RESEARCH IN REVIEW 2012-5

CollegeBoard

International Comparative Assessments:

Broadening the Interpretability, Application and Relevance to the United States

By F. Tony Di Giacomo, Bethany G. Fishbein and Vanessa W. Buckley

F. Tony Di Giacomo is a director in Research at the College Board.

Bethany G. Fishbein is a graduate student intern at the College Board.

Vanessa W. Buckley was an intern for the College Board.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators and schools. For further information, visit www.collegeboard.org.

TRENDS

For more information on College Board research and data, visit research.collegeboard.org.

Contents

Figures

Executive Summary

Many articles and reports have reviewed, researched, and commented on international assessments from the perspective of exploring what is relevant for the United States' education systems.[1] Researchers make claims about whether the top-performing systems have transferable practices or policies that could be applied to the United States. However, looking only at top-performing education systems may omit important knowledge that could be applied from countries with similar demographic, geographic, linguistic, or economic characteristics — even if these countries do not perform highly on comparative assessments. Moreover, by exploring only the top performers, a presumption exists that these international assessments are in alignment with a country's curricular, pedagogic, political, and economic goals, which may falsely lead to the conclusion that by copying top performers, test scores would invariably increase and also meet the nation's needs. While international comparative assessments can be valuable when developing national or state policies, the way in which they are interpreted can be broadened cautiously to better inform their interpretability, relevance, and application to countries such as the United States — all while considering the purpose of each international assessment in the context of a nation's priorities. Ultimately, this report serves as a reference guide for various international assessments, as well as a review of literature that explores a possible relationship between national economies and international assessment performance. In addition, this review will discuss how policymakers might use international assessment results from various systems to adapt successful policies in the United States.

1. We intentionally refer to the United States education systems as plural, as any reforms made would have to be applied individually to all 50 states, the District of Columbia, and the U.S. territories.

Introduction

Over the last several administrations of various international assessments that measure academic proficiency in reading, mathematics, and science, the United States performed around average. The results of international assessments have driven concern among United States policymakers since 1964, when the First International Mathematics Study (FIMS) was conducted (Baker, 2007), and continued with the administration of international assessments such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA), both of which are described in this review.

Many researchers with varied perspectives on how to improve the United States performance have explored these results. One major approach is to explore the education landscapes and contexts of top-performing systems. For example, Tucker (2011) presented the frameworks of the various top-performing systems for adaptation in the U.S. While he offered an important lens and framework for guiding policy decision making at the state level, Biddle's (2012) critical review of this publication highlighted a few limitations of this approach.

First, assuming that the U.S. would join the ranks of high-performing systems by adopting features of their education systems undervalues certain aspects of U.S. education, such as creativity and teamwork, which have not been assessed in international comparative studies, as well as epistemological differences, because different cultures value different knowledge. In addition, Biddle (2012) pointed to the fact that the U.S. cannot necessarily "join the ranks" of these systems without addressing societal problems, such as youth poverty. Tucker (2011) pointed to two features that he felt are especially important for improving U.S. education: a high-quality teaching force and coherence in the design of the overall education system. Tucker posited that the United States lacks high standards for teaching and logically ordered curricula that are connected to national standards. While Biddle agreed in this respect, he further argued that U.S. education has some advantages, such as the large variety of subjects for both academics and career-related work. Biddle claimed that international assessments have never assessed the breadth of student interests. He questioned whether an education system can be considered "superior" because its students perform well on a few assessments. Biddle further argued that while Tucker's book provided great insight to features of high-performing education systems, other factors beyond top performance are valuable for informing policy. Alternatives to solely exploring the systems and practices of top performers take into account the methodological and logical limitations of international assessments as discussed by Theisen, Achola, and Boakari (1983), who presented the three most important functions of cross-national studies:

1. Comparisons of relative achievement status by subject and country;
2. Gleaning policy implications in one nation from what has been found to be related to the achievement in others; and
3. Reassessments of in-country expenditure priorities to boost achievement scores.

The researchers, like Biddle (2012), also warned against formulating policy based on achievements in other nations, claiming that analysts frequently fail to acknowledge that differences in cultural context may affect the causal variables. Theisen et al. (1983) suggested analyzing indicators of achievement in relation to context and individual factors related to education.

