



# Two decades of havoc: A synthesis of criticism against PISA

Yong Zhao<sup>1</sup>

Published online: 22 January 2020  
© Springer Nature B.V. 2020

## Abstract

The Programme for International Students Assessment (PISA) has become one of the most influential forces in global education. The growing influence has been accompanied by growing criticism. For nearly two decades since the first round of PISA was conducted in 2000, the global assessment program has been roundly scrutinized and criticized by education researchers all over the world. But the mounting criticism seems to have had little impact on PISA's influence as evidenced by its growing power in global education policy and practice. The lack of impact of criticism does not mean the criticism is not valid or PISA has improved. It simply means that the criticism has been largely ignored. The lack of impact is no reason to give up exposing PISA as a flawed business that has great power to misguide education. The expanding influence of the PISA enterprise makes it more even more important to be critical of this juggernaut today. It is also important to consider more effective and more straightforward approaches to present the criticism. The purpose of this article is to present a summary of criticisms that reveal the most fundamental flaws of PISA in non-technical language in one place. Specifically, the article focuses on criticisms of PISA's three fundamental deficiencies: its underlying view of education, its implementation, and its interpretation and impact on education globally.

**Keywords** PISA · International assessment · Educational policy · International and comparative education · Purpose of education · Future of education · Education reform · Educational change

## Introduction

The Programme for International Student Assessment (PISA) has been in existence for more than two decades (OECD 2010). This triennial international education assessment enterprise has been operated by the Organisation for Economic

---

✉ Yong Zhao  
yongzhao@ku.edu

<sup>1</sup> School of Education, University of Kansas, 419 JRP, 1122 West Campus Rd., Lawrence, KS 66045-3101, USA

Co-operation and Development (OECD), an intergovernmental organization with the mission to stimulate economic growth and international trade. OECD was founded in 1961 and has grown to include 36 members, most of which are among the wealthiest countries in the world that practice market economy.

For the nearly two decades since its first implementation, PISA has been roundly scrutinized and criticized by education researchers all over the world (Hopfenbeck et al. 2018; Lewis 2017; Meyer and Benavot 2013; Sjøberg 2015b; Stewart 2013). The criticisms reveal a wide range of problems with PISA that span from its fundamental conceptualizations, technical implementations, statistical analyses and policy interpretations, misuses in policy and practice, and oversized influence on education (e.g. Bieber and Martens 2011; Feniger and Lefstein 2014; Grek 2009; Hopfenbeck et al. 2018; Hopmann et al. 2007; Kreiner and Christensen 2014; Labaree 2014; Meyer and Benavot 2013; Morrison 2013; Pereyra et al. 2011; Rutkowski and Rutkowski 2016; Sjøberg 2012; Stewart 2013; Zhao 2016b). The litany of criticism makes it clear that PISA is fundamentally flawed.

Yet, this fundamentally flawed enterprise has dominated the agenda of educational discussions in government cabinet meetings, academic conferences and journals, and the media (e.g. Arzarello et al. 2015; Baird et al. 2016; Baroutsis and Lingard 2017; Domínguez et al. 2012; Figazzolo 2009; Hopfenbeck et al. 2018; Meyer et al. 2014; Ninomiya 2016). Its distorted view of education has been promoted as the ideal of the twenty-first Century (Schleicher 2018) and the global gold standard for educational quality (Sjøberg 2015a). Its erroneous findings have been interpreted as ways to build such a system (e.g. Schleicher 2018; Tucker 2011) and created false idols of educational excellence for the world to worship (Sahlberg 2011; Tucker 2011; Zhao 2014, 2016b). As a result, it has been recognized “as a form of international and transnational governance and as a disciplinary technology, which aims to govern education in the twenty-first century” (Pereyra et al. 2011, p. 3) and “a model for the governing of national school development in a global world” (Lundgren 2011, p. 28). As such, it has delivered “shocks” to many education systems (Gruber 2006; Lundgren 2011) and led to drastic, but not necessarily productive, changes to education policies and practices in a number of countries (Arzarello et al. 2015; Grek 2009; Gruber 2006; Kuramoto and Koizumi 2016).

It seems that the criticism has had little effect on raising awareness about the truth of PISA among policy makers, educators, and the public. It has yet to dismantle the marketing propaganda that PISA is an accurate, reliable, and valid measure of educational quality and essential abilities to live successfully in the twenty-first Century. More worrisome is that PISA has seen its power expanding amidst the criticism.

The expanding power of PISA is evidenced by the increase of education jurisdictions participating in the test, from 32 in 2000 (OECD/UNESCO 2003) to 72 in 2015 (OECD 2016b). The increase in the number of participating education systems does not only mean PISA has power over more education systems but also makes PISA more powerful because it shows the majority of the world’s education systems have accepted PISA as a legitimate entity. PISA’s expanding power is also evidenced and buttressed by its expansion in the coverage of domains, from reading, math, and science to the addition of financial literacy, collaborative problem solving, global competency (OECD 2018d), and possibly creativity (Gewertz 2018), as well

as its expanding line of products such as PISA for Schools (OECD 2018c), PISA for Development (OECD 2018b), and the “Baby PISA” (OECD nd; Pence 2016; Urban 2017).

The Baby PISA, nickname for the International Early Learning and Child Well-being Study, which aims to assess 5 year olds around the world in the same way as PISA does with 15 year olds, is set to start collecting data in 2018 and release its report in 2020. This means that the PISA enterprise will be the arbiter of not only quality of schools but also quality of pre-schools, communities, and families. If the history of PISA is any indication of OECD’s expertise in amassing political and media attention, the world can expect no less of a frenzy than that caused by the release of PISA results. The Baby PISA is likely to wield as much influence as PISA has over the world of early childhood education. Together with the Programme for International Assessment of Adult Competencies (PIAAC), the OECD product that claims to assess skills of 16–65 year olds (OECD 2018a), the PISA enterprise will be the most influential single institution in global education politics, policy, and practice from pre-school to high school to retirement.

The growing power of PISA does not mean the criticism is not valid or PISA has improved. It simply means that the criticism has not had its intended effect for many reasons such as the global political context (Lundgren 2011; Sjøberg 2012; Tröhler 2013), the natural human fallacy to seek ranking (Gould 1996), and certainly the ways the criticisms have been presented, which may have been overly technical, appeared mostly in academic circles, and in typical academic fashion, presented with too much concerns for balance and humility. Regardless, the lack of impact is no reason to give up exposing PISA as a flawed business that has great power to misguide education. The expanding influence of the PISA enterprise makes it even more important to be critical of this juggernaut today. It is also important to consider more effective and more straightforward approaches to present the criticism.

This article is an attempt to present the criticism in a different way from most PISA critiques. The purpose is to present a summary of criticisms that reveal the most fundamental flaws of PISA in non-technical language in one place. It aims to be straightforward and thus deliberately chose not to present a balanced view of pros and cons of PISA. Furthermore, the powerful PISA operator has already successfully shared its pros with the world of education and the cons have been ignored. In other words, this article is intentionally biased against PISA. The remainder of this article has three major sections, each including criticisms of one of PISA’s three fundamental deficiencies: its underlying view of education, its implementation, and its interpretation and impact on education globally.

## **Illusion of excellence: Distorted view of education**

The success of PISA is an excellent lesson for students of marketing. It starts by tapping into the universal anxiety about the future. Humans are naturally concerned about the future and have a strong desire to know if tomorrow is better than, or at least as good as, today. Parents want to know if their children will have a good life; politicians want to know if their nations have the people to build a

more prosperous economy; the public wants to know if the young will become successful and contributing members of the society.

