



Article

Trends in International PISA Scores over Time: Which Countries Are Actually Improving?

Kristie J. Rowley ^{1,*}, Shelby M. McNeill ² , Mikaela J. Dufur ³ , Chrisse Edmunds ⁴ and Jonathan A. Jarvis ³

¹ Independent Researcher, 429 North 750 West, Springville, UT 84663, USA

² Department of Leadership, Policy, and Organizations, Vanderbilt University, PMB #414 230 Appleton Place, Nashville, TN 37203, USA

³ Department of Sociology, Brigham Young University, 2008 JFSB, Provo, UT 84602, USA

⁴ Department of Sociology, Ohio State University, 238 Townshend Hall The Ohio State University 1885 Neil Avenue Mall, Columbus, OH 43210, USA

* Correspondence: kristie.j.rowley@gmail.com

Received: 30 June 2019; Accepted: 27 July 2019; Published: 2 August 2019



Abstract: Many countries attempt to increase their Program for International Student Assessment (PISA) rankings and scores over time. However, despite providing a more accurate assessment of the achievement-based improvements across countries, few studies have systematically examined growth in PISA scores over multiple assessments. Using data from the 2006, the 2009, and the 2012 PISA, we analyzed which countries experienced significant increases in their country-level average PISA scores between 2006 and 2012. To facilitate improved policy decisions, we also examined what country-level conditions were associated with such increases. Contrary to expectations, we found that few countries significantly increased their PISA scores over time. Countries that did experience meaningful improvements in PISA scores were more likely to have had lower PISA scores in 2006 and experienced country-level foundational advancements more recently, such as advancing to a more democratic form of government and/or a higher income classification.

Keywords: PISA; mathematics achievement; education policy; PISA trends over time; comparative education; institutional theory

1. Introduction

Considered by many to be education's "World Cup" (Coughlan 2013; Wilby 2013), the Program for International Student Assessment (PISA) has become a prominent international benchmarking assessment of national education systems. While the efficacy of using PISA results to inform educational policy and practice continues to be debated (Hanushek et al. 2013; Hess 2013; Wilby 2013), few doubt PISA's political significance. Countries that perform well on the assessment—which measures the academic performance of 15-year-old students in mathematics, reading, and science every three years [Organization for Economic Co-operation and Development (OECD) 2012]—are seen as having model education systems and are held in high prestige by governments, education scholars, and policymakers around the world (Department for Education 2013; Duncan 2013; Rutkowski and Rutkowski 2016). As such, the release of PISA scores and the associated international rankings routinely generates much attention in both media and education policy forums (Coughlan 2013; Hess 2013; Wilby 2013).

Given nations' and policymakers' emphases on PISA scores, PISA has stimulated change in many participating educational systems since its first cycle in 2000. For example, many countries that underperform relative to other high-performing countries or drop in their PISA rankings have attempted to improve their rankings and scores from one PISA assessment to the next (Dobbins and

Martens 2012; Department for Education 2013; Duncan 2013). However, despite countries' emphases on improving their rankings and scores over time, the question of growth in PISA scores over multiple assessments has rarely been systemically examined.

In order to better understand international patterns and trends in PISA scores, this study sought to examine meaningful improvements in PISA scores over time. Examining which countries significantly increase their PISA scores over time is a better indicator of educational improvement than PISA rankings or changes in rankings, as it provides a more accurate assessment of achievement-based advancements across countries. Using data from the 2006, the 2009, and the 2012 PISA, we analyzed which countries significantly increased their country-level average PISA scores between 2006 and 2012. We also examined what country-level conditions were associated with such increases in student achievement over time to facilitate improved policy decisions.

2. Trends in International PISA Scores over Time

Among the many international assessments, PISA has one of the largest influences on educational discourse, policy, and practice (Takayama 2008). Originally developed by the Organization for Economic Cooperation and Development (OECD), PISA assessments of mathematics, reading, and science are administered internationally every three years in both OECD member and participating non-member countries, with participation increasing from 35 countries in 2000 to 65 countries in 2012 (Organization for Economic Co-Operation and Development OECD). PISA assessments are intended to measure “how well 15-year-old students approaching the end of compulsory schooling are prepared to meet the challenges of today’s knowledge societies” (Organization for Economic Co-Operation and Development OECD, p. 22). However, the proper interpretation and use of PISA results is often debated by researchers, policymakers, and education correspondents around the world. For example, some skeptics question the use of PISA results as evidence of the quality of national education systems due to geographic, social, cultural, socioeconomic, ethnic, and religious differences between and within countries (Hess 2013; Meyer and Benavot 2013; Wilby 2013). In contrast, other scholars argue for a more nuanced approach to analyzing and utilizing PISA, while acknowledging the increased and potential policy influence of international assessments of education (Elliott et al. 2019; Rutkowski and Rutkowski 2016).

2.1. Benchmarking PISA Performance

While much debate surrounds the usefulness and the practical significance of PISA (Elliott et al. 2019; Hanushek et al. 2013; Hess 2013; Meyer and Benavot 2013; Rutkowski and Rutkowski 2016; Wilby 2013), few doubt its *political* significance. On one hand, countries that perform in the top handful of countries are seen as having model education systems and are held in high regard by governments, policymakers, and education scholars around the world (Department for Education 2013; Duncan 2013; Rutkowski and Rutkowski 2016). Prime examples of high-performing countries include Finland, which consistently ranked among the highest-performing countries for the first three PISA assessments between 2000 and 2006 (Hargreaves and Shirley 2009; Darling-Hammond and Lieberman 2012; Sahlberg 2011), Singapore, and administrative regions in China, all of which excelled on the 2009 and the 2012 PISA examinations (Organization for Economic Co-Operation and Development OECD; Sellar and Lingard 2013; Deng and Gopinathan 2016). On the other hand, countries that do not live up to the high educational performance of top-performing countries are heavily criticized, both within their own countries and by other nations, resulting in their education systems being labeled as “declining”, “lagging”, or “deteriorating” (Takayama 2008; Dobbins and Martens 2012; Hanushek et al. 2013). Patterns of “PISA shock” have been observed in many countries that either underperformed relative to other high-performing countries or dropped in their PISA rankings, including Denmark (Egelund 2008), France (Dobbins and Martens 2012), Germany (Ertl 2006), Israel (Feniger et al. 2012), Japan (Takayama 2008), Spain (Choi and Jerrim 2016), Turkey (Gür et al. 2012), the United Kingdom (Department for Education 2013), and the United States (Duncan

2013). As a result of “PISA shock”, many countries have been motivated to significantly improve their rankings on future PISA examinations (Dobbins and Martens 2012; Gür et al. 2012; Department for Education 2013; Duncan 2013). A desire to climb in the rankings has in turn led countries to prioritize improving their PISA scores over time.

