parental background, educational background as well as parental involvement in children's learning. Unfortunately, only a small number of countries chose to collect data from parents: in 2009 only 14 countries and in 2012 only 11 countries out of 65 countries covered in each PISA round.

Similarly, since the 2015 PISA test, organisers made a teacher questionnaire optional. The teachers' and parents' questionnaire open the possibility to investigate a greater variety of factors associated with students' achievement. Some of the most important skills, competences and attitudes that matter in quality education, such as critical thinking, motivation, resilience, self-management, resourcefulness, and relationship skills, cannot be easily measured by statistical models and computer algorithms, but they can be assessed by teachers using human judgment. (Araujo, Saltelli and Schnepf, 2017)

The data collected is used by the OECD to contextualise the findings and to explore the potential effects of background variables on student achievement. (Baird & al., 2011) According to Sellar & Lingard (2014) linking performance data to other data sets is a mean to increase the explanatory power of PISA.

With unrepresentative samples

PISA surveys are sample-based. The sample size per country was originally intended to be around 4,500 15-year-olds. Some countries have experienced difficulties meeting PISA's sampling criteria, because of school recruiting problems. Others have not encountered such challenges, and have chosen to increase their samples from the very first survey so that they could profit from the opportunity to gather comparative data for their different regions. (Baird & al., 2011)

School students "are selected using a conventional two-stage stratified sampling procedure. The school population is first stratified, usually by school type and size, but other variables can be brought in, depending on the country. Schools are selected at random from within the various strata to produce a school sample representative of the population. Within each school around 30 students are randomly selected for testing from a list provided by the school; where schools are too small to provide 30 students all the 15-year-olds are tested". (Baird & al., 2011)

A problematic issue is whether the sample of students used to estimate the country average is representative of the target population. PISA allows states to exclude students with special educational needs and newly arrived immigrants. This becomes problematic because some countries exclude more students than the 5% threshold imposed by the PISA designers. (Araujo, Saltelli and Schnepf, 2017) Similarly, PISA only assesses students who are currently undertaking schooling. This can cause comparability difficulties when developing countries are concerned as stated by Baird & al. (2011): *"In some countries the PISA target population can be much smaller than the actual population of 15-year-olds, and since students not in schooling tend to be from low socioeconomic groups, the "population attainment" results provided by PISA will not be directly comparable with those of countries with almost 100% coverage"*.

Once selected, each student does not sit exactly the same PISA test. In each survey the number of total items would consume several hours of testing time per student, a situation that neither the students nor their schools would tolerate. To maximize the use of the two allocated hours for the test, the items are bundled into several 30 minute 'clusters' within each domain, and individual students are randomly allocated four of these clusters in the form of two pencil and paper test booklets. Cluster allocation to students uses 'matrix sampling', and allows cluster overlap from one cluster grouping to another to facilitate the placing of all the items within a particular domain onto the same 'Rasch scale'. (OECD, 2011)

Furthermore, there is general agreement in academia that the main limitation of PISA is its reliance on cross-sectional data: data that refers in each round to different student groups. Researchers have repeatedly argue that PISA results should not be used for outlining specific policy recommendations for improving education because the rounds do not follow the same students over the course of their school careers. (Araujo, Saltelli and Schnepf, 2017)

Using Opaque Statistical Model

PISA uses the Rasch model to deliver cross-countries comparison. *"The Rasch model is an Item Response Theory (IRT) model that depends on valid application based on the very strong assumption of 'item invariance': individual students are assumed to have inherent fixed 'abilities', and items have inherent fixed 'difficulties'. In principle, item difficulty cannot vary across students or student groups, irrespective of interests, culture or curriculum experience. Using relatively opaque techniques, the item analysis eventually locates both students and items on a common scale – a scale, by the way, with no absolute 'zero'. In PISA the scale is transformed and 'stretched', to have a mean of 500 and a standard deviation of 100. The effect of this is to exaggerate even the smallest differences in country performance estimates, differences that even if 'statistically' significant are not always educationally significant, but which have at times been over-interpreted by policy makers and politicians, leading to unnecessary concern about their countries' relative 'standards'."* (Baird & al., 2011)

The method has been criticised by well-qualified statisticians, even among those who actually worked on PISA data. According to many authors, the margin of error on the scores provided by the OECD is underestimated, and the rankings of countries are more open to interpretation than one would understand from OECD analyses. (Araujo, Saltelli and Schnepf, 2017) For example, Kreiner & Christensen (2014) argued that by simply changing some of the parameters in the complicated statistical analysis, Denmark's rank can fluctuate from as high as number 2 to as low as number 42.

Another main criticism in the literature concerns the multidimensionality of the items being measured. The items used do not only measure, for example, 'maths' ability but also a student's ability to comply with the test structure. In PISA, it is very unclear which dimensions are being measured and which removed. (Araujo, Saltelli and Schnepf, 2017)

The method used is so technical that it is often unintelligible for the vast majority of people affected by PISA's conclusions and recommendations. (Sjøberg, 2016) Moreover, PISA methods of statistical calculations are published only in a general form, making detailed critique and replication difficult. (Kreiner & Christensen 2014) In this context, with undisclosed statistical procedures, how can PISA creators possibly be held accountable? And how can policy-makers trust the data?