PISA brilliantly exploits the anxiety and desire of parents, politicians, and the public with three questions:

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? (OECD 1999, p. 7).

These words begin the document that introduced PISA to the world in 1999 and have been repeated in virtually all PISA reports ever since (Sjøberg 2015b). The document then states the obvious: “Parents, students, the public and those who run education systems need to know” (OECD 1999, p. 7). And as can be expected, PISA offers itself as the fortuneteller by claiming that:

PISA assesses the extent to which 15-year-old students, near the end of their compulsory education, have acquired key knowledge and skills that are essential for full participation in modern societies. ... The assessment does not just ascertain whether students can reproduce knowledge; it also examines how well students can extrapolate from what they have learned and can apply that knowledge in unfamiliar settings, both in and outside of school. This approach reflects the fact that modern economies reward individuals not for what they know, but for what they can do with what they know. (OECD 2016a, p. 25).

This claim not only offers PISA as *a* tool to sooth anxiety but also, and perhaps more importantly, makes it *the* tool for such purpose because it helps to knock out its competitors. As an international education assessment, PISA came late. Prior to PISA, the International Association for the Evaluation of Educational Achievement (IEA) had already been operating international assessments since the 1960s, offering influential programs such as TIMSS and PIRLS. For a start-up to beat the establishment, it must offer something different and better. That’s exactly what PISA promised: a different and better assessment.

The IEA “surveys have concentrated on outcomes linked directly to the curriculum and then only to those parts of the curriculum that are essentially common across the participating countries” (OECD 1999, p. 10) and that’s a problem according to PISA because:

School curricula are traditionally constructed largely in terms of bodies of information and techniques to be mastered. They traditionally focus less, within curriculum areas, on the skills to be developed in each domain for use generally in adult life. They focus even less on more general competencies, developed across the curriculum, to solve problems and apply one’s ideas and understanding to situations encountered in life. (OECD 1999, p. 10).

PISA overcomes the limitations by assessing “what skills are deemed to be essential for future life,” which may or may not be covered by school curriculum. So

it claims. In other words, PISA asserts that other international surveys measure how well students have mastered the intended school curriculum of education systems, but the school curriculum could be misaligned with what is needed for future life.

To make the offer even better, PISA makes another seductive claim to education policy makers: “By directly testing for knowledge and skills close to the end of basic schooling, OECD/PISA examines the degree of preparedness of young people for adult life and, to some extent, the effectiveness of education systems,” (OECD 1999, p. 11). To paraphrase, PISA not only tells you if your children are prepared for future life, but also tells you that you have control over it through improving “the effectiveness of education.” Thus, “if schools and education systems are to be encouraged to focus on modern challenges,” PISA is needed.

PISA’s claim to measure essential life skills needed for the future and the effectiveness of education systems in instilling these skills was especially compelling because of the zeitgeist of the 1990s (Lundgren 2011). The arrival of the “flat world” (Friedman 2007) and the rise of new global powers such as China and India intensified anxiety about an uncertain future and one’s competitiveness in the new global society. Competitiveness in the new society is less associated with natural resources and more dependent on human intellectual resources. Worried political leaders were eager to make sure that their nations were cultivating the right intellectual resources needed to win the global competition. A tool that allows them know in advance how well their nations were doing and tell them what corrections to make to their education systems was more than a godsend. Thus it did not take much for PISA to be accepted as the global gold standard of future preparedness and in turn educational excellence. PISA scores and standings on the league tables are equated with the level of preparedness of a country’s youth for the future world. High performing countries logically become models to emulate or beat. “Every minister of education realized or believed in the necessity to be better than Finland” (Lundgren 2011, p. 28).

However, the claim, the foundation upon which PISA has built its success, has been roundly criticized since the beginning. The criticism falls into three categories. First, there is no evidence to justify, let alone prove, the claim that PISA indeed measures skills that are essential for life in modern economies. Second, the claim is an imposition of a monolithic and West-centric view of societies on the rest of the world. Third, the claim distorts the purpose of education.

### **Made-up claim**

The claim that PISA measures knowledge and skills essential for the modern society or the future world is not based on any empirical evidence:

There is no research available that proves this assertion beyond the point that knowing something is always good and knowing more is better. There is not even research showing that PISA covers enough to be representative of the school subjects involved or the general knowledge-base. PISA items are based on the practical reasoning of its researchers and on pre-tests of what works

in most or all settings – and not on systematic research on current or future knowledge structures and needs. (Hopmann 2008, p. 438).

In other words, the claim was just a fantasy, an illusion, entirely made up by the PISA team. But PISA keeps repeating its assertion that measures skills needed for the future. The strategy worked. PISA successfully convinced people through repetition (Labaree 2014; Sjøberg 2015b).

Furthermore, there is empirical evidence that suggests what PISA measures is not significantly different from other international assessments or intelligence tests (Hanushek and Woessmann 2010; Hopfenbeck et al. 2018; Nyborg 2007). For example, despite PISA's claim to measure something different from IEA-sponsored studies such as TIMSS, performance on PISA is significantly correlated with TIMSS (Wu 2009). Eric Hanushek, an economist who coauthored the influential report that supports PISA's claim admits: "in fact, the TIMSS tests with their curricular focus and the PISA tests with their applied focus are highly correlated at the country level" (Hanushek and Woessmann 2010, p. 38). Furthermore, a large analysis of correlations between PISA, IEA-studies, and other measures of cognitive ability found:

The cross-national correlations between different scales, between different studies (e.g. grades/ages, measurement points, used scales) and between different approaches (e.g. IEA vs. OECD, grade-level vs. age-level, student assessment vs. intelligence tests) were generally high. Factor analyses supported a strong g-factor. Different scales of student assessment studies and different cognitive test approaches appear to have measured essentially the same construct, namely general national cognitive ability. (Rindermann 2007, p. 697).

And ironically, the PISA project used results from other studies to support its case. PISA published an influential report aimed at demonstrating the importance of what it measures for economic development (Hanushek and Woessmann 2010). The report made a number of stunning claims about the long term economic impact of improving PISA outcomes, including, for example, "having all OECD countries boost their average PISA scores by 25 points over the next 20 years ... implies an aggregate gain of OECD GDP of USD 115 trillion over the lifetime of the generation born in 2010" (Hanushek and Woessmann 2010, p. 6).

The report has been challenged by a number of scholars since its release (Kamens 2015; Klees 2016; Komatsu and Rappleye 2017; Stromquist 2016). One of the most devastating problems with the conclusion of significant relationship between test scores and economic growth is the logic underlying the analysis utilized to reach the conclusion. The report compared test scores in a given period (1964–2003) with economic growth during roughly the same period (1960–2000), which is logically flawed because the students who took the test were not in the workforce at the time. It takes time for the students to enter the workforce and make up a significant portion of the workforce. Thus "test scores of students in any given period should be compared with economic growth in a subsequent period" (Komatsu and Rappleye 2017, p. 170). Studies that compared test scores with economic growth in the subsequent periods using the same dataset and method found no "consistently strong nor strongly consistent" relationship between test scores and economic growth and

“that the relationship between changes in test scores in one period and changes in economic growth for subsequent periods were unclear at best, doubtful at worst (Komatsu and Rappleye 2017, p. 183), essentially invalidating the claims made in the report.