2.2. Significantly Increasing PISA Scores over Time

Despite countries’ emphases on improving their scores and rankings over time, few studies have systematically examined meaningful growth in PISA scores over multiple assessments. This may be due to governments’ and policymakers’ preoccupation with PISA rankings, as highlighted by the attention given to high-performing countries (Coughlan 2013). We suggest that focusing solely on countries that are high-performing on the most recent PISA assessments does not capture improvements in PISA scores over time. PISA rankings and changes in these rankings compared to other countries also do not provide accurate or complete assessments of the achievement-based improvements in any given country. Rather, an improved examination of countries’ educational successes would include an examination of actual increases in their PISA scores *over time* and how common it is for countries to improve their scores over time. Therefore, we ask the following research question concerning improvements in PISA scores:

Q1: Which countries significantly improve their country-level average PISA scores over time?

2.3. Country-Level Conditions Associated with Significant Improvements

Examining which countries significantly improve their PISA scores over time is perhaps one of the most meaningful benchmarks for determining educational improvement; however, such an examination would be incomplete without an attempt to identify the conditions that are associated with these meaningful improvements. While few studies have examined growth in PISA scores over time or the conditions associated with such growth, many countries have tried to increase their scores by implementing micro-level educational policies or reforms used in high-performing countries, such as hiring more qualified teachers or implementing new curriculum (Dobbins and Martens 2012; Department for Education 2013; Duncan 2013). However, micro-level policy approaches often ignore the macro-level contexts that influence all students’ performance in a country. Because significantly increasing country-level average student performance is a considerable undertaking, we argue that policy efforts that facilitate large-scale improvements to macro-level contexts are likely motivating significant increases in test scores.

A prime model for understanding the macro-level conditions associated with educational development and improvement is institutional theory. Institutional theorists have hypothesized that significant improvements across educational, economic, social, and political institutional infrastructures are motivated by country-level foundational advancements, such as democratization and increased economic wealth (Meyer 1977; Ramirez and Boli 1987; Baker and LeTendre 2005). Educational research has consistently provided support for institutional theory by demonstrating that significant developments in national education systems are often predicated on country-level foundational advancements. For example, past studies have found that recently independent or democratized countries are more likely to experience rapid advancements in their education systems (Ramirez and Boli 1987; Meyer et al. 1992), suggesting that political foundational advancements are associated with educational improvement. Similarly, institutional theorists Ramirez and Boli (1987) used historical records to demonstrate that educational development is also commonly preceded by economic, political, and social foundational advancements, such as the expansion of market economies, international relations, and wide-spread notions that education leads to national progress and success.

While these studies demonstrate how national foundational advancements are associated with educational development and improvements in general, less is known regarding how national foundational advancements are related to increases in country-level student achievement. This is particularly true for large, international assessments such as PISA, where improvements in country-level

achievement outcomes are often anecdotally attributed to micro-level policy changes ([Department for Education 2013](#); [Duncan 2013](#)). Because governments and policymakers often view improvements in test scores as a prime indicator of educational development and improvement, it is important to understand whether national foundational advancements are also associated with significant increases in PISA scores. Therefore, we ask the following research question:

Q2: What national foundational advancements are associated with significant improvements in PISA scores over time?

2.4. Additional Tests of Country-Level Conditions

An additional test of whether national foundational advancements are legitimately driving improvements in PISA scores would be whether the same advancements are concurrently driving improvements in other social contexts. According to institutional theory, foundational advancements motivate improvements across not only countries' educational infrastructures but also their economic, their social, and their political institutional infrastructures ([Meyer 1977](#); [Ramirez and Boli 1987](#); [Baker and LeTendre 2005](#)). Thus, institutional theory implies that if national foundational advancements are motivating significant improvements in educational infrastructures, as shown by increases in test scores, significant improvements should also have taken place in other country-level institutional infrastructures. If such patterns are evident not only in educational outcomes but also other infrastructures related to other social contexts, such as health care or human development, we increase our confidence in the connection between national foundational advancements and educational improvements. To provide further evidence of the association between country-level conditions and significant increases in PISA scores over time, we test the relationship between national foundational advancements and other country-level institutional infrastructures. Therefore, we examine the following research question as a further check of country-level conditions:

Q3: Do other country-level institutional infrastructures significantly improve during the same time period that PISA scores significantly improve?

3. Materials and Methods

3.1. Data and Analytic Sample

To inform this study, we created an international database containing over 200 country-level indicators of educational, economic, and sociopolitical contexts.¹ The database includes information for over 70 countries and administrative regions from the mid-1900s to 2012. Because our study focused on country educational achievement, we used aggregate country-level PISA scores ([Organization for Economic Co-Operation and Development OECD](#)).² We specifically focused on mathematics scores for our analysis, because math is the subject most likely to be taught uniformly and sequentially across international contexts ([Akiba et al. 2007](#)) and is least likely to be influenced by non-school factors ([Lee and Bryk 1989](#)). To allow for appropriate comparisons across nations and over time, we limited our analyses to the 55 countries that participated in the 2006, the 2009, and the 2012 PISA mathematics

¹ These indicators originated from the Luxembourg Income Study (LIS), OECD, PISA, World Bank, and other online and print resources.

² PISA creates a representative, generalizable sample for each nation by systematically targeting schools with 15-year-old students. While all students sampled were age 15, they represented grade levels ranging from 7th to 12th grade, with the vast majority of students enrolled in the 10th grade. A minimum of 150 schools were selected within each country, and approximately 35 students were sampled from each school ([National Center for Education Statistics 2015](#)).

examinations.³⁴ We examined scores between 2006 and 2012, because these years include the largest group of countries with three consecutive PISA scores.

3.2. Measures

To address the research questions outlined above, we used measures of country mean achievement scores to determine meaningful cut points for assessing significant growth in achievement over time. We also created more than 200 additional measures of mathematics achievement, national foundational advancements, and country-level institutional infrastructures. We report below on the foundational advancements and institutional infrastructures that were associated with significant increases in PISA scores over time. Each of these measures is discussed in more detail below. Means and standard deviations for each of our measures are listed in Table 1.

³ The PISA 2006, 2009, and 2012 assessments use the same mathematics performance scale, which means that points on this scale are directly comparable over time. For additional information regarding the comparability of PISA scores over time, see ([Organization for Economic Co-operation and Development OECD](#), Annex A5): Technical Details of Trend Analysis.

⁴ Because PISA is a widely followed international assessment, student sampling is generally intended to produce educational outcomes that are representative of each participating country. However, it is necessary to note that China does not participate in PISA as an entire country. Therefore, our reporting for China includes two special administrative regions (Hong Kong and Macao) that are reported separately and are not intended to be representative of China as a country. We excluded the results from Shanghai, the only province in mainland China that volunteered to fully participate in PISA, because its extraordinarily unusual social and educational characteristics (see [Tucker 2013](#)) make it difficult to compare it to other high-income countries.

Table 1. Description of variables.

