

# Multi-Interpretation of Program for International Student Assessment (PISA)<sup>\*</sup>

Jiang Lijing

Qufu Normal University, Qufu, China;  
Columbia University, New York, USA

Ji Yingnan

Columbia University, New York, USA

With Program for International Student Assessment (PISA) is becoming the largest and most influential international comparative education test in the world, critiques involving its premise, statistical methods, samples, and test items have aroused at the same time. Multi-interpretation of PISA could be divided into three main parties. First are supporters of PISA, who insist PISA scores are indicators of the future economic growth of one country and provide broad prognostics on one country's educational system. Second are opponents of PISA that could be subdivided into two sub-parties: modest critics and radical critics. Third are neutrals of PISA who analyze pros and cons of PISA, suggest not over-believing PISA scores, and misusing PISA results.

*Keywords:* Program for International Student Assessment (PISA), test score, critiques

## Introduction

Since the Organization for Economic Cooperation and Development (OECD) first implemented Program for International Student Assessment (PISA), conducted every three years and with the result released the following year, surveying the competencies and skills of 15 years old in reading, math, and science. In 2000, PISA had mushroomed to 72 countries and non-national entities in 2015, and became the largest and most influential international comparative education test in the world.

However, with its growth of influence, critiques involving its premise, statistical methods, samples, and test items, etc. have aroused at the same time, which provide us with totally different views of PISA. In general, these interpretations could be divided into three main parties. First are supporters of PISA. They believe that PISA test score is predictor of a nation's future economic growth and provides broad prognostics on what is wrong or right in a nation's educational system. Second are opponents of PISA, who question the validity of PISA score and the relevance of policy-making based on PISA results. These opponents could be subdivided into two sub-parties. One are modest critics that point out the shortages and passive influence of PISA and argue to revise and improve it, the other are radical critics, who claim to pause PISA based on its enormous damage to the world's education. Third are neutrals of PISA. Being aware of the pros and cons of PISA, they

---

\* **Acknowledgement:** Project funded by China Postdoctoral Science Foundation (No. 2016M590648).

Jiang Lijing, Ph.D., associate professor, School of Education Science (SES), Qufu Normal University; visiting scholar, Teachers College, Columbia University.

Ji Yingnan, Ph.D., student, Teachers College, Columbia University.

suggest educators, politicians, and policy analysts not to misunderstand and misuse PISA results lest causing misleading in practice.

### **Supporters' Premise: PISA Scores are Indicators of Economic Growth Potential**

According to studies of Eric Hanushek, Paul Peterson, and Ludgar Woessmann (2013), average national mathematics test scores were the single best predictor of national economic growth in the period of 1960-2010.

In their representative publication *Endangering Prosperity* (Hanushek, Peterson, & Woessmann, 2013), they analyzed the relationship between previous international test results and current PISA test scores, then connected it with economic growth worldwide from 1960 to 2000 based on a vast body of persuasive statistical work, finally they concluded that the relationship between PISA-style test scores and gross domestic product (GDP) during the period of 1960-2010 was "causal" (p. 65).

Hanushek and his colleagues use this international "evidence" to claim that the United States (U.S.) is missing out on economic growth and that a revamping of the public education system is urgent for the country's economic health. They warn that American students' poor performance in PISA will soon weigh heavily on the economic future of U.S.. To avert this, they argue overturning structural reforms are now needed.

Though American students have sat in the middle of the range among developing countries, and even ranked about average for OECD countries since its debut in PISA 2000, it seems no one, except only a few scholars like Hanushek and his colleagues, concerns about PISA during 2000 to 2009. However, upon release of the 2009 results, which finding teenagers in Shanghai, China to be the best-educated in the world, it prompted officials elsewhere across the globe to question their own educational systems, and PISA results were also called "a wake-up call" by the U.S. education secretary (Guttenplan, 2010).

Arne Duncan (2010), the U.S. education secretary, said,

Every three years, PISA assesses the reading, mathematics, and scientific literacy of 15 years old students. It provides crucial information about how well our students are prepared to do the sorts of reading, mathematics, and science that will be demanded of them in postsecondary education or the job market, and as young adults in modern society. Unfortunately, the 2009 PISA results show that American students are poorly prepared to compete in today's knowledge economy.