In this review, we will explore how the U.S. and other countries can make use of assessment results to improve their education systems at the state or national levels, in addition to considering some possible pitfalls of examining achievement in relation to other nations. In support of Biddle's (2012) argument, Theisen et al.'s (1983) warnings, and other warnings of researchers presented in this review, we also consider successful educational practices of nations that may not be high-performing but have characteristics similar to the U.S., such as geographical size, ethnic diversity, and economy. In addition, we consider the cultural contexts of those successful features of various nations indicated by international assessment results. By providing a thorough overview of the assessments, we will lay the foundation for exploring the interpretability, application, and relevance of these assessments to the United States.

In the first section, "Overview of International Assessments," we provide a brief reference table for four common international assessments, followed by individual sections for each assessment. In the "National Assessments" section, assessments used in high-performing systems are given brief mention, followed by a discussion of the national assessments used in the United States. Later, in the "Suggestions for Using International Assessments" section, we suggest possible ways to make use of the results of international assessments for approaching education policy decisions. Further, in "International Assessments: Economic Value," we explore the possible existence of a relationship between nations' economies and international assessment performances. We present the middling performance of the United States in the section titled "Summary: U.S. and State Performance on a Global Level," while also making reference to the markedly high performance of some individual states. Lastly, in the "International Assessments and Common Core in a Decentralized System" section, we consider the benefits of examining policies of particular systems, as well as the Common Core State Standards initiative for application in the United States.

Overview of International Assessments

Table 1 provides a brief overview of the assessments in regard to general test information, purpose, population, and administration. In-depth information about the various international assessments will follow.

Table 1.

Overview of International Assessments

	PISA	TIMSS	PIRLS	CIVED
Assessment Name	Programme for International Student Assessment	Trends in International Mathematics and Science Study	Progress in International Reading Literacy Study	Civic Education Study
Primary Purpose	• Evaluates education systems of various countries; • Assesses the extent to which students have acquired the knowledge and skills that are crucial for participating fully in society; • Provides a knowledge base for policy analysis and research; and • Measures trends over time related to student and school characteristics	• Measures trends in student achievement in mathematics and science; • Gathers information about learning contexts for mathematics and science; • Gathers data about the mathematics and science curricula in each country; and • Provides countries with information to improve teaching and learning	• Measures trends in reading literacy achievement in primary school to help strengthen the teaching and learning of reading skills; • Measures change in reading achievement; and • Investigates experiences children have at home and in school when learning to read	• Examines the context and meaning of civic education in several countries; • Gathers information about civic knowledge, attitudes, and engagement of students; and • Informs education practitioners and policymakers, parents, and citizens about the status of civic education
Subject Areas Tested	Reading, mathematics, science	Mathematics, science	Reading	Democracy and citizenship, national identity, social cohesion and diversity
Responsible Organization	Organisation for Economic Co-operation and Development (OECD)	International Association for the Evaluation of Educational Achievement (IEA)	International Association for the Evaluation of Educational Achievement (IEA)	International Association for the Evaluation of Educational Achievement (IEA)
Years of Administration	2000, 2003, 2006, 2009, 2012	1995, 1999, 2003, 2007, 2011	2001, 2006, 2011	1996–1997, 1999
Grade/Age Assessed	15-year-olds	Grades 4 and 8	Grade 4	14-year-olds, upper-secondary students
Type of Test	Criterion-referenced	Criterion-referenced	Criterion-referenced	Criterion-referenced
Achievement Levels Reported	Reading 1a–5, Mathematics 1–6, Science 1–6	Low, intermediate, high, advanced	Low, intermediate, high, advanced	Not Applicable

Note: This table is adapted from Egan, Beattie, Byrd, Chadwick, and DeCandia (2011). Additional information for CIVED, PIRLS, and TIMSS is from the International Association for the Evaluation of Educational Achievement (2011), and additional information for PISA is from the Organisation for Economic Co-operation and Development (2009).

In this section, we provide information on four assessments: the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), and the Civic Education Study (CIVED). Later, we provide information about the National Assessment of Educational Progress (NAEP), which, while not international, can provide important information for state benchmarking purposes. For each assessment, we discuss its purpose, methods, and participants, and offer some critique as well as examples of how the assessment data have been used in further studies. The results discussed for each assessment are not intended to be exhaustive; rather, they are intended to provide a glimpse into what the assessment studies have found and how future research may use the international assessment data.