<i>Variables ^a</i>	2006		2009		2012	
	Mean	SD	Mean	SD	Mean	SD
PISA Mathematics Achievement Scores						
<i>Country Mean Achievement Scores</i>	472	56.33	474	51.37	476	50.43
Mean achievement scores for all students nationwide						
<i>Mean Achievement Scores for High-SES Students</i>	532	52.93	535	49.79	536	49.21
Mean achievement scores for students in the top socioeconomic (SES) decile (the 90th to 100th percentile) nationwide						
<i>Mean Achievement Scores for Low-SES Students</i>	416	54.86	419	48.85	422	47.30
Mean achievement scores for students in the bottom SES decile (the 0 to 10th percentile) nationwide						
<i>Mean SES-Based Achievement Gaps</i>	116	27.12	116	27.32	114	26.94
Difference between mean achievement scores for high-SES students and mean achievement scores for low-SES students						
Country-Level Institutional Infrastructures						
Educational Infrastructures						
<i>Educational Spending as a Percentage of GDP</i>	4.69%	1.29%	5.10%	1.34%	5.29%	1.19%
Total government expenditure on education as a percentage of gross domestic product (GDP)						
<i>Educational Spending in US Dollars</i>	410	1080	449	1080	544	1270
Total government expenditure on education in units of 100 million US dollars						
<i>Secondary School Enrollment</i>	94.68%	7.94%	95.25%	6.15%	96.28%	4.81%
Total enrollment in secondary education as a percentage of the population of official secondary education age						
Economic Infrastructures						
<i>Income Inequality</i>	33.96	7.96	34.00	7.64	33.87	7.43
Difference between a nation's actual income distribution and an equal income distribution, as measured by the Gini coefficient. This measure ranges from 1 to 100, with 1 indicating that all citizens have equal wealth in a nation and 100 indicating that one person has all the wealth in a nation						
<i>Gross Domestic Product (GDP)</i>	7910	20,100	8760	21000	10,400	23,900
Sum of goods and services generated within a nation in units of 100 million US dollars						
<i>Gross Domestic Product (GDP) per Capita</i>	27,200.85	23,782.46	29,866.56	25,446.40	33,969.3	28,626.63
Gross domestic product (GDP) divided by national population						
Sociopolitical Infrastructures						
<i>Infant Mortality</i>	8.28	6.48	7.27	5.64	6.44	4.96
Number of deaths of infants under one year old per 1000 live births						
<i>Human Development ^b</i>	0.82	0.07	0.83	0.07	0.84	0.06
Composite statistic of life expectancy, education, and per capita income indicators in a nation. This measure ranges from 0 to 1, with larger values indicating higher human development.						
Sample Size (N) ^c	55		55		55	

Notes: ^a Included in our analyses are indicators of economic and governmental foundational advancements. Because the measurement of these indicators does not align with the measurement of the indicators listed here, they are not included in the table; ^b Statistics were calculated in 2005, 2008, and 2012. ^c Some statistics were computed using fewer countries due to missing data. PISA = Program for International Student Assessment.

3.3. Country Mean Achievement Scores and Significant Increases in Scores

To determine country mean achievement, we calculated each country's mean score on the PISA mathematics examination in 2006, 2009, and 2012 by using STATA's PISA-specific plausible values aggregation command (Macdonald 2008). To identify significant increases in PISA scores over time, we calculated the standard deviation of the achievement average of all 55 countries included in our sample in 2012. We used one-third of a standard deviation, or 17 points on the PISA test, to identify significant increases in country mean achievement over time. Thus, we considered countries that improved their scores by 17 points or more between 2006 and 2012 to have significantly increased their achievement scores (hereafter "SI countries"), whereas countries that improved their scores by fewer than 17 points between 2006 and 2012 or decreased their scores were considered to have *not* significantly increased their achievement scores (hereafter "NSI countries").

We used one-third of a standard deviation to identify significant increases in PISA scores for several reasons: first, it is larger than every country's standard error in our sample, which demonstrates that we were measuring real and significant change; second, one-third of a standard deviation has historically represented the upper range of effect sizes necessary for interventions to be considered "educationally significant" (Tallmadge 1977; Bloom et al. 2008); third, meta analyses have demonstrated that one-third of a standard deviation is on par with the average effect size associated with randomized studies of educational interventions (Hill et al. 2008); and lastly, by more contemporary educational standards, it is generally considered a sizable achievement increase between groups of students (Lipsey et al. 2012). A 17-point increase in PISA scores is also roughly equivalent to an increase of one-half year of schooling, as supported by OECD documentation indicating that 39 points on the PISA mathematics assessment is roughly equivalent to one year of schooling (Organization for Economic Co-Operation and Development OECD).

Some researchers have made the case that lower thresholds could be used to evaluate both the educational effectiveness of interventions (Lipsey et al. 2012) and significant increases in achievement on the PISA assessments (Carnoy and Rothstein 2013). For example, Carnoy and Rothstein (2013) used an 8-point cut off to define "better" performance on the PISA examination. However, similar to our 17-point threshold, they defined "substantially better" performance using an 18-point cutoff (Carnoy and Rothstein 2013, p. 6). Therefore, while we agree that our one-third standard deviation may be on the high end of what might be considered "educationally significant", we suggest that it is a responsible and appropriate benchmark for determining significant increases in country mean achievement scores. If we were going to claim that 15-year-old students significantly improved their PISA scores over a short time span, we wanted such findings to be based on meaningful and sufficiently large increases in PISA scores to ensure that these increases were not attributed to artifacts of sampling or margins of error.

3.4. Mean Achievement Scores for High- and Low-SES Students and SES-Based Achievement Gaps

To further understand differences in PISA achievement between countries that significantly increase their scores and those that do not, our analyses also included measures of mean achievement scores for high-socioeconomic status (SES) students, mean achievement scores for low-SES students, and mean SES-based achievement gaps as defined by differences in scores between high- and low-SES students. To create our measures of mean achievement scores for high- and low-SES students, we first defined high- and low-SES students using PISA's index of economic, social, and cultural status.⁵ Based on this SES measure, we used decile dispersion to measure socioeconomic inequality because

⁵ PISA's index of economic, social, and cultural status is calculated from information regarding parental income, education status, and home possessions, including family wealth, cultural possessions, and home educational resources. Parental income, education status, and home possessions are all equally weighted in this measure (Organization for Economic Co-Operation and Development OECD). The measure is then standardized to have an average of 0 and standard deviation of 1. Thus, the average score for each country varies. However, principal component analysis derives that each of the factors

it compares the share of the highest-SES students within a nation to the lowest-SES students ([World Bank 2015](#)). Thus, within each country, students in the bottom SES decile (the 0 to 10th percentile) were considered low-SES students, and students in the top SES decile (the 90th to 100th percentile) were considered high-SES students. We then averaged mathematics achievement scores for high- and low-SES students in each country. Our measure of the SES-based achievement gaps was calculated using the differences between the aggregated mathematics scores of high- and low-SES students within each country. Countries with a larger difference in scores between high- and low-SES students had a larger mean SES-based achievement gap, whereas countries with a smaller difference between high- and low-SES students had a smaller mean SES-based achievement gap. Similar to standards for country-level average PISA scores, we defined significant improvement in mean achievement scores for high-SES students, mean achievement scores for low-SES students, and mean SES-based achievement gaps using a 17-point threshold.