He claimed, "In a highly-competitive knowledge-based economy, maintaining the educational status quo means American students are effectively losing ground." He pointed out, "Americans need to wake up to this educational reality."

Two years later, when 2012 PISA results were announced on "PISA Day" 2013, Duncan (2013) first admitted to positive reductions in high school dropouts and increasing numbers of students' attendance to college, but he still concluded that the big picture was straightforward and stark: It was a picture of educational stagnation. He recommended, "In a knowledge-based, global economy, where education is more important than ever before, both to individual success and collective prosperity, our students are basically losing ground. We are running in place, as other high-performing countries start to lap us" (Duncan, 2013).

Besides, in a speech to a college audience in North Carolina, President Obama (2010) recalled how the Soviet Union's 1957 launching of Sputnik stimulated U.S. to increase investment in math and science education, helping America win the space race. President Obama said, "Fifty years later, our generation's Sputnik moment is back." With billions of people in India and China "suddenly plugged into the world economy," he pointed out that nations with the most educated workers will prevail. As it stands right now, America is in danger of falling behind.

It is very clear when Arne Duncan (2012; 2013) and other researchers, such as Marc Tucker (2015) emphasized that American students performed far below their “competitors” in Shanghai, Singapore, and Korea, and even when the former President Obama (2010) claimed that “Our Generation’s Sputnik Moment is Now,” they were explicitly drawing on the Hanushek’s premise that American student’s PISA result is an accurate predictor of U.S. future economic growth (Carnoy, 2015, pp. 5-6).

### **Opponents’ Critiques: Modest and Radical**

#### **Modest Critics**

**Are PISA test scores indicators of future economic growth?** Some researchers like Diane Ravitch, are sharply critical of Hanushek-Peterson-Woessmann premise that the U.S. is doomed to a future of slower progress because of its low PISA mathematics scores by using the experiences of U.S., an exceptional case in Hanushek’s study (Hanushek, Peterson, & Woessmann, 2013).

Diane Ravitch (2013) pointed out directly, “U.S. students have never been top performers on the international tests. We are doing about the same now on PISA as we have done for the past half-century.”

Actually, there has been a long history for American students to take part in the international tests since they first participated the First International Mathematics Study (FIMS) in the mid-1960s. The FIMS result showed American students, a large sample (6,700) of 13 years old in 395 schools in U.S. with both numbers larger than PISA samples today, ranked next to the last (Sweden was lower) of the 12 countries that took the test (Medrich & Griffith, 1992, pp. 96-103).

Then, the Second International Mathematics Study (SIMS), held in 1987, showed that, American 13 years old students performed relatively well on algebra, arithmetic, and descriptive statistics items, but still near the bottom on geometry and measurement items while students in other countries like Japan, France, Netherlands, Hungary, and Belgium all performed better than American students (Medrich & Griffith, 1992, pp. 104-111).

Besides, Ravitch, according to the study of Keith Baker (2007, pp. 101-104) claimed that there was no relationship between a nation’s economic productivity and its test scores. Nor did the test scores bear any relationship to the quality of life or democratic institutions. And when it came to creativity, the U.S. clobbered the world with more patents per million people than any other nation (p. 2).

**The validity of PISA rankings and the relevance for policy-making.** A number of European academics criticize that PISA rankings are based on a “profound conceptual error,” which questions the relevance of international test results for policy-making (Stewart, 2013).

Stewart (2013) listed three main shortages of the statistical methods of PISA.

1. The structure of questions of PISA might fail to show differences in students’ subject knowledge because PISA questions have to be designed to estimate math, reading, and science knowledge across a wide range of cultures.

2. It is common that PISA imputes “plausible values,” which Rasch model is used to estimate, for each student’s score since not all of students answer the same questions on the test. However, Rasch model depends on items that have similar degrees of difficulty in different countries. Though PISA seeks to eliminate questions that are biased toward particular countries, some researchers say they could not find two items in PISA’s tests that function in exactly the same way in different countries, which invalidates the Rasch model (Stewart, 2013).