Even if the claims were valid, they primarily relied on results of international assessments besides PISA. While the report states that it “uses recent economic modeling to relate cognitive skills—as measured by PISA and other international instruments—to economic growth” (Hanushek and Woessmann 2010, p. 6), the fact is that results from PISA constituted a very small portion of the data used in the modeling. Only three rounds of PISA had been offered by the time the report was released. Moreover, the economic data covered the period of 1960 to 2000, the year when PISA was first implemented. Only one round of PISA data was included but the report relied on “data from international tests given over the past 45 years in order to develop a single comparable measure of skills for each country that can be used to index skills of individuals in the labour force” (Hanushek and Woessmann 2010, p. 14).

Hanushek and others (Hanushek 2013; Hanushek and Woessmann 2008, 2012) have repeated similar claims about the economic impact of improving PISA. Whether the conclusions are correct is a different matter. The point is that PISA’s claim to measure something different from other international assessments is a lie. It indeed measures the same construct as others. The claim to better measure what matters in the modern economy or the future world than other tests that had been in existence prior to the invention of PISA is but a made-up illusion.

## A monolithic view of education

Underlying PISA’s claim is the assumption that there is a set of skills and knowledge that are universally valuable in all societies, regardless of their history and future. “A fundamental premise for the PISA project is that it is indeed possible to—measure the quality of a country’s education by indicators that are common, i.e. universal, independent of school systems, social structure, traditions, culture, natural conditions, ways of living, modes of production etc.” (Sjøberg 2015b, p. 116). But this assumption is problematic.

The first problem is that there is more than one society in the world and societies are different from each other. For all sorts of reasons—cultural, political, religious, and economical—different societies operate differently and present different challenges. Meeting different challenges requires different knowledge and skills. As a result, “one can hardly assume that the 15-year olds in e.g. USA, Japan, Turkey, Mexico and Norway are preparing for the same challenges and that they need identical life skills and competencies” (Sjøberg 2015b, p. 116).

The second and a bigger problem with PISA’s assumption of a universal set of valuable skills and knowledge for all countries is its imposition of a monolithic, primarily Western view of societies. PISA was first and foremost developed to serve member states of OECD, the majority of which are the world’s most advanced economies with only a few exceptions such as Mexico, Chile and Turkey. The 35

OECD members in no way represent the full spectrum of diversity across the nearly 200 countries in the world today. The assumptions supporting PISA are primarily based on the economic and education reality of OECD members. Not surprisingly, “the PISA framework and its test are meant for the relatively rich and modernized OECD-countries. When this instrument is used as a ‘benchmark’ standard in the 30+ non-OECD countries that take part in PISA, the mismatch of the PISA test with the needs of the nation and its youth may become even more obvious” (Sjøberg 2015b, p. 116).

### **Distorted view of education**

Although PISA claims that it does not assess according to national curricula or school knowledge, its results have been interpreted as a valid measure of the quality of educational systems. PISA reports have not been shy about making education policy recommendations either (Loveless 2012; Sjøberg 2015b). As a political force, PISA has successfully worked to harmonize and universalize its definition of education quality (Lundgren 2011; Popkewitz 2011; Sjøberg 2015b). Consequently, what PISA claims to measure essentially defines what school systems should aim to cultivate or the purpose of education. As a result, national education systems no longer need to discuss or reflect on the purpose of education. They only need to find ways to improve what PISA measures (Uljens 2007).

But the view of education promoted by PISA is a distorted and extremely narrow one (Berliner 2011; Sjøberg 2015b; Uljens 2007). PISA treats economic growth and competitiveness as the sole purpose of education. Thus it only assesses subjects—reading, math, science, financial literacy, and problem solving—that are generally viewed as important for boosting competitiveness in the global economy driven by science and technology. PISA shows little interest in other subjects that have occupied the curricula of many countries such as the humanities, arts and music, physical education, social sciences, world languages, history, and geography (Sjøberg 2015b).

While preparing children for economic participation is certainly part of the responsibility of educational institutions, it cannot and should not be the only responsibility (Labaree 1997; Sjøberg 2015b; Zhao 2014, 2016b). The purpose of education in many countries includes a lot more than preparing economic beings. Citizenship, solidarity, equity, curiosity and engagement, compassion, empathy, curiosity, cultural values, physical and mental health, and many others are some of the frequently mentioned purposes in national education goal states. But these aspects of purpose of education “are often forgotten or ignored when discussions about the quality of the school is based on PISA scores and rankings” (Sjøberg 2015b, p. 113).

The distorted and narrow definition of the purpose of education is one of the major reasons for some of the peculiar and seemingly surprising discoveries associated with PISA. There is the persistent pattern of negative correlation between PISA scores and students’ interest and attitude. Many researchers have found that higher PISA scoring countries seem to have students with lower interest in and less



positive attitude toward the tested subject (Bybee and McCrae 2011; Zhao 2012, 2014, 2016b). For example, PISA science score has a significant negative correlation with future science orientation and with future science jobs (Kjærnsli and Lie 2011). Higher PISA scores have also been found to be associated with lower entrepreneurship confidence and capabilities (Campbell 2013; Zhao 2012). Moreover, high PISA scoring education systems seemed to have a more authoritarian orientation (Shirley 2017; Zhao 2014, 2016b). Additionally, PISA scores have been found to have a negative correlation with student wellbeing (Shirley 2017; Zhao 2014, 2016b), a finding that was finally openly acknowledged by PISA in a 2017 report (OECD 2017). These findings basically suggest that PISA only measures a very narrow aspect of education and neglects to pay attention to the broader responsibilities of educational systems. Furthermore, pursuing the narrowly defined purpose of education may come at the cost of the broader purpose of education (Zhao 2017b, 2018c). “There are very few things you can summarise with a number and yet PISA claims to be able to capture a country’s entire education system in just three of them. It can’t be possible. It is madness” (Morrison 2013).

In summary, PISA successfully marketed itself as an indicator of educational excellence with the claim to measure skills and knowledge that matters in modern economies and in the future world. Upon closer examination, the excellence defined by PISA is but an illusion, a manufactured claim without any empirical evidence. Furthermore, PISA implies and espouses a monolithic, distorted, and narrow view of purpose for all education systems in the world. The consequence is a trend of global homogenization of education and celebration of authoritarian education systems for their high PISA scores, while ignoring the negative consequences on important human attributes and local cultures of such systems.

## Illusion of science: Flawed implementation

Another reason for PISA’s rapid rise as the gold standard of quality of education is its claim to be scientific. And indeed, PISA appears to be a scientific enterprise, following a seemingly scientific approach to developing test items, sampling, collecting data, analyzing results, and reporting findings. It uses sophisticated psychometric theories, complex statistic modeling, established sampling methods, and rigorous implementation procedures to arrive at the findings. More impressively, PISA masterfully employs data representation methods to make its reporting attractive and look scientific, with excellent visuals and all sorts of numbers. According to Sjøberg:

It [PISA] has many of the characteristics of what is called—Big science and—techno-science: It is costly; it involves the cooperation of around 70 countries. The logistics of the project is complicated, and there are piles of documents with detailed instructions to the national groups who are responsible in the participating countries. Hundreds of experts from several fields of expertise are involved, contracts with subcontractors are given by bids, thousands of schools and teachers, nearly half a million of students spend 2½ hours answer-

ing the test and the questionnaire, data are carefully coded by thousands of specially trained markers etc. etc. (Sjøberg 2015b #5159, p. 121).