3.5. National Foundational Advancements

Based upon the findings of institutional theorists that large-scale governmental and economic advancements in a country are associated with educational development and improvement ([Ramirez and Boli 1987](#); [Meyer et al. 1992](#)), our measures of national foundational advancements included transitions to more democratic forms of government and higher country income classifications. To examine transitions to more democratic forms of government, we first compiled data on the year each country in our sample transitioned to a more democratic form of government. We then used Huntington's waves of democratization ([Huntington 2012](#)) to categorize each country's date of transition into three "waves" or time periods. Based on Huntington's categorizations, the first wave ranged in date from 1828–1926, the second wave from 1943–1962, and the third wave from 1974–present. For parsimony, we combined the first and the second wave (hereafter referred to as the pre-third wave) and ended the third wave in 2012 due to it being the last year of our PISA data.

Our measure of advancing to a higher country income classification was created using the World Bank's country income classifications, which document each country as low-income, lower-middle income, upper-middle income, or high income between the years 1987 and 2012 based on their gross national income (GNI). We categorized countries into three advancement types: advanced to a high-income classification pre-1987, advanced to an upper middle income country classification post-2000, or advanced to a high income country classification post-2000. We used the World Bank's joint-issuance of the Millennium Development Goals in the year 2000 ([United Nations 2000](#)) as a benchmark for more recent transitions and the year 1987 as a benchmark for earlier transitions due to it being the first year country income classifications were calculated by the World Bank.

3.6. Country-Level Institutional Infrastructures

Lastly, measures of country-level institutional infrastructures were primarily gathered from the OECD and the World Bank and were categorized into three types of infrastructures: educational, economic, and sociopolitical. Corresponding to our measures of country mean achievement scores, we primarily used measures of infrastructures from 2006, 2009, and 2012 and defined significant improvements in infrastructures using a one-third standard deviation cut-off.

Our measures of educational infrastructures included educational spending as a percentage of gross domestic product (GDP), total educational spending in US dollars, and secondary school enrollment. Educational spending as a percentage of GDP was calculated by expressing the total government expenditure on education as a percentage of GDP. Education spending in US dollars represented total government expenditure on education in units of 100 million US dollars. We computed

included in this measure similarly load on to the measure across nations ([Organization for Economic Co-Operation and Development OECD](#)).

secondary school enrollment as the total enrollment in secondary education as a percentage of the population of official secondary education age.

Our measures of economic infrastructures included income inequality, GDP, and GDP per capita. Income inequality was measured by the Gini index. As a measure of the distribution of household income within a nation, the Gini index estimates how a nation's actual income distribution varies from an equal income distribution. This measure ranges from 1 to 100, with 1 indicating that all citizens have equal wealth in a nation and 100 indicating that one person has all the wealth in a nation. GDP represents the sum of goods and services generated within a nation over the course of a year in units of 100 million US dollars. GDP per capita represents gross domestic product divided by national population and is a common measure of living standards within a country.

Our sociopolitical infrastructure measures included infant mortality and human development, measures often seen as indicators of health and well-being in a country. The infant mortality rate was computed as the number of deaths of infants under one year old per 1000 live births. Human development was measured as a composite statistic of life expectancy, education, and per capita income indicators. This measure ranged from 0 to 1, with larger values indicating higher human development.

3.7. Analyses

To better understand which countries significantly improved their country-level average PISA scores over time, we analyzed each country's mean achievement scores in 2006, 2009, and 2012 and then compared the scores over time to identify countries in which performance grew by a minimum of one-third of a standard deviation over the six years included in our study. To further understand differences in PISA achievement between countries that significantly increased their scores and those that did not, we compared across NSI and SI countries measures of country mean achievement scores, mean achievement scores for high-SES students, mean achievement scores for low-SES students, and mean SES-based achievement gaps in 2006, 2009, and 2012. These comparisons offered descriptive information about how PISA scores changed over time and how student achievement differed across NSI and SI countries.

To examine what national foundational advancements were related to significant improvements in PISA scores, we compared foundational advancements to more democratic forms of government and higher income classifications for SI countries. In addition, we also averaged multiple indicators of national educational, economic, and sociopolitical infrastructures across SI countries in 2006, 2009, and 2012 to test whether other country-level institutional infrastructures significantly increased during the same time period as PISA scores. Such comparisons allow governments and policymakers to better understand the underlying country-level conditions that are associated with the significant increases in PISA scores experienced by SI countries.

4. Results

4.1. Significant Increases in PISA Mathematics Achievement Scores

To address our first research question and assess which countries significantly improved their country-level average PISA scores over time, we first rank ordered countries that participated in the 2006, the 2009, and the 2012 PISA from highest to lowest average increase in scores between 2006 and 2012. We then categorized countries according to whether or not they significantly increased their scores during this time period (see Table 2). Based on our criteria, 45 of the 55 countries in our sample did not significantly increase their scores between 2006 and 2012 and were categorized as NSI countries. However, ten countries did significantly increase their scores between 2006 and 2012 and were categorized as SI countries. Ordered by largest to smallest significant increase in scores, these countries were Qatar, Romania, Bulgaria, Israel, Turkey, Poland, Italy, Tunisia, Portugal, and Brazil. The country with the largest increase in mean achievement was Qatar, with an increase in average score of 58 points between 2006 and 2012. Based on OECD documentation, 39 points on

the PISA mathematics assessment is roughly equivalent to one year of schooling ([Organization for Economic Co-Operation and Development OECD](#)). Thus, an average 15-year-old student in Qatar in 2012 performed about one and a half years (or grades) above an average 15-year-old student in Qatar in 2006. The nine other SI countries in our sample also significantly increased their PISA scores by at least half a school year.

Table 2. PISA mathematics achievement scores and differences by country, 2006–2012.

Country	Country Mean Achievement Scores				Type of Difference
	2006	2009	2012	Difference 2006–12	
Qatar	318	368	376	58	Countries that increased 17 points or more (Significant Increasers)
Romania	415	427	445	30	
Bulgaria	413	428	439	26	
Israel	442	447	466	24	
Turkey	424	445	448	24	
Poland	495	495	518	23	
Italy	462	483	485	23	
Tunisia	365	371	388	23	
Portugal	466	487	487	21	
Brazil	370	386	389	19	
Hong Kong (China)	547	555	561	14	Countries that did not increase 17 points or more (Not Significant Increasers)
Serbia	435	442	449	14	
Macao (China)	525	525	538	13	
Japan	523	529	536	13	
Chile	411	421	423	12	
Chinese Taipei	549	543	560	11	
Montenegro	399	403	410	11	
Germany	504	513	514	10	
Lichtenstein	525	536	535	10	
Thailand	417	419	427	10	
Argentina	381	388	388	7	
Korea	547	546	554	7	
Mexico	406	419	413	7	
United States	474	487	481	7	
Colombia	370	381	376	6	
Estonia	515	512	521	6	
Russia	476	468	482	6	
Latvia	486	482	491	5	
AVERAGE	472	474	476	4	
Spain	480	483	484	4	
Croatia	467	460	471	4	
Jordan	384	387	386	2	
Austria	505	496	506	1	
Switzerland	530	534	531	1	
Ireland	501	487	501	0	
Luxembourg	490	489	490	0	
France	496	497	495	−1	
Norway	490	498	489	−1	
United Kingdom	495	492	494	−1	
Slovenia	504	501	501	−3	
Belgium	520	515	515	−5	
Greece	459	466	453	−6	
Lithuania	486	477	479	−7	
Netherlands	531	526	523	−8	
Canada	527	527	518	−9	
Slovakia	492	497	482	−10	
Czech Republic	510	493	499	−11	
Denmark	513	503	500	−13	
Iceland	506	507	493	−13	
Hungary	491	490	477	−14	
Australia	520	514	504	−16	
Indonesia	391	371	375	−16	
Uruguay	427	427	409	−18	
New Zealand	522	519	500	−22	
Sweden	502	494	478	−24	
Finland	548	541	519	−29	