3. There are usually minor students taking the test when math, reading, or science is in its “off years” since each round of PISA has its own focus among these three subjects. Reading was key focus in 2000 and 2009

PISA, mathematics was in 2003 and 2012, and science was in 2006 and 2015. In the “off years,” one nation’s reading, math, or science scores were estimated on the basis of only 40% of the student sample actually taking the test, and not all those answering the same questions, which raises serious questions about the credibility and validity of PISA rankings (Stewart, 2013).

**The representativeness of PISA samples.** With Shanghai’s debut in PISA 2009, Shanghai achievement, turning out to be the top of the PISA rankings, has stunned experts all over the world by outscoring their counterparts in dozens of other countries, in reading, math, and science.

Due to these results and the publicity given them by the OECD’s media team, the performance of Shanghai students has become a benchmark for how well students worldwide should be able to achieve academically and Shanghai scores also become conflated with China’s national performance.

However, two researchers, such as Tom Loveless (2013; 2014) and James Harvey (2015) argued sharply that the Shanghai sample was not representative even of the Shanghai 15 years old students and certainly not of China, with almost 66% of rural students, since Shanghai schools excluded most of Shanghai’s large rural migrant students of 15 years old due to China’s Hukou policy. Besides, Loveless (2013; 2014) continued to question why the OECD agreed to publish the results from only one Chinese city—Shanghai, when 12 other Chinese provinces, with lower scores, had participated at the same time. Furthermore, Loveless (2013; 2014) pointed out that OECD’s endorsement of the Shanghai results was an implicit endorsement of China’s Hukou policy of forcing rural migrants’ teen age children to return to their provinces of origin to attend school. They insisted that these issues should have kept Shanghai scores out of any OECD comparison group and raised serious questions about the OECD’s brand as an international testing agency.

### **Radical Critiques: Claiming to Stop PISA**

In 2014, PISA’s 13th year, academics from around the world wrote an open letter (Andrews, Atkinson, & Ball, 2014), first published on May 2014 in the Guardian, to Dr. Schleicher, OECD’s director of the PISA, expressing deep concern about the impact of PISA test and calling for a halt to the next round of testing. In the letter, they pointed out negative consequences of PISA test and rankings.

1. While standardized testing has been used in many nations for decades (despite serious reservations about its validity and reliability), PISA has contributed to an escalation in such testing and a dramatically increased reliance on quantitative measures.
2. In education policy, PISA with its three-year assessment cycle, has caused a shift of attention to short-term fixes designed to help a country quickly climb the rankings, despite research showing that enduring changes in education practice take decades, not a few years, to come to fruition.
3. By stressing on a narrow scope of measurable aspects of education, PISA neglects the less measurable or immeasurable educational objectives, such as physical, moral, civic, and artistic development, thereby, dangerously narrowing our collective imagination regarding what education is and ought to be about.
4. While OECD emphasizes the economic role of public schools, the main role of public education is to prepare students for participation in democratic self-government, moral action, a life of personal development, growth, and wellbeing.
5. To carry out PISA and a host of follow-up services, OECD has embraced “public-private partnerships” and entered into alliances with multi-national for-profit companies, which stand to gain financially from any

deficits—real or perceived—unearthed by PISA. Some of these companies provide educational services to American schools and school districts on a massive, for-profit basis, while also pursuing plans to develop for-profit elementary education in Africa, where OECD is now planning to introduce the PISA programme.

6. Unlike United Nations organizations, such as United Nations Educational, Scientific, and Cultural Organization (UNESCO) or United Nations International Children's Emergency Fund (UNICEF), OECD has neither mandate to improve education nor mechanisms of effective democratic participation in its education decision-making process.

7. The new PISA Regime harms our children and impoverishes our classrooms, as it involves more and longer batteries of multiple-choice testing, more scripted “vendor”-made lessons, and less autonomy for teachers. From this point of view, PISA has deeply increased the current high pressure on schools, consequently endangering the wellbeing of students and teachers.

These researchers argue that these above developments are in overt conflict with widely accepted principles of good educational and democratic practice.