Despite the appearance of being scientific, PISA has a slew of inherent and implementation problems that threaten the quality of its findings. These problems have been reported since the beginning of PISA and finally caught the attention of the media lately. The popular UK education magazine *TES* published an article entitled *Is PISA Fundamentally Flawed* in 2013. The article put forth a number of sobering questions about the PISA data:

what if there are “serious problems” with the Pisa data? What if the statistical techniques used to compile it are “utterly wrong” and based on a “profound conceptual error”? Suppose the whole idea of being able to accurately rank such diverse education systems is “meaningless”, “madness”? What if you learned that Pisa’s comparisons are not based on a common test, but on different students answering different questions? And what if switching these questions around leads to huge variations in the all-important Pisa rankings...? What if these rankings—that so many reputations and billions of pounds depend on, that have so much impact on students and teachers around the world—are in fact “useless”? (Stewart 2013 #3473).

### Sampling problems

PISA’s claimed representativeness has been challenged due to a number of sampling problems. The first problem with PISA’s sampling method is using age as the criterion instead of grade level “because the sample would include a sizeable proportion of students who had either repeated a class or skipped one” (Hopfenbeck et al. 2018). This means that the sample would include students at different grade levels, which means different exposure to the school curriculum. Furthermore, it does not take into account school start age, which varies from country to country. While PISA claims not to assess curriculum-related knowledge and skills, exposure to curriculum and schooling certainly matters in the results.

The second sampling problem is the issue of representativeness. Researchers have found that PISA samples do not have the level of representativeness that supports its claim. For example, low participation rates and lack of representativeness in the 2003 cycle of PISA in England were raised as a serious concern in the interpretation of PISA (Prais 2003). It was found that nearly half of the population of 15-year-olds was not included in the sampling frame in the 2012 cycle.

And although this is not an inherent sampling problem (as indicated by a well-covered target population), it certainly precludes any generalization of PISA results to the entire population of 15-year-olds (that eventually enter the workforce). Such low coverage weakens OECD claims that the average level of skills measured by PISA is an “important indicator of human capital, which in turn has an impact on the prosperity and well-being of society as a whole” (OECD 2013, p. 169). Clearly, as an overall indicator of human

capital, PISA will necessarily be limited by the fact that 15-year-olds not enrolled in schools are outside of the target population. (Rutkowski and Rutkowski 2016, p. 253).

The third problem is the exclusion of students (Hopfenbeck et al. 2018). PISA allows students with certain characteristics to be excluded from participation but it stipulates that the overall exclusion rate be below 5%. However, the exclusion rates varied a great deal in different education systems in previous cycles of the assessment. For example, eight education systems reported an exclusion rate over 5% in the 2012 cycle (Rutkowski and Rutkowski 2016). Another issue is that PISA excludes students with disabilities. The exclusion of students with disabilities belies PISA's claim to capture the quality and equity of education of education systems and further marginalize students with special needs (Schuelka 2013). Furthermore, criticism has been raised with regard to school systems that actively exclude certain students from attending the schools PISA sampled. For example, Loveless has raised serious concerns about the exclusion of migrant children from the sample in Shanghai (Loveless 2014).

### **Biased assessments**

PISA assessments have been criticized for being biased. The bias stem from multiple sources including item formats, constructs, language, culture, and text types (Hopfenbeck et al. 2018; Solheim and Lundetræ 2018). For example, PISA reading assessment has been found to be girl-friendly and thus biased against boys in Nordic countries because “PISA measures reading literacy based on a mix of fiction and non-fiction texts, and it includes a high proportion of items that require respondents to document their reading comprehension through writing” (Solheim and Lundetræ 2018, p. 121). PISA instruments have been found to be more comparable across Western countries than they are across Middle Eastern or Asian countries for linguistic and cultural reasons (Grisay et al. 2007; Grisay and Gonzalez 2009).

A comprehensive review of studies that examined the biases of PISA found that “[m]ost of the studies reported a substantial amount of DIF when comparing differential language versions” (Hopfenbeck et al. 2018, p. 344). DIF, short for differential item functioning, is a statistical characteristic of a test item that shows the degree to which the item might be measuring different abilities for members of different subgroups. In this case, it indicates that PISA may be measuring different abilities for students of different languages. For example, the linguistic discrepancy between PISA items and Greek textbooks may be a cause for Greek students' low performance in science (Hatzinikita et al. 2008). PISA is also biased for or against students of different language groups because of its variation in length, i.e., the assessment can be longer in some languages than the original. The German versions, for instance, are 18% longer than the English versions (Eivers 2010). But PISA allocated the same amount of time for all languages. As a result, students using the longer versions face more time pressure.

## Inappropriate approach

PISA uses the Rasch model, a widely used psychometric model named after the late Danish mathematician and statistician Georg Rasch (1901–1980), to derive its results. For this model to work properly, certain requirements must be met. But according to Kreiner, who studied under Rasch and has worked with his model for over 40 years, PISA's application does not meet those requirements. Kreiner and Christensen found that the Rasch model does not fit the reading literacy data of PISA, and thus the resulting country rankings are not robust (Kreiner and Christensen 2014). As a result, rankings of countries can vary a great deal over different subsets. "That means that [PISA] comparisons between countries are meaningless," according to Kreiner (Stewart 2013). Other analyses found that the PISA data fit a multidimensional model with two factors better (Goldstein et al. 2007; Hopfenbeck et al. 2018).

In summary, PISA's serious technical flaws challenge its claim of being scientific. In 2007, a collection of nearly 20 researchers from multiple European countries presented their critical analyses in the book *PISA According to PISA: Does PISA Keep What It Promises* (Hopmann et al. 2007). Independent scholars took apart PISA's methodology, examining how it was designed; how it sampled, collected, and presented data; and what its outcomes were. Almost all of them "raise[d] serious doubts concerning the theoretical and methodological standards applied within PISA, and particularly to its most prominent by-products, its national league tables or analyses of school systems" (Hopmann et al. 2007, p. 10). Among their conclusions:

- PISA is by design culturally biased and methodologically constrained to a degree which prohibits accurate representations of what actually is achieved in and by schools. Nor is there any proof that what it covers is a valid conceptualization of what every student should know.
- The product of most public value, the national league tables, are based on so many weak links that they should be abandoned right away. If only a few of the methodological issues raised in this volume are on target, the league tables depend on assumptions about the validity and reliability which are unattainable.
- The widely discussed by-products of PISA, such as the analyses of "good schools," "good instruction" or differences between school systems ... go far beyond what a cautious approach to these data allows for. They are more often than not speculative... (Hopmann et al. 2007, pp. 12–13)

## Illusion of progress: Flawed interpretation and two decades of havoc

Andreas Schleicher, the orchestrator of PISA, proudly shares his view of the progress and influence of PISA in his latest book *World Class: How to Build a twenty-first-Century School System*:

Over the years, PISA established itself as an influential force for education reform. The triennial assessment has helped policy makers lower the cost of political action by backing difficult decisions with evidence. But it has also raised the political cost of inaction by exposing areas where policy and practice were unsatisfactory. Two years after that first meeting around a table in Paris, 28 countries signed on to participate. Today, PISA brings together more than 90 countries, representing 80% of the world economy, in a global conversation about education. (Schleicher 2018, p. 20)

Schleicher should be proud of PISA's success. But the success is only measured in terms of its expanding influence over global education. Whether this success has led global education to a better place is questionable. Proponents of PISA argue that the international assessment has successfully caught the attention of politicians. The program has been able to pressure politicians into action with its international league tables, which publicly celebrate the best and shame the worst PISA performers. As evidence, Schleicher (2018) puts multiple examples of countries that have launched massive changes in education, including his home country Germany. The changes are often about emulating "excellent" education systems according to PISA results such as Finland, Shanghai, and Singapore.