Finding that ten countries significantly increased their scores between 2006 and 2012 confirms that countries can meaningfully improve their scores over time. Considering all the time, the policy efforts, and the resources that go into improving students' educational development around the world, this is good news for countries that are seeking to improve their test scores. However, because fewer than 20 percent of countries in our sample were SI countries, our findings also suggest that it is rare for countries to experience meaningful growth in scores over time. Furthermore, considering that our overall sample of countries only experienced an average increase of four points in country mean achievement between 2006 and 2012, we question the reasonableness of expectations of significant growth over short time periods. Raising country-level average student performance on the PISA examination by a minimum of half a school year within a six-year time period is a significant undertaking and difficult to accomplish. Therefore, we conclude that countries can and do significantly increase their PISA scores over time; however, experiencing such increases in scores is not likely and is not an easily-achieved expectation. As such, increased focus should be placed on the countries that significantly increase their scores over time and on the conditions that are related to such improvements, regardless of whether the countries are high-performing or highly-ranked.

4.2. Comparing PISA Mathematics Achievement Scores across NSI and SI Countries

In order to further understand how PISA achievement differed between countries that significantly increased their scores and those that did not, we compared additional measures of mean PISA mathematics achievement scores between NSI and SI countries. In addition, many policy approaches to improving country-level performance focus on pulling up the performance of the poorest-performing students in a country, in turn diminishing achievement gaps between the highest- and the lowest-performing students. To assess whether SI countries were accomplishing their growth in this fashion, we averaged measures of country mean achievement scores, mean achievement scores for high-SES students, mean achievement scores for low-SES students, and mean SES-based achievement gaps in 2006, 2009, and 2012 across NSI and SI countries (see Table 3). In Table 3, we found that SI countries were more likely to have significantly lower country mean achievement scores in 2006, 2009, and 2012 compared to NSI countries. Similarly, we saw that SI countries were also more likely to have significantly lower mean achievement scores for both high- and low-SES students across these three time points. However, SI countries significantly increased both their high- and their low-SES students' mean achievement scores at similar rates between 2006 and 2012, whereas NSI countries' scores remained relatively stagnant for both groups. We also found that SES-based achievement gaps did not greatly differ in size between NSI and SI countries over time and that neither group of countries significantly reduced their achievement gaps.

Table 3. Mean PISA mathematics achievement scores of NSI and SI countries, 2006–2012.

Mathematics Achievement Scores	NSI Countries				SI Countries			
	2006	2009	2012	Difference, 2006–2012	2006	2009	2012	Difference, 2006–2012
Country Mean Achievement Scores	484	484	483	−1	417	434	444	27
Mean Achievement Scores for High-SES Students	542	542	542	0	487	500	511	24
Mean Achievement Scores for Low-SES Students	428	428	429	1	363	379	390	27
Mean SES-Based Achievement Gaps	114	114	113	−1	124	121	121	−3
Sample Size (N)				45				10

Notes: bold print indicates significant increase. NSI = not significant increaser. SI = significant increaser.

Three important findings emerged from these analyses. First, SI countries had significantly lower PISA scores on average than NSI countries across all three time points in our study. This trend suggests that countries with lower test scores were more likely to significantly improve their test scores over

time. This finding is encouraging for low-performing countries that desire to increase their PISA scores; however, it also undermines the expectations for growth for higher-performing countries. Countries that are already performing well may experience ceiling effects that cap their potential growth in PISA scores over time. Second, SI countries were more likely to have significantly increased both their high- and their low-SES students' mean achievement scores than NSI countries. This finding suggests that improving the test scores of both high- and low-performing students is associated with increased country-level average PISA scores. As a result, meaningful growth in PISA scores over time is likely linked to policy efforts that are influencing all students' performance in a country.

Lastly, we found that SI and NSI countries were both likely to record large mean SES-based achievement gaps that remained fairly stable over time. Because high- and low-SES students' mean performance increased at similar rates in SI countries and remained stagnant in NSI countries, our finding that neither group of countries significantly decreased their mean achievement gap over time was not initially surprising. However, this finding is interesting in that it does not support the common policy expectation that significant increases in test scores are associated with reducing achievement gaps between high- and low-SES students. One possible reason for why SI countries experienced significant increases in PISA scores without decreasing their SES-based achievement gaps could be that both high- and low-SES students still had ample room for improvement. In other words, decreasing SES-based achievement gaps may only be associated with meaningful improvements in scores after countries have achieved higher levels of academic performance overall, especially when those countries are low-performing to begin with. In relatively high performing countries, the highest-performing students may be maximizing their achievement and are therefore less able to experience growth in PISA test scores over time. Subsequently, these countries may be better able to improve their overall PISA performance by focusing policy efforts on poor-performing, typically disadvantaged students. However, countries with average or relatively low average PISA scores might be better served by focusing on more fundamental policy issues that uniformly improve the educational outcomes for all of their students.

4.3. National Foundational Advancements across SI Countries

Having determined which countries significantly increased their PISA scores over time, we focused the remainder of our analyses on exploring why SI countries may have significantly increased their scores over time. To do so, we examined our second research question concerning the country-level conditions that were related to significant improvements in PISA scores through scrutinizing foundational advancements. We report below on the two foundational advancements that were associated with significant increases in PISA scores over time, namely, advancing to more democratic forms of government and higher income classifications. Considering that a typical country takes about 55 years to advance from a lower middle income to an upper middle income classification and about 15 years to advance from an upper middle to a high income classification (Felipe et al. 2017), moving to a higher income classification represents substantial economic advancement in a country. We grouped countries according to the specific time period in which they experienced foundational advancements. Such time periods were based on Huntington's waves of democratization (2012) and other notable dates in history (United Nations 2000), with advancements to a democratic form of government during the Third Wave (1974–2012) and advancements to a higher income classification post-2000 representing the most recent advancements.