Given PISA's considerable impact on policy, the OECD should make an effort to be more transparent with the choices made to produce data. This should include information on the representativeness of data and how the modelling choices impact on the results. (Araujo, Saltelli and Schnepf, 2017)

3. How PISA results influence policy-making?

With over 60 countries regularly participating in the survey, the release of the PISA performance tables attracts significant attention from both the media and policy makers, and large-scale educational reforms have even been launched in direct response.

The media, policy makers and the public accept PISA as a reliable and legitimate assessment of educational systems. We believe that PISA has the power to shape the way in which they judge systems' performances, define policy problems, set improvement targets and make policy reforms. In other words, Simon Breakspear stated that "*PISA indicators become the lens through which we come to understand our systems*".

With a dominant narrative and presence in the media

Every three years, PISA performance tables are widely published in international and national media across OECD and partner countries. PISA indicators and recommendations receive a high level of reliability. Though the attention given to PISA results in national media varies between countries, in most countries it has increased after several rounds of PISA testing. (Breakspear, 2012)

The fact that the OECD admitted that "*large variation in single country ranking positions is likely*" and therefore recognises the limits of its method, has made no difference to how tables are reported in the media. For example, in 2009 the OECD said that the UK's PISA ranking out of a total of 74 countries was between 19th and 27th for reading, between 23rd and 31st for maths, and between 14th and 19th for science. However, these variations were not stressed in the media.

The media often presents tabloid-like and oversimplified rankings that often lead to a misunderstanding of the results, and even worse, to an unjustified and detrimental public image of the quality of education in schools. This media attention exacerbates the pressure for educational reforms and contributes to the idealisation of particular countries and their educational systems. (Waldow, Takayama & Sung, 2014)

Defining policy problems and new targets

According to Breakspear (2014), the most important effect of the PISA survey is that its results define policy problems. PISA indicators simplify and quantify the policy problems that need to be solved in the area of education. However, educational outcomes are complex and difficult to track over time and by simplifying them, we may lose many elements and limit the national view of what matters in education.

Every three years, PISA reports state if a country is improving, declining or staying stable regarding its own performance but also in comparison with other countries. By doing this, PISA is shaping and framing how policy makers understand the performance of their national education system and the areas that require more attention.

Improving educational quality has become synonymous of improving PISA scores. Breakspear identifies in his study, three targets being set by participating countries:

- The country rank in the PISA tables, which are dependent on other countries performances.
- National PISA scores
- Equity goals that may include a decrease in the percentage of students that score at or below PISA baseline proficiency level or a decrease in the differences between schools.

The new standard is that for a reform to be acknowledged as successful in improving the quality of education, it must be evidenced, at least partly, by an improvement in PISA rankings.

Stanley (2013) questions the phenomenon of setting rank-based targets: *"Rank position is inappropriate as a goal for improvement, as national rankings can be influenced by quite small differences in student cohort scores. Such differences may not represent meaningful learning and/or skill differences in standards attained, which should be the main focus"*.

Encouraging Policy Borrowing

PISA rankings lead to classification of education systems as high, average and low performing systems. This classification has consequences for policy borrowing behaviour. Average and low performing countries are seen as countries where reform effort must be taken in order to improve performance. These countries become the target audience of policy lessons that should be learned from the high performers or the successful PISA improvers. The OECD has identified consistent high PISA performers, being Finland, Canada, Japan and Korea. These countries are the new references and become targets of policy borrowing. They set the goals to which other countries aspire. By giving attention to top performing countries, PISA is narrowing the scope of sources of learning and inspiration. (Breakspear, 2014)

The policy borrowing behaviour can be dangerous and lead to harmful reforms. Policy makers should engage in policy learning rather than in policy copying, especially if little attention is given to the cultural context and the overall policy coherence. (Breakspear, 2014)

Producing Interpretations and Recommendations

PISA creates country-specific reports for all participating OECD countries.

With PISA data as input for its calculations, OECD gives advices on how to make schools better. The operational definition of a "better school" is a school that is "cost-effective," that could give more PISA points per euro spent on education. OECD definition of a good school thereby ignores the national priorities set for the school system. (Sjøberg, 2016)

From the beginning of the PISA project, the OECD has produced graphs and indicators that have shown small or negligible correlations between a country's PISA scores and its spending on education (OECD, 2001). This, of course, has led to the OECD advice that more spending on education will not improve its quality.

Legitimizing Harmful Reforms

PISA rankings are often used as evidence to argue that educational systems need to be improved overall or in a specific area, therefore serving the debate about a potential policy reform. However, policy reforms undertaken in a crisis context often lead to short term reforms attempting to increase PISA scores instead

of improving the education system in a sustainable manner. This phenomenon is now called “PISA Shock” (Breakspear, 2014)

PISA has caused a shift of attention to short-term policies, that are often not even empirically founded, despite research shows that sustainable changes in educational systems take decades. (Open Letter to Andreas Schleicher, 2014)

4. Conclusion

One should remember that the E in OECD stands for Economic, not Education. As the Organisation for Economic Cooperation and Development, the OECD is naturally biased in favour of the economic role of educational systems. Furthermore, preparing the young generation to access the global labour market is not the only, and not even the main goal of national education, which has to prepare students for active participation in their society. It is important that national educational systems state clearly what they value in education. What matters in education must be debated in the public sphere through democratic engagement and not be imposed by the OECD.