There is no question that PISA has led to massive changes in global education but there is the question that whether these changes have made or will make education better in the world. PISA critics would agree with its proponents that the program has been a success, but argue that the success has not made, and will not make, education better. In fact, many have pointed out that PISA has successfully made education worse. It has wrecked havoc in the world of education in a number of significant ways.

### Homogenizing education

PISA has been criticized for successfully putting the world's education on a path of homogenization (Lingard et al. 2015; Sellar and Lingard 2014; Zhao and Gearin 2016). Education systems, processes, and values around the world are becoming increasingly homogenous and standardized (Carney et al. 2012; Green and Mostafa 2013; Tröhler 2013; Zhao 2009, 2012), partly due to the Global Education Reform Movement (GERM) (Sahlberg 2012), which has been buttressed by PISA and other international assessments. Educational changes in the world in recent years are strikingly similar: centralizing curriculum and standards making, global benchmarking in curriculum and standards, strengthening testing and accountability measures, and encouraging local autonomy and market competition so as to achieve the same outcomes as defined by PISA (Breakspear 2012). PISA has also been promoting similar global efforts to standardize teacher recruitment, preparation, and incentivization (Barber and Mourshed 2007; Tucker 2011; Zhao 2018a). As a result, students' experiences are also increasingly homogenized globally as countries emulate or attempt to emulate successes of high performing nations: more time on academic studies and narrower focus on PISA subjects—reading, math, and science (Zhao 2017a, 2018b).

It is certainly not a problem that countries learn from each other but it is dangerous to have all countries, regardless of their local needs and contexts, to implement the same policies and practices because educational policies and practices are contextual, culturally sensitive. What works in one context may not work or even cause damage in another (Harris and Jones 2015; Zhao 2018b, c). What is needed for 15 year olds in one country can differ drastically from another (Sjøberg 2015b).

Furthermore, the policies and practices promoted by PISA are problematic. They can be either mistakenly drawn from PISA's flawed data or have negative consequences while leading to high PISA performance. For example, the policy recommendation that teachers should be recruited from top high school graduates because high performing systems do so (Barber and Mourshed 2007) resulted from erroneous observations (Zhao 2018a). Finland, for example, does not have such a policy or practice (Sahlberg 2017). This policy can also do damage because research shows that top graduates do not necessarily benefit all students (Zhao 2018a, c). In fact, teachers with high academic performances in secondary schools benefit high performing students but hurt low performing students (Grönqvist and Vlachos 2016).

Another danger of a homogenized education worldwide is the reduction of diversity in educational approaches, values, processes, and most important, talents. Even if all PISA promoted policies and practices were perfect, and obviously they are not, and they were universally adopted, they could become outdated as societies change. If societies meet a transformational change such the waves of technological revolutions, education in the world would become obsolete and new ideas must be sought. But there would be no source of new ideas because everyone would have been implementing PISA promoted policies and practices.

The same could happen with talents. If everyone became a high PISA performer while sacrificing individual uniqueness, when societies change, they would all become unfit for the new society. A homogenized and standardized education has been found to “squeeze out” creative talents as well (Zhao 2014; Zhao and Gearin 2016). This is akin to growing only one type of potato, which may be the type that is most productive. But it can be wiped out completely when attacked by a disease.

## Stifling innovation

PISA has also been criticized for stifling innovation. East Asian countries have been working hard to introduce significant reforms since the 1990s to overcome the apparent shortcomings of its education system in order to cultivate a more diverse, creative, and entrepreneurial citizenry (Zhao 2015; Zhao and Wang 2018). But PISA has weakened the desire of politicians to make drastic changes because it has put these education systems on a pedestal. Putting someone on a pedestal is an effective way to ensure he does not veer far from his previous behaviors because any deviation could tarnish the bestowed honor.

For example, China has taken on its rigid exam system as a target of reform because it has long been recognized as the culprit for limiting capacity for producing creative and diverse talents (Zhao 2009, 2014). But just as China's education reforms began to touch the core of the system—the *gaokao* or College Entrance

Exam, PISA proclaimed that Chinese education is the best in the world. And the exam system, including the *gaokao*, is glorified as a major contributor to China's success (Schleicher 2018; Tucker 2011, 2014, 2016). This made it very difficult for the Chinese government to continue the battle against testing. Even Marc Tucker, one of the most prominent PISA proponents who has on many occasions expressed unequivocal admiration of China's education, admits:

... many people in China are upset about the success of Shanghai on the PISA league tables, because they think that success will blunt the edge of their fight to dethrone the *Gaokao* from its premier position as the sole determinant of advancement in Chinese society. They see the *Gaokao* as enforcing an out-dated ideal of education, one that rewards memorization and rote learning over understanding and the ability to apply mastery of complex skills to real world problems, particularly problems requiring innovation and creativity (Tucker 2014, p. 4).

PISA had a similar effect on Japan's education reform. PISA "played a role in the decision to reverse, at least in part, the *yutori* reform launched at the beginning of the decade" (Jones 2011, p. 22). *Yutori* kyoiku (roughly translated "relaxed education" or education with some freedom) was a major education reform movement started in the 1980s in Japan. "The *yutori* reform was based on an emerging consensus that the school system was too rigid and that a new approach was needed to encourage creativity" (Jones 2011, p. 13). The major changes included reduction in school days and a 30% cut in the school curriculum (Jones 2011; Schleicher 2018). "In addition, the government relaxed grading practices and introduced "integrated learning classes" without textbooks in an effort to help students think independently and reduce the importance of rote learning" (Jones 2011, p. 13). The changes were announced in 1998 and implemented in 2002.

In 2003, Japan's PISA rankings fell, resulting in a public panic over Japan's decline in international academic standing. Opponents of the *yutori* reform seized the moment and blamed the reform for the decline. In response, Japan decided to water down the previous reforms with an increase in required topics in standard academic subjects, increasing time devoted to these subjects, and introducing national standardized testing in math and Japanese for the first time in 2007 (Jones 2011). Schleicher himself writes about how PISA reversed the reform, despite its positive educational outcomes:

When results from PISA showed a decline in mathematics performance in 2003, parents lost confidence that the reformed curriculum would prepare their children for the challenges that lay ahead... But pressure mounted to reverse the reform, and over the past few years curriculum content became more dominant again. (Schleicher 2018, p. 76).

### **Propagating harmful half-truths**

PISA has been criticized for promoting myths or half-truths about education that can be harmful (Zhao 2016b). Thanks to its distorted view of education and flawed

implementation, PISA has resulted in a number of myths and it has not been shy about spreading them. One example is the aforementioned myth that high performing education systems recruit top graduates into the teaching force, which has been debunked (Sahlberg 2017; Zhao 2018a). Another myth that is that students in high performing systems take more responsibility for their own learning, thus blaming themselves for academic failures, which is essentially a form of self-condemnation.

### **Romanticizing misery**

Schleicher believes that self-condemnation is a valuable trait and worth cultivating (Zhao 2016b). He uses self-condemnation to explain the superb Chinese PISA performance. On many occasions, Schleicher promoted the idea that Chinese students take responsibilities for their own learning, while in “many countries, students were quick to blame everyone but themselves” (Schleicher 2013). French students blamed their teachers, for instance. Schleicher maintains that this difference in attitude contributed to the gap between Shanghai, ranked first, and France, ranked 25th. “The fact that students in some countries consistently believe that achievement is mainly a product of hard work, rather than inherited intelligence, suggests that education and its social context can make a difference in instilling the values that foster success in education” (Schleicher 2013).