From these comparisons, we found that all ten SI countries advanced to a more democratic form of government and/or a higher income classification (see Table 4). Interestingly, the majority of SI countries experienced one or both of these national foundational advancements during the most recent time periods. One SI country (Tunisia) transitioned to a higher income classification post-2000, and six SI countries (Brazil, Bulgaria, Poland, Portugal, Romania, and Turkey) transitioned to both a higher income classification post-2000 and a more democratic form of government during the Third Wave. The three remaining SI countries (Israel, Italy, and Qatar) experienced one or more foundational

advancements during earlier time periods.⁶ It is interesting to note that Qatar is the only SI country in our sample that has not transitioned to a democratic form of government. However, while Qatar still maintains an authoritarian government, they have taken significant steps towards democratization during the Third Wave, including expanding freedom of expression in the media and increasing opportunities for political participation by women (Rathmell and Schulze 2000). Such steps towards an open and participatory government may have contributed to their significant increase in PISA scores.

Table 4. Advancements to a democratic form of government and higher income classification.

	Non-Democratic	Advanced to a Democratic Form of Government, Pre-Third Wave ^a	Advanced to a Democratic Form of Government, Third Wave ^b
Advanced to a high-income classification, pre-1987 ^c	Qatar	Israel, Italy	—
Advanced to an upper-middle income classification, post-2000 ^d	Tunisia ^e	—	Brazil, Bulgaria, Poland ^f , Romania, Turkey
Advanced to a high-income classification, post-2000 ^d	—	—	Poland ^f , Portugal ^g

^a Based on Huntington's waves of democratization (1991), the pre-Third Wave ranges in date from 1828–1962; ^b Based on Huntington's waves of democratization (1991), the Third Wave ranges in date from 1974–2012; ^c 1987 was the first year country income classifications were calculated by the World Bank; ^d Due to country income classifications originally being calculated by the World Bank we used their joint-issuance of the Millennium Development Goals in 2000 (United Nations 2000) as a benchmark for more recent transitions; ^e Tunisia advanced to a democratic form of government after the 2012 PISA examination; ^f Poland advanced to an upper middle income classification in 1996 and a high income classification in 2009; ^g Portugal advanced to a high income classification in 1994.

Finding that all of the SI countries in our sample advanced to a democratic form of government and/or a higher income classification suggests that national foundational advancements are associated with significant increases in PISA scores over time. These findings support the hypotheses of institutional theorists that significant educational development and improvement are motivated by country-level foundational advancements in a country (Meyer 1977; Ramirez and Boli 1987; Baker and LeTendre 2005). While we cannot provide causal explanations regarding significant increases in test scores, the time ordering of national foundational advancements preceding significant increases in test scores suggests there is an association. Further, our findings suggest that the time period in which these foundational advancements occur is also meaningful. The majority of SI countries advanced to a democratic form of government and/or a higher income classification during the most recent time periods, suggesting that experiencing foundational advancements in more recent time periods is associated with more recent increases in test scores. Therefore, we expect that countries that concentrate on such foundational advancements will be more likely to experience significant improvements in their PISA scores in the future. Likewise, we would expect that most countries that advanced to a more democratic form of government and/or a higher income classification at earlier time periods also experienced subsequent increases in educational performance prior to the time frame of this study (and the time frame covered by the PISA data). Future studies should examine these relationships further.

4.4. Country-Level Institutional Infrastructures across SI Countries

Given our finding that national foundational advancements are associated with significant increases in country-level average PISA scores, it is likely that macro-level foundational advancements may be motivating improvements in other country-level institutional infrastructures during the same

⁶ We also examined foundational advancements to more democratic forms of government and higher income classifications for NSI countries. Compared to SI countries, fewer of the NSI countries experienced one or more foundational advancements during the most recent time periods.

time periods. As an additional check of the conditions associated with significant increases in test scores, we examined what country-level institutional infrastructures significantly increased during the same time frame as PISA scores. To examine our third research question, we averaged multiple indicators of national educational, economic, and sociopolitical infrastructures across SI countries in 2006, 2009, and 2012 (see Table 5). We found that SI countries significantly improved their infant mortality rate and human development on average between 2006 and 2012, both of which are classified as sociopolitical institutional infrastructures. Our five other indicators of educational and economic institutional infrastructures did not significantly increase between 2006 and 2012.

Table 5. Country-level institutional infrastructures of SI countries, 2006–2012 ^a.

Institutional Infrastructures	2006	2009	2012	Difference 2006–2012
Educational				
Educational Spending as a Percentage of GDP	4.49%	4.74%	4.86%	0.37%
Educational Spending in US Dollars ^b	208	274	310	102
Secondary School Enrollment	93.75%	93.69%	94.55%	0.80%
Economic				
Income Inequality	37.29	36.92	36.35	−0.94
GDP ^b	4520	5670	6560	2040
GDP per Capita	17,382.7	19,494.7	23,357.7	5975
Sociopolitical				
Infant Mortality	11.81	10.01	8.65	−3.16
Human Development ^c	0.78	0.79	0.80	0.02
Sample Size (N) ^d				10

Notes: bold print indicates significant increase; ^a Mean statistics for SI countries were also computed from the mid-1970s to 2012. Similar results were found for most statistics and thus not reported; ^b Measured in units of 100 million USD; ^c Statistics were calculated in 2005, 2008, and 2012; ^d Some mean statistics were computed using fewer countries due to missing data.

The fact that SI countries experienced significant improvements in two sociopolitical infrastructures on average between 2006 and 2012 suggests that other country-level institutional infrastructures significantly improved during the same time period as PISA scores. We argue that it is not a coincidence that all ten of the SI countries in our study experienced macro-level foundational advancements. While most countries that participate in PISA examinations implement various micro-level policies to improve their scores, the countries that actually experienced significant growth over time started low and experienced macro-level improvements in health and development for children at the same time they experienced educational improvements. This increases our confidence that national foundational advancements are related to improvements in PISA scores in SI countries. Therefore, we conclude that macro-level conditions are associated with growth in PISA scores over time.

5. Discussion/Conclusions

Our study contributes to the comparative international education and educational policy literature by examining which countries experienced significant increases in their average PISA scores over time and what country-level conditions were associated with such increases. We found that *few* countries significantly increased their PISA scores over time. While this finding is in contrast to most countries' expectations of growth, it is not nearly as surprising when one considers that PISA does not examine growth in individual students' academic performance longitudinally but rather tests the academic performance of different samples of 15-year-old students within countries every three years. Therefore, raising country-level average student performance on the PISA examination by half of a school year within a six-year time period is a considerable undertaking, and countries that do

meaningfully increase their test scores over time have likely experienced significant and noteworthy improvements in not just their educational systems but also their social systems.

Our study also confirms that national foundational advancements are associated with meaningful improvements in PISA scores. All of the SI countries in our sample experienced the foundational advancements of transitioning to a more democratic form of government and/or a higher income classification. Because of the time ordering, and because of the magnitude of these economic and political shifts, these foundational advancements are likely related to the significant increases in PISA scores found in our study. This relationship between national foundational advancements and improvements in PISA scores is further supported by our finding that other country-level infrastructures also improved during the same time period. Therefore, we conclude that experiencing significant improvements in country-level test scores is associated with equally significant advancements in a country's governmental and economic foundations.