PISA is not a neutral statistical model producing objective numbers and facts. PISA claims to measure “*how well young adults are prepared to meet the challenges of the future*” but by stating its purpose, PISA implies that these challenges are already known and are similar for young people all around the world, denying the existence of different realities, cultures and well-functioning societies.

To measure its ambitious goal, PISA uses traditional pen-and-paper test that does not reflect the challenges that young people may face in the future. The meaning of the assessment materials is often lost in translation which can impact the test’s results.

PISA surveys are sample-based and these samples are not always representative of the target population. Furthermore, PISA relies on cross-sectional data, data that refers in each round to different student groups, which limits PISA relevance and reliability.

PISA uses the Rasch model to deliver cross-countries comparison. This method has been criticised by well-qualified statisticians because the margin of error on the scores provided by the OECD is underestimated, and the rankings of countries are more open to interpretation than one would understand from OECD analyses, making PISA results unreliable. Furthermore, the method used is so technical that it is often unintelligible for the vast majority of people affected by PISA rankings.

With over 60 countries regularly participating in the survey, PISA’s rankings attract significant attention from both the media and policy makers, and large-scale educational reforms have even been launched in direct response. The media, policy makers and the public accept PISA as a reliable and legitimate assessment of educational systems. PISA has the power to shape the way in which they judge systems’ performances, define policy problems, set improvement targets and top performing systems, that should be seen as reference, and make policy reforms. This leads to policy borrowing behaviours that can be dangerous when the cultural context is not taken into consideration and can create harmful reforms. But this also conducts to short-term policies despite research shows that sustainable changes in educational

systems take decades. In order to avoid PISA misleading policies, PISA results should be positioned as a resource for policy-making rather than a key policy goal.

Finally, the high attention given to PISA rankings leads to an increase of the already high level of stress in classrooms and schools endangering the wellbeing of students and teachers. A greater inclusion of students and teachers in policy-making in education will guarantee more sustainable reforms.

References:

Araujo, L., Saltelli, A., Schnepf, S. (2017). Do PISA data justify PISA-based education policy? *International Journal of Comparative Education and Development*, Vol. 19.

Arffman, I. (2010). Equivalence of translations in international reading literacy studies. *Scandinavian Journal of Educational Research*, 54(1), 37–59.

Baird, J-A., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Terra, S. & Daugherty, R. (2011). Policy Effects of PISA. Pearson, UK.

Breakspear, S. (2014). How does PISA shape education policy making? Why how we measure learning determines what counts in education. Issue 240 of Seminar Series. Centre for Strategic Education.

Hutchison, D., & Schagen, I. (2007). Comparisons between PISA and TIMSS – are we the man with two watches? National Foundation for Educational Research, 1-32.

Kreiner, S. & Christensen, K-B. (2014). Analyses of Model Fit and Robustness. A New Look at the PISA Scaling Model Underlying Ranking of Countries According to Reading Literacy. *Psychometrika*. 79 (2).

Loveless, T. (2009). How well are American students learning? [The 2008 Brown Center report on American education] (Vol. II, No. 3). Washington, DC: Brookings Institution.

Lundgren, U. P. (2011). PISA as a political instrument: One history behind the formulating of the PISA programme. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 17–30). Rotterdam, Holland: Sense Publishers.

Maslen, G. (2013, November 29). Andreas Schleicher and the PISA phenomenon. *The Sydney Morning Herald*.

Open Letter to Andreas Schleicher, OECD, Paris. (2014). *Policy Futures in Education*, 12(7), 872–877.

Organisation for Economic Co-operation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris, France: OECD Publications.

Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris, France: OECD Publications.

Organisation for Economic Co-operation and Development. (2009) *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2010). *PISA 2009 results: What makes a school successful? Resources, policies and practices* (Vol. IV). Paris, France: OECD Publications.

Organisation for Economic Co-operation and Development. (2011). *PISA 2009 Technical Report*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD Publishing.

Papadopoulos, G. (1994). *L'OCDE face à l'éducation 1960-1990*, Paris: OCDE (collection historique de l'OCDE).

Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*, London, UK. Routledge.

Sellar, S. & Lingard, B. (2014). The OECD and the expansion of PISA: new global modes of governance in education. *British Educational Research Journal* Vol. 40, No. 6, December 2014, pp. 917–936

Sjøberg, S. (2016). OECD, PISA and Globalization: The influence of the international assessment regime. In Tienken, C. H. & Mullen, C. A. (Eds). *Education Policy Perils. Tackling the Tough Issues*. Routledge.

Stanley, G. (2013). "Foreword" to GN Masters (2013) *REforming Educational Assessment: Imperatives, Principles and Challenges*. Australian Council for Educational Research. Camberwell, Victoria.