However, Schleicher’s observation is inconsistent with reality. There are plenty of countries that have higher PISA rankings than France, yet reported similar attitudes (Zhao 2016b). Moreover, the PISA report contradicts Schleicher’s reasoning because it finds that students with lower scores tend to take more responsibility:

Overall, the groups of students who tend to perform more poorly in mathematics—girls and socio-economically disadvantaged students—feel more responsible for failing mathematics tests than students who generally perform at higher levels (OECD 2013, p. 62).

The fact is that the degree to which students take responsibility for failing in math or blaming outside factors does not have much to do with their PISA performance. The percentage of students who attribute their failing in math to teachers: countries with low percentages of students saying “my teacher did explain the concepts well this week” or “my teacher did not get students interested in the material” do not necessarily have the best ranking. Conversely, countries where students are more likely to blame teachers are not necessarily poor performers (Zhao 2016b).

### **Glorifying authoritarian education**

Self-condemnation is more likely a result of authoritarian education than a contributor to high performance. The countries whose students are least likely to blame their teachers have a more authoritarian cultural tradition than the countries whose students are most likely to blame their teachers. Students in more authoritarian education systems are more likely to blame themselves and less likely to question the authority—the teacher—than students in more democratic educational systems. Authoritarian



education demands obedience and does not tolerate questioning of authority. They have externally defined high expectations that are not necessarily accepted by students intrinsically. They require mandatory conformity through rigid rules and severe punishment for noncompliance. More important, the authority works very hard to convince children to blame themselves for failing to meet the expectations. As a result, they produce students with low confidence and low self-esteem (Zhao 2016b).

### **A ploy to deny responsibility**

Blaming themselves for not putting in enough efforts to achieve academic success, while praised by PISA proponents as Shanghai's secret to educational excellence, is in fact a desired outcome of authoritarian societies. In an authoritarian system, the ruler and the ruling class have much to gain when people believe it is their own effort, and nothing else, that makes them successful. The authority wants people to believe that their innate abilities or social circumstances do not matter in educational outcomes. If they cannot succeed, they only have themselves to blame. This is an excellent and convenient way for the authorities to deny any responsibility for social inequity and injustice, and to avoid accommodating differently talented people and those born into disadvantaged situations. It is an effective ploy that helps the elite and authority to fool people into accepting the inequalities they were born into (Zhao 2016a, b).

### **The future of PISA: Final words**

It is clear from the criticism raised against PISA that the entire enterprise is problematic and does not deserve the attention it enjoys from all over the world. Its flawed view of education, flawed implementation, and flawed reasoning are sufficient to discredit its findings and recommendations. Its negative impact on global education is emerging. Yet, its influence continues to expand.

What should happen to PISA? Views differ. Many critics would like to save PISA, despite its flaws. They would like to see PISA improving (Meyer et al. 2014) and also provide advice to PISA consumers on how to use PISA more wisely (Hargreaves and Sahlberg 2015). However, it is unlikely that PISA is going to improve significantly; at least not in the directions critics want. Technical improvement can be made. But technical fixes do little to alter PISA's view of education, the very basis upon which the entire PISA enterprise is built. Thus, a more reasonable action is probably to just ignore it entirely or wish for PISA to end itself, which cannot happen until and unless OECD members take actions to reject PISA.

### **References**

- Arzarello, F., Garuti, R., & Ricci, R. (2015). The impact of PISA studies on the Italian national assessment system. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy* (pp. 249–260). Berlin: Springer.

- Baird, J.-A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., et al. (2016). On the supranational spell of PISA in policy. *Educational Research*, 58(2), 121–138.
- Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. Retrieved August 3, 2010 from New York <https://www.mckinsey.com/industries/social-sector/our-insights/how-the-worlds-best-performing-school-systems-come-out-on-top>.
- Baroutsis, A., & Lingard, B. (2017). Counting and comparing school performance: An analysis of media coverage of PISA in Australia, 2000–2014. *Journal of Education Policy*, 32(4), 432–449.
- Berliner, D. C. (2011). The context for interpreting PISA results in the USA: Negativism, chauvinism, misunderstanding, and the potential to distort the educational systems of nations. In M. A. Peryera, H.-G. Kotthoff, & R. Cowen (Eds.), *Pisa under examination* (pp. 77–96). New York: Springer.
- Bieber, T., & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons from Switzerland and the US. *European Journal of Education*, 46(1), 101–116. <https://doi.org/10.1111/1j.1465-3435.2010.01462.x>.
- Breakspear, S. (2012). *The policy impact of Pisa: An exploration of the normative effects of international benchmarking in school system performance*. Retrieved July 3, 2016 from Paris [http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/WKP\(2012\)8&docLanguage=En](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=EDU/WKP(2012)8&docLanguage=En).
- Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, 33(1), 7–26.
- Campbell, M. (2013). West vs Asia education rankings are misleading: Western schoolchildren are routinely outperformed by their Asian peers, but worrying about it is pointless. *New Scientist* (2898). Retrieved January 10, 2015 from <https://www.newscientist.com/article/mg21728985-800-west-vs-asia-education-rankings-are-misleading/>.
- Carney, S., Rappleye, J., & Silova, I. (2012). Between faith and science: World culture theory and comparative education. *Comparative Education Review*, 56(3), 366–393.
- Domínguez, M., Vieira, M.-J., & Vidal, J. (2012). The impact of the programme for international student assessment on academic journals. *Assessment in Education: Principles, Policy & Practice*, 19(4), 393–409.
- Eivers, E. (2010). PISA: Issues in implementation and interpretation. *The Irish Journal of Education/Iris Eireannach an Oideachais*, 38, 94–118.
- Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: An ironic investigation. *Journal of Education Policy*, 29(6), 845–855. <https://doi.org/10.1080/02680939.2014.892156>.
- Figazzolo, L. (2009). *Impact of PISA 2006 on the education policy debate*. Retrieved September 9, 2015 from <http://download.ei-ie.org/docs/IRISDocuments/ResearchWebsiteDocuments/2009-00036-01-E.pdf>.
- Friedman, T. L. (2007). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus and Giroux.
- Gewertz, C. (2018). PISA working toward exam to gauge creative thinking—Built by ACT. *Education Week*, 38(6), 4.
- Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel structural equation models for the analysis of comparative data on educational performance. *Journal of Educational and Behavioral Statistics*, 32(3), 252–286. <https://doi.org/10.3102/1076998606298042>
- Gould, S. J. (1996). *The mismeasure of man*. New York: Norton.
- Green, A., & Mostafa, T. (2013). The dynamics of education systems: Convergent and divergent trends, 1990–2010. In J. Janmaat, M. Duru-Bellat, P. Méhaut, & A. Green (Eds.), *The dynamics and social outcomes of education systems* (pp. 15–45). Berlin: Springer.
- Grek, S. (2009). Governing by numbers: The PISA ‘effect’ in Europe. *Journal of Education Policy*, 24(1), 23–37.
- Grisay, A., De Jong, J., Gebhardt, E., Berezner, A., & Halleux-Monheur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement*, 8(3), 249.
- Grisay, A., & Gonzalez, E. (2009). Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 2, 63–83.
- Grönqvist, E., & Vlachos, J. (2016). One size fits all? The effects of teachers’ cognitive and social abilities on student achievement. *Labour Economics*, 42, 138–150.
- Gruber, K. H. (2006). *The German ‘PISA-Shock’: Some aspects of the extraordinary impact of the OECD’s PISA study on the German education system*. Paper presented at the Cross-national Attraction in Education: Accounts from England and Germany. Oxford: Symposium Books.

- Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review*, 37, 204–212.
- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46, 607–668.
- Hanushek, E. A., & Woessmann, L. (2010). *The high cost of low educational performance: The long-run economic impact of improving PISA Outcomes*. Retrieved March 20, 2011 from Paris [http://books.google.com/books?id=k7AGPo0NvfYC&pg=PA33&lpg=PA33&dq=hanushek+pisa+gdp&source=bl&ots=2gCfzF-f1\\_&sig=wwe0XLL5Eb1VWK9e7RJfb5MyhIU&hl=en&sa=X&ei=MLPCUqaOD8-JogS6v4C4Bw&ved=0CGcQ6AEwBjgK-v=onepage&q=hanushek%20pisa%20gdp&f=false](http://books.google.com/books?id=k7AGPo0NvfYC&pg=PA33&lpg=PA33&dq=hanushek+pisa+gdp&source=bl&ots=2gCfzF-f1_&sig=wwe0XLL5Eb1VWK9e7RJfb5MyhIU&hl=en&sa=X&ei=MLPCUqaOD8-JogS6v4C4Bw&ved=0CGcQ6AEwBjgK-v=onepage&q=hanushek%20pisa%20gdp&f=false).
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267–321.
- Hargreaves, A., & Sahlberg, P. (2015). *The tower of PISA is badly leaning. An argument for why it should be saved*. Retrieved April 5, 2017 from [https://www.washingtonpost.com/news/answer-sheet/wp/2015/03/24/the-tower-of-pisa-is-badly-leaning-an-argument-for-why-it-should-be-saved/?noredirect=on&utm\\_term=.69275346b5d3](https://www.washingtonpost.com/news/answer-sheet/wp/2015/03/24/the-tower-of-pisa-is-badly-leaning-an-argument-for-why-it-should-be-saved/?noredirect=on&utm_term=.69275346b5d3).
- Harris, A., & Jones, M. S. (2015). *Leading futures: Global perspectives on educational leadership*. New Delhi: SAGE Publications India.
- Hatzinikita, V., Dimopoulos, K., & Christidou, V. (2008). PISA test items and school textbooks related to science: A textual comparison. *Science Education*, 92(4), 664–687.
- Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J.-A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353. <https://doi.org/10.1080/00313831.2016.1258726>.
- Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417–456.
- Hopmann, S. T., Brinek, Gertrude, & Retzl, M. (Eds.). (2007). *PISA zufolge PISA—PISA according to PISA*. Berlin: Lit Verlag.
- Jones, R. S. (2011). *Education reform in Japan*. Paris: OECD. <https://doi.org/10.1787/5kg58z7g95np-en>.
- Kamens, D. H. (2015). A maturing global testing regime meets the world economy: Test scores and economic growth, 1960–2012. *Comparative Education Review*, 59(3), 420–446.
- Kjærnsli, M., & Lie, S. (2011). Students' preference for science careers: International comparisons based on PISA 2006. *International Journal of Science Education*, 33(1), 121–144.
- Klees, S. J. (2016). Human capital and rates of return: Brilliant ideas or ideological dead ends? *Comparative Education Review*, 60(4), 644–672.
- Komatsu, H., & Rapplepe, J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, 53(2), 166–191.
- 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*. <https://doi.org/10.1007/s11336-013-9347-z>.
- Kuramoto, N., & Koizumi, R. (2016). Current issues in large-scale educational assessment in Japan: Focus on national assessment of academic ability and university entrance examinations. *Assessment in Education: Principles, Policy & Practice*. <https://doi.org/10.1080/0969594X.2016.1225667>.
- Labaree, D. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal*, 34(1), 39–81.
- Labaree, D. F. (2014). Let's measure what no one teaches: PISA, NCLB, and the shrinking aims of education. *Teachers College Record*, 116(9), 1–14.
- Lewis, S. (2017). Policy, philanthropy and profit: The OECD's PISA for schools and new modes of heterarchical educational governance. *Comparative Education*, 53(4), 518–537. <https://doi.org/10.1080/03050068.2017.1327246>.
- Lingard, B., Rezai-Rashti, G., Martino, W., & Sellar, S. (2015). *Globalizing educational accountabilities*. London: Routledge.
- Loveless, T. (2012). *The 2012 brown center report on American education: How well are American students learning?* Retrieved January 20, 2015 from Washington DC [http://www.brookings.edu/~media/Files/rc/reports/2012/0216\\_brown\\_education\\_loveless/0216\\_brown\\_education\\_loveless.pdf](http://www.brookings.edu/~media/Files/rc/reports/2012/0216_brown_education_loveless/0216_brown_education_loveless.pdf).

- Loveless, T. (2014). *PISA's China problem continues: A response to Schleicher, Zhang, and Tucker*. Retrieved January 20, 2015 from <http://www.brookings.edu/research/papers/2014/01/08-shanghai-pisa-loveless>.
- 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* (pp. 17–30). New York: Springer.
- Meyer, H.-D., et. al. (2014). *OECD and Pisa tests are damaging education worldwide—academics*. Retrieved February 3, 2015 from <https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>.
- Meyer, H.-D., & Benavot, A. (2013). *PISA, power, and policy: The emergence of global educational governance*. Oxford: Oxford University Press.
- Morrison, H. (2013). *Pisa 2012 major flaw exposed*. Retrieved February 3, 2015 from <https://pacen.i.wordpress.com/2013/12/01/pisa-2012-major-flaw-exposed/>.
- Ninomiya, S. (2016). The impact of PISA and the interrelation and development of assessment policy and assessment theory in Japan. *Assessment in Education: Principles, Policy & Practice*, 26, 1–20.
- Nyborg, H. (2007). Do recent large-scale cross-national student assessment studies neglect general intelligence g for political reasons? *European Journal of Personality*, 21, 739–741.
- OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Retrieved March 15, 2019 from Paris: <http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33693997.pdf>.
- OECD. (2010). *OECD programme for international student assessment*. Retrieved March 15, 2019 from [http://www.pisa.oecd.org/pages/0,2987,en\\_32252351\\_32235731\\_1\\_1\\_1\\_1\\_1\\_1\\_1,00.html](http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1_1_1,00.html).
- OECD. (2013). *Ready to learn: Students engagement, drive, and self-beliefs*. Retrieved March 15, 2019 from Paris: <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-III.pdf>.
- OECD. (2016a). *PISA 2015 results (volume I): Excellence and equity in education*. Retrieved March 15, 2019 from Paris: <http://dx.doi.org/10.1787/9789264266490-en>.
- OECD. (2016b). *PISA 2015: Results in focus. PISA*. Retrieved March 15, 2019 from <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>.
- OECD. (2017). *PISA 2015 results: Students' well-being*. Retrieved March 15, 2019 from Paris [http://www.keepeek.com/Digital-Asset-Management/oecd/education/pisa-2015-results-volume-iii\\_9789264273856-en-Wk1WGrQ-fOQ#page1](http://www.keepeek.com/Digital-Asset-Management/oecd/education/pisa-2015-results-volume-iii_9789264273856-en-Wk1WGrQ-fOQ#page1).
- OECD. (2018a). *About PIAAC*. Retrieved March 15, 2019 from <http://www.oecd.org/skills/piaac/about/>.
- OECD. (2018b). *PISA for development*. Retrieved March 15, 2019 from <https://www.oecd.org/pisa/pisa-for-development/>.
- OECD. (2018c). *PISA for schools*. Retrieved March 15, 2019 from <http://www.oecd.org/PISA/pisa-for-schools/>.
- OECD. (2018d). *PISA: About*. Retrieved March 15, 2019 from <http://www.oecd.org/pisa/aboutpisa/>.
- OECD. (nd). *The international early learning and child well-being study—The study*. Retrieved March 15, 2019 from <http://www.oecd.org/education/school/the-international-early-learning-and-child-well-being-study-the-study.htm>.
- OECD/UNESCO. (2003). Literacy skills for the world of tomorrow—Further results from Pisa 2000. *PISA*. Retrieved October 23, 2017 from <http://www.oecd.org/education/school/2960581.pdf>.
- Pence, A. (2016). Baby PISA: Dangers that can arise when foundations shift. *Journal of Childhood Studies*, 41(3), 54–58.
- Pereyra, M. A., Kotthoff, H.-G., & Cowen, R. (2011). *PISA under examination*. New York: Springer.
- Popkewitz, T. (2011). PISA: Numbers, standardizing conduct, and the alchemy of school subjects. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *Pisa under examination* (pp. 31–46). New York: Springer.
- Prais, S. J. (2003). Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education*, 29, 139–163.
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21, 667–706.
- Rutkowski, L., & Rutkowski, D. (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher*, 45(4), 252–257.
- Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?*. New York: Teachers College Press.