Because our study was primarily concerned with significant improvements in PISA scores over time, we did not focus on individual countries that did not significantly increase their PISA scores in our analyses. However, it is worth briefly noting that a few countries in our sample (specifically Uruguay, Sweden, New Zealand, and Finland) did experience significant decreases in their PISA scores over the six-year time span of our study. Although we suspect that such decreases were motivated by entirely different country-level conditions than those motivating significant improvements in PISA scores, explaining what conditions were associated with decreases in scores was beyond the scope of this paper. Nevertheless, these patterns of significant decrease do provide additional support for the importance of examining trends in PISA scores over time when seeking to assess achievement-based improvements. By comparing differences in PISA scores from only one assessment to the next, many countries and educational scholars mistakenly equate one-time improvements in PISA scores with long-lasting and significant improvements in scores over time. A prime example of how this approach is problematic is evident in the significant decline in PISA scores of Finland. Finland was ranked among the top five countries for the first three PISA cycles between 2000 and 2006, yet it experienced the largest decrease in country-level average PISA scores of any country in our sample between 2006 and 2012. Thus, examining trends in PISA scores over multiple assessments not only provides a more accurate picture of significant and non-significant improvements in educational achievement over time but also the ability to distinguish between the two.

While this study draws important conclusions about international trends in PISA scores over time, it is not without limitations. Foremost, we recognize that PISA assessments are not intended to comprehensively assess all aspects of students' scholastic experiences and abilities. In addition, we acknowledge that the associations we identify here do not provide causal explanations regarding significant increases in test scores over time. We also recognize that neither our study nor the PISA data fully accounts for geographic, social, cultural, socioeconomic, ethnic, and religious differences between and within countries. Nevertheless, our study does offer a useful framework for interpreting and comparing international achievement scores over time, as well as for assessing general shifts that are associated with improvements in test scores. Finally, we acknowledge that our data are unable to look at fine distinctions among countries' advancements to more democratic forms of government or higher income classifications. Examining more meso-level conditions associated with democracy or national wealth could provide a more nuanced understanding of the conditions associated with educational improvement. Such meso-level conditions might also help explain why Qatar, for example, experienced significant improvements in PISA scores even though it has only taken steps towards democratization and not fully transitioned to a democratic form of government. Future studies should examine these relationships further.

Our results have important implications for future social and educational practice, policy, and research. First, the finding that the majority of countries do not significantly increase their PISA scores over time indicates the need for countries to reassess their expectations regarding test scores and the ways in which they interpret differences over time. Given that PISA assessments are administered

to a new cohort of 15-year-old students every three years, most countries should not expect their scores to significantly increase from one PISA assessment to the next, or even over multiple PISA assessments. Rather, countries should consider their own historic patterns of growth and plateaus in growth when setting outcome-based goals for improvement, as well as reasonable timeframes for when such results should be observed. Furthermore, when seeking to improve test scores over time, our findings suggest that policy efforts outside of the educational sphere that improve the general social and economic well-being of students are more likely to be effective. Many countries, including the United States, the United Kingdom, and Japan, try to increase test scores by implementing micro-level educational reforms and policies patterned after high performing countries (Takayama 2008; Department for Education 2013; Duncan 2013). However, our findings suggest that high-performing countries are not the ones to look to when seeking to improve test scores, because they are not experiencing significant growth in scores. Based on the experiences of the ten countries in our sample that did significantly increase their PISA scores over time, policy efforts tailored towards improving the sociopolitical and the economic contexts of students are more likely to improve academic performance.

With regard to research, our findings point to the need for further study of international trends in test scores over time. In particular, further examination of differences in test scores over longer periods of time and with larger samples of countries may yield new insights into which groups of countries are making long-lasting improvements in their education systems and the processes that are contributing to such changes. Examining other international assessments along with PISA, such as the Trends in International Mathematics and Science Study (TIMSS), could also lead to greater understanding regarding significant improvements in academic performance. Future research may also benefit from analyzing whether interactions between macro- and micro-level policies foster meaningful educational improvements. For example, while we find that macro-level policies are associated with significant increases in test scores, the implementation of various micro-level policies may influence the magnitude of such improvements or the degree to which such improvements can be sustained over longer time periods. Studies focused on individual countries in which both micro- and macro-level policies could be identified would be especially useful in examining these issues. Such studies that are more narrowly focused could fill in the details of the general policy framework provided by our study regarding the mechanisms that lead to academic development and improvement.

Author Contributions: Conceptualization: K.J.R., M.J.D., C.E., and J.A.J.; Methodology: K.J.R., M.J.D., C.E., and J.A.J.; Software: K.J.R. and C.E.; Validation: K.J.R. and C.E.; Formal Analysis: K.J.R., S.M.M., and C.E.; Investigation: K.J.R., S.M.M., M.J.D., and C.E.; Data Curation: K.J.R., S.M.M., and C.E.; Writing—Original Draft Preparation: K.J.R. and S.M.M.; Writing—Review & Editing: K.J.R., S.M.M., M.J.D., C.E., and J.A.J.; Visualization: K.J.R. and S.M.M.; Supervision: K.J.R., M.J.D., and J.A.J.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Akiba, Motoko, Gerald K. LeTendre, and Jay P. Scribner. 2007. Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries. *Educational Researcher* 36: 369–87. [\[CrossRef\]](#)
- Baker, David, and Gerald K. LeTendre. 2005. *National Differences, Global Similarities: World Culture and the Future of Schooling*. Stanford: Stanford University Press.
- Bloom, Howard S., Carolyn J. Hill, Alison Rebeck Black, and Mark W. Lipsey. 2008. Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions. *Journal of Research on Educational Effectiveness* 1: 289–328. [\[CrossRef\]](#)
- Carnoy, Martin, and Richard Rothstein. 2013. *What Do International Tests Really Show about U.S. Student Performance?* Washington, DC: Economic Policy Institute, Available online: <http://www.epi.org/publication/us-student-performance-testing/> (accessed on 29 July 2019).
- Choi, Álvaro, and John Jerrim. 2016. The Use (and Misuse) of PISA in Guiding Policy Reform: The Case of Spain. *Comparative Education* 52: 230–45. [\[CrossRef\]](#)