Although they also provide with some constructive ideas and suggestions that may help to alleviate the negative affects of PISA, they concluded that OECD’s narrow focus on standardised testing risks turning learning into drudgery and killing the joy of learning. Since they are deeply concerned that measuring a great diversity of educational traditions and cultures using a single, narrow, and biased yardstick could, in the end, do irreparable harm to our schools and our students, they call for slowing down the testing juggernaut.

### **Neutrals' Opinions: Pros and Cons of PISA**

#### **Taking PISA Scores With a Grain of Salt**

Dr. Henry M. Levin (2016), who had served as a consultant to PISA and also as Teachers College (TC)'s William H. Kilpatrick, professor of Economics and Education, wrote a paper “PISA 2015: A misleading test?” on the best countries pages of the *U.S. News & World Report Website*, published on November 30, 2016, only one week before the PISA Day, reminding the public of taking PISA result with a grain of salt.

On the one hand, Levin (2016) affirmed the positive function of PISA. He insisted, PISA represented a massive evaluation of schools and student achievement across the world, and the PISA test administered every three years and invited comparison of U.S. students and schools to those in other countries. Moreover, since PISA studies are careful and scientific, requiring representative samples of students, it provides “insights and detailed data” that can be helpful, and often be used by educators, politicians, and policy analysts to provide broad prognostics on what is wrong or right in the U.S. educational system and elsewhere.

On the other hand, Levin (2016) pointed out that there were dangers in believing that PISA scores could be raised by simply adopting the practices of the high-performing countries. This is because the statistical approach of PISA cannot provide causal links between school differences and student performance, only descriptive information on education. PISA is not a statistical study that isolates cause and effect.

Furthermore, Levin (2016) listed several shortages of test items of PISA, which make PISA fall short in diagnosing the reasons why performance varies from country to country:

1. It does not measure government assistance for poor families that accounts for student success. Most OECD countries provide much greater assistance than the U.S. for family housing, health care, family income, and universal early childhood education, all of which improve student achievement.

2. School data is collected only on a student's current school and does not measure the characteristics or results of previous schools attended. Almost all 15 years old in Korea and Japan are in the first year of the high school that they are attending when PISA data are obtained, so it is unlikely that the data on these schools can be used to account for their high performance.

3. Korea's exemplary results may not reflect schools as much as outside assistance. Parents spend almost as much sending their children to after-school programs as the government does on schools, a pattern found in Japan, China, and much of Asia.

4. PISA does not measure the soft-skills acquired in school, such as collaboration and persistence, which have been identified recently as contributing as much to employment and earnings as test scores, nor does it assess the skills that enable one to be effective as citizens—a major goal of schools.

The quality of an educational system can be based partially on its achievement scores, but comprehensive comparisons require a much deeper evaluation of causes and effects as well as performance on broader purposes of education, concludes by Levin (2016) (as cited in Jiang, 2017).

### **Turning Focus From International Perspective to National Perspective**

Recently, a number of American researchers suggest policymakers in the U.S. to turn their focus away from international comparisons and toward the U.S.'s own state-based systems to gain insights into school system improvement. Using international test results, such as PISA and the Trends in International Mathematics and Science Study (TIMSS), to conduct inward, across states comparisons. Their main reasons are as follows.

1. There are great differences of student performance among U.S. states. Some states, such as Massachusetts and Connecticut show high scores and great improvement in PISA and TIMSS while other states perform not as well as them.

2. The conditions and context of education are more similar in U.S. states than between the United States and other countries. Just as some scholars argue there are far greater contextual differences between countries, including the varying role that out of school tutoring and "cram" courses play in many countries (Bray, 2006; Ripley, 2013), as well as the national emphasis placed on international test performance (Carnoy, Khavenson, & Ivanova, 2014).

So, critics argue that from the standpoint of U.S. educational analysts and politicians, it would seem much more relevant, interesting, and helpful to use state-level test results over time to understand the policies high-gaining states implemented in the past 20 years than to examine other countries' educational policies. In other words, "learning from our own states rather than from other countries" is a better way to make use of PISA and TIMSS's results (Carnoy, 2015, p. 13).