- Sahlberg, P. (2012). *How GERM is infecting schools around the world*. Retrieved October 23, 2017 from [http://www.washingtonpost.com/blogs/answer-sheet/post/how-germ-is-infecting-schools-around-the-world/2012/06/29/gJQAVELZAW\\_blog.html-comments](http://www.washingtonpost.com/blogs/answer-sheet/post/how-germ-is-infecting-schools-around-the-world/2012/06/29/gJQAVELZAW_blog.html-comments).
- Sahlberg, P. (2017). *FinnishED leadership*. Thousand Oaks: Corwin.
- Schleicher, A. (2013). *Are the Chinese cheating in PISA or are we cheating ourselves?* Retrieved February 3, 2015 from <http://oecdeducationtoday.blogspot.com/2013/12/are-chinese-cheating-in-pisa-or-are-we.html>.
- Schleicher, A. (2018). *World class: How to build a 21st-century school system*. Paris: OECD.
- Schuelka, M. J. (2013). Excluding students with disabilities from the culture of achievement: The case of the TIMSS, PIRLS, and PISA. *Journal of Education Policy*, 28(2), 216–230.
- Sellar, S., & Lingard, B. (2014). The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*, 40(6), 917–936.
- Shirley, D. (2017). *The new imperatives of educational change: Achievement with integrity*. New York: Routledge.
- Sjøberg, S. (2012). PISA: Politics, fundamental problems and intriguing results. *Recherches en Education*, 14, 1–21.
- Sjøberg, S. (2015a). OECD, PISA, and globalization: The influence of the international assessment regime. In C. H. Tienken & C. A. Mullen (Eds.), *Education policy perils* (pp. 114–145). London: Routledge.
- Sjøberg, S. (2015b). PISA and global educational governance—A critique of the project, its uses and implications. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(1), 111–127.
- Solheim, O. J., & Lundetrae, K. (2018). Can test construction account for varying gender differences in international reading achievement tests of children, adolescents and young adults? A study based on Nordic results in PIRLS, PISA and PIAAC. *Assessment in Education: Principles, Policy and Practice*, 25(1), 107–126.
- Stewart, W. (2013). Is Pisa fundamentally flawed? Retrieved February 3, 2015 from <http://www.tes.co.uk/article.aspx?storycode=6344672>.
- Stromquist, N. P. (2016). Using regression analysis to predict countries' economic growth: Illusion and fact in education policy. *Real-World Economics Review*, 76, 65–74.
- Tröhler, D. (2013). The OECD and cold war culture: Thinking historically about PISA. In *PISA, power, and policy: The emergence of global educational governance* (pp. 141–161).
- Tucker, M. (Ed.). (2011). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Boston: Harvard Education Press.
- Tucker, M. (2014). *Chinese lessons: Shanghai's rise to the top of the PISA league tables*. Retrieved October 23, 2017 from Washington DC. <http://www.ncee.org/wp-content/uploads/2013/10/ChineseLessonsWeb.pdf>.
- Tucker, M. (2016). Asian countries take the U.S. to school. *The Atlantic*. Retrieved October 23, 2017 from <https://www.theatlantic.com/education/archive/2016/02/us-asia-education-differences/471564/>.
- Uljens, M. (2007). The hidden curriculum of PISA: The promotion of neo-liberal policy by educational assessment. In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA zufolge PISA—PISA According to PISA* (pp. 295–303). Berlin: Lit Verlag.
- Urban, M. (2017). We need meaningful, systemic evaluation, not a preschool PISA. *Global Education Review*, 4(2), 18–24.
- Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *Prospects*, 39(1), 33.
- Zhao, Y. (2009). *Catching up or leading the way: American education in the age of globalization*. Alexandria: ASCD.
- Zhao, Y. (2012). *World class learners: Educating creative and entrepreneurial students*. Thousand Oaks: Corwin.
- Zhao, Y. (2014). *Who's afraid of the big bad dragon: Why China has the best (and Worst) education system in the world*. San Francisco: Jossey-Bass.
- Zhao, Y. (2015). *Lessons that matter: What we should learn from Asian school systems*. Retrieved July 20, 2017 from Melbourne. <http://www.mitchellinstitute.org.au/reports/lessons-that-matter-what-should-we-learn-from-asias-school-systems/>.
- Zhao, Y. (2016a). From deficiency to strength: Shifting the mindset about education inequality. *Journal of Social Issues*, 72(4), 716–735.

- Zhao, Y. (2016b). Who's afraid of PISA: The fallacy of international assessments of system performance. In A. Harris & M. S. Jones (Eds.), *Leading futures* (pp. 7–21). Thousand Oaks: Sage.
- Zhao, Y. (2017a). Fatal attraction: Why the west must stop copying China's flawed education system. *New Internationalist*, 505, 24–25.
- Zhao, Y. (2017b). What works can hurt: Side effects in education. *Journal of Educational Change*, 18(1), 1–19.
- Zhao, Y. (2018a). The changing context of teaching and implications for teacher education. *Peabody Journal of Education*, 93, 1–14.
- Zhao, Y. (2018b). Shifting the education paradigm: Why international borrowing is no longer sufficient for improving education in China. *ECNU Review of Education*, 1(1), 76–106.
- Zhao, Y. (2018c). *What works may hurt: Side effects in education*. New York: Teachers College Press.
- Zhao, Y., & Gearin, B. (2016). Squeezed out. In D. Ambrose & R. J. Sternberg (Eds.), *Creative intelligence in the 21st century* (pp. 121–138). Berlin: Springer.
- Zhao, Y., & Wang, Y. (2018). Guarding the past or inventing the future: Education reforms in East Asia. In Y. Zhao & B. Gearin (Eds.), *Imagining the future of global education: Dreams and nightmares* (pp. 143–159). New York: Routledge.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.