- Coughlan, Sean. 2013. How PISA Became the World's most Important Exam. *BBC News*. Available online: <https://www.bbc.com/news/business-24988343> (accessed on 29 July 2019).
- Darling-Hammond, Linda, and Ann Lieberman, eds. 2012. *Teacher Education around the World: Changing Policies and Practices*. New York: Routledge.
- Deng, Zongyi, and S. Gopinathan. 2016. PISA and High-Performing Education Systems: Explaining Singapore's Education Success. *Comparative Education* 52: 449–72. [CrossRef]
- Department for Education. 2013. Secretary of State for Education Michael Gove's Statement in the House on the OECD's 2012 PISA Results. Available online: <https://www.gov.uk/government/speeches/2012-oecd-pisa-results> (accessed on 29 July 2019).
- Dobbins, Michael, and Kerstin Martens. 2012. Towards an Education Approach À La Finlandaise? French Education Policy after PISA. *Journal of Education Policy* 27: 23–43. [CrossRef]
- Duncan, Arne. 2013. The Threat of Educational Stagnation and Complacency: Remarks of US Secretary of Education Arne Duncan at the Release of the 2012 Program for International Student Assessment (PISA). Available online: <https://www.ed.gov/news/speeches/threat-educational-stagnation-and-complacency> (accessed on 29 July 2019).
- Egelund, Niels. 2008. The Value of International Comparative Studies of Achievement—A Danish Perspective. *Assessment in Education: Principles, Policy & Practice* 15: 245–51.
- Elliott, Julian, Lazar Stankov, Jihyun Lee, and Jens F. Beckmann. 2019. What Did PISA and TIMSS Ever Do for Us?: The Potential of Large Scale Datasets for Understanding and Improving Educational Practice. *Comparative Education* 55: 133–55. [CrossRef]
- Ertl, Hubert. 2006. Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany. *Oxford Review of Education* 32: 619–34. [CrossRef]
- Felipe, Jesus, Utsav Kumar, and Reynold Galope. 2017. Middle-Income Transitions: Trap or Myth? *Journal of the Asia Pacific Economy* 22: 429–53. [CrossRef]
- Feniger, Yariv, Idit Livneh, and Abraham Yogev. 2012. Globalisation and the Politics of International Tests: The Case of Israel. *Comparative Education* 48: 323–35. [CrossRef]
- Gür, Bekir S., Zafer Celik, and Murat Özoğlu. 2012. Policy Options for Turkey: A Critique of the Interpretation and Utilization of PISA Results in Turkey. *Journal of Education Policy* 27: 1–21. [CrossRef]
- Hanushek, Eric A., Paul E. Peterson, and Ludger Woessmann. 2013. *Endangering Prosperity: A Global View of the American School*. Washington, DC: Brookings Institution Press.
- Hargreaves, Andy P., and Dennis L. Shirley, eds. 2009. *The Fourth Way: The Inspiring Future for Educational Change*. Thousand Oaks: Corwin Press.
- Hess, Rick. 2013. 7 Reasons I Don't Care About the PISA Results. *Education Week*. Available online: https://blogs.edweek.org/edweek/rick_hess_straight_up/2013/12/7_reasons_i_dont_care_about_the_pisa_results.html (accessed on 29 July 2019).
- Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives* 2: 172–77. [CrossRef]
- Huntington, Samuel P. 2012. *The Third Wave: Democratization in the Late 20th Century*, 4th ed. Norman: University of Oklahoma Press.
- Lee, Valerie E., and Anthony S. Bryk. 1989. A Multilevel Model of the Social Distribution of High School Achievement. *Sociology of Education* 62: 172–92. [CrossRef]
- Lipsey, Mark W., Kelly Puzio, Cathy Yun, Michael A. Hebert, Kasia Steinka-Fry, Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick. 2012. *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms*. NCSE 2013-3000. Washington, DC: National Center for Special Education Research, U.S. Department of Education.
- Macdonald, Kevin. 2008. PV: Stata Module to Perform Estimation with Plausible Values. Available online: <https://econpapers.repec.org/software/bocbocode/s456951.htm> (accessed on 29 July 2019).
- Meyer, John W. 1977. The Effects of Education as an Institution. *American Journal of Sociology* 83: 55–77. [CrossRef]
- Meyer, Heinz-Dieter, and Aaron Benavot, eds. 2013. *PISA, Power, and Policy: The Emergence of Global Educational Governance*. Oxford: Symposium Books Ltd.
- Meyer, John W., Francisco O. Ramirez, and Yasemin Nuhoğlu Soysal. 1992. World Expansion of Mass Education, 1870–1980. *Sociology of Education* 65: 128–49. [CrossRef]

- National Center for Education Statistics. 2015. Methodology and Technical Notes—International Requirements. Available online: http://nces.ed.gov/surveys/pisa/pisa2012/pisa2012highlights_9a.asp (accessed on 29 July 2019).
- Organization for Economic Co-Operation and Development (OECD). 2010. Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States. doi:10.1787/9789264096660-en. Available online: <http://www.oecd.org/pisa/46623978.pdf> (accessed on 29 July 2019).
- Organization for Economic Co-Operation and Development (OECD). 2012. PISA 2009 Technical Report. doi:10.1787/9789264167872-en. Available online: <https://www.oecd.org/pisa/pisaproducts/50036771.pdf> (accessed on 29 July 2019).
- Organization for Economic Co-Operation and Development (OECD). 2014a. PISA 2012 Results in Focus: What 15-Year-Olds Know and What They Can Do with What They Know. Available online: <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf> (accessed on 29 July 2019).
- Organization for Economic Co-Operation and Development (OECD). 2014b. PISA 2012 Results: What Students Know and Can Do—Student Performance in Mathematics, Reading and Science, Vol 1. doi:10.1787/9789264201118-en. Available online: <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf> (accessed on 29 July 2019).
- Ramirez, Francisco O., and John Boli. 1987. The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization. *Sociology of Education* 60: 2–17. [CrossRef]
- Rathmell, Andrew, and Kirsten Schulze. 2000. Political Reform in the Gulf: The Case of Qatar. *Middle Eastern Studies* 36: 47–62. [CrossRef]
- Rutkowski, Leslie, and David Rutkowski. 2016. A Call for a More Measured Approach to Reporting and Interpreting PISA Results. *Educational Researcher* 45: 252–57. [CrossRef]
- Sahlberg, P. 2011. *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* New York: Teachers College Press.
- Sellar, Sam, and Bob Lingard. 2013. Looking East: Shanghai, PISA 2009 and the Reconstitution of Reference Societies in the Global Education Policy Field. *Comparative Education* 49: 464–85. [CrossRef]
- Takayama, Keita. 2008. The Politics of International League Tables: PISA in Japan's Achievement Crisis Debate. *Comparative Education* 44: 387–407. [CrossRef]
- Tallmadge, G. Kasten. 1977. *The Joint Dissemination Review Panel Ideabook*. Washington, DC: National Institute of Education & U.S. Office of Education.
- Tucker, M. 2013. Response to the Brookings Institution Attack on PISA. *Education Week*. Available online: https://blogs.edweek.org/edweek/top_performers/2013/12/response_to_the_brookings_institution_attack_on_pisa.html (accessed on 29 July 2019).
- United Nations. 2000. United Nations Millennium Declaration. Available online: <http://www.un.org/millennium/declaration/ares552e.pdf> (accessed on 29 July 2019).
- Wilby, Peter. 2013. Don't Let Dubious Pisa League Tables Dictate How We Educate Our Children. *The Guardian*. Available online: <https://www.theguardian.com/commentisfree/2013/dec/01/dont-let-pisa-league-tables-dictate-schooling> (accessed on 29 July 2019).
- World Bank. 2015. Measuring Inequality. Available online: <http://go.worldbank.org/3SLYUTVY00> (accessed on 29 July 2019).