PISA

Purpose. In 1997, the Organisation for Economic Co-operation and Development (OECD) began conducting an international study called the Programme for International Student Assessment (PISA) to evaluate education systems across the world (OECD, n.d.).[2] Most recently administered in 2012, PISA tests 15-year-old students from various countries and economies to assess the extent to which these students have acquired the knowledge and skills necessary for participating successfully within society and solving real-life problems.[3] PISA assesses reading literacy, mathematics literacy, science literacy, and problem solving in terms of achievement but also in terms of skills essential for solving what they refer to as "life's problems." Data collection in 2012 focused on mathematics, and countries could participate in an optional assessment of financial literacy (OECD, n.d.). PISA results can be useful in a number of ways. In addition to assessing students' capacity to apply knowledge and skills, PISA now also assesses students' ability to analyze, reason, and communicate effectively (OECD, 2010a). PISA allows for examining differences in performance patterns across countries and identifying common features among high-performing students, schools, and education systems. Countries can also use PISA results to monitor their progress in meeting curricular goals (OECD, 2009; 2010a). PISA is meant to be a long-term, ongoing program that will allow readers to examine trends in knowledge and skills of students in various countries/economies[4] and with various demographic characteristics (OECD, 2010a).

PISA allows for examining differences in performance patterns across countries and identifying common features among high-performing students, schools, and education systems.

Participants. Over 70 countries and economies now participate in PISA, and cycles were completed in 2000, 2003, 2006, 2009, and 2012 (OECD, n.d.). Between 4,500 and 10,000 students from each country/economy participated in each administration (OECD, 2009). Figure 1 shows a map of the participating countries and economies of PISA 2009, and Table 2 lists them, distinguishing OECD countries from non-OECD countries and economies.

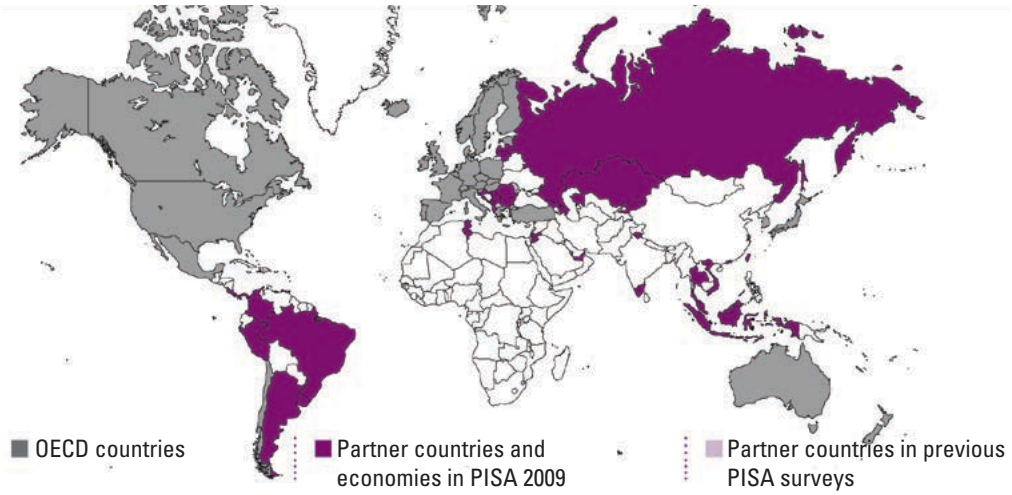
2. The OECD is an international organization that assists governments facing economic, social, and governance issues in a globalized economy. Visit www.oecd.org for more information.

3. Results from the PISA 2012 data collection will be released in December of 2013.

4. The OECD and the authors of the current paper use "country/economy" to refer to any PISA participant.

Figure 1.

A map of PISA countries and economies.



Source: OECD (2009).

Table 2.

PISA Participants 2009

OECD Countries		Partner Countries and Economies	
Australia	Japan	Albania	Macao-China
Austria	Korea	Argentina	Republic of Montenegro
Belgium	Luxembourg	Azerbaijan	Panama
Canada	Mexico	Brazil	Peru
Chile	Netherlands	Bulgaria	Qatar
Czech Republic	New Zealand	Colombia	Romania
Denmark	Norway	Croatia