### **Conclusion**

As long as the increasing influence of PISA in the world, lots of critiques have grown up at the same time, showing us with multiple versions of PISA. In general, multi-interpretation of PISA could be classified into three key parties: supporters, opponents, and neutrals. Although, up till now, those three have made no agree-on views on PISA involving with its premise, statistical methods, samples, test items, and the relevance of policy-making, etc., the debate has helped educators, politicians, and policy analysts recognize the advantages and disadvantages of PISA and hold a relatively more rational view of it, which consequently leads to less misuse of PISA results and less misleading in educational practice.

## References

- Andrews, P., Atkinson, L., Ball, S. (2014). OECD and PISA tests are damaging education worldwide-academics. *The Guardian*, 5(6).
- Baker, K. (2007). Are international tests worth anything? *Phi Delta Kappan*, 89(2), 101-104.
- Bray, M. (2006). Private supplementary tutoring: Comparative perspectives on patterns and implications. *Compare*, 36(4), 515-530.
- Carnoy, M., Khavenson, T., & Ivanova, A. (2014). Using TIMSS and PISA results to inform educational policy: A study of Russia and its neighbors. *Compare*, 45(2), 248-271.
- Carnoy, M. (2015). *International test score comparisons and educational policy: A review of the critiques*. Boulder, C.O.: National Education Policy Center. Retrieved May 1, 2017, from <http://nepc.colorado.edu/publication/international-test-scores>
- Duncan, A. (2010). *Secretary Arne Duncan's remarks at OECD's release of the PISA 2009 results*. Retrieved May 1, 2017, from <https://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-oecd-s-release-program-international-student-assessment>
- Duncan, A. (2012). *Statement by U.S. secretary of education Arne Duncan on the release of the 2011 TIMSS and Progress in International Reading Literacy Study (PIRLS) assessments*. Retrieved May 1, 2017, from <http://www.ed.gov/news/press-releases/statement-us-secretary-education-arne-duncan-release-2011-timss-and-pirls-assess>
- Duncan, A. (2013). *The threat of educational stagnation and complacency: Remarks of U.S. Secretary of Education Arne Duncan at the release of the 2012 Program for PISA*. Retrieved May 1, 2017, from <http://www.ed.gov/news/speeches/threat-educational-stagnation-and-complacency>
- Guttenplan, D. D. (2010). Western nations react to poor education results. *New York Times*, 12(8).
- Hanushek, E., Peterson, P., & Woessmann, L. (2013). *Endangering prosperity*. Washington, D.C.: Brookings Institution Press.
- Harvey, J. (2015). Ten things you need to know about international assessments. *Washington Post*, 2(3).
- Henry, M. L. (2016). PISA 2015: A misleading test? *U.S. News & World Report Website*, 11(30).
- Jiang, L. J. (2017). PISA test result: A grain of salt. *China Education Daily*, 1(19).
- Loveless, T. (2013). PISA's China problem. *Brookings Institution, Brown Center Chalkboard*, 10(9). Retrieved May 1, 2017, from <http://www.brookings.edu/research/papers/2013/10/09-pisa-china-problem-loveless>
- Loveless, T. (2014). Lessons from the PISA-Shanghai controversy. *Brookings Institution, Brown Center Chalkboard*, 3(18).
- Medrich, E. & Griffith, J. (1992). *International mathematics and science assessment: What have we learned?* Washington, D.C.: National Center for Educational Statistics.
- Obama, B. (2010). *President Obama in North Carolina: Our generation's sputnik moment is now*. Retrieved May 1, 2017, from <https://obamawhitehouse.archives.gov/blog/2010/12/06/president-obama-north-carolina-our-generation-s-sputnik-moment-now>
- Ravitch, D. (2013). My view of the PISA Scores. *Diane Ravitch's Blog*, p. 2. Retrieved May 1, 2017, from <https://dianeravitch.net/2013/12/03/my-view-of-the-pisa-scores/>
- Ripley, A. (2013). *The smartest kids in the world*. New York: Simon and Shuster.
- Stewart, W. (2013). Is PISA fundamentally flawed? *Times Education Supplement Magazine*, 6(26).
- Tucker, M. (2015). Are we just fooling ourselves? Is American education a colossal failure? *Education Week*, 4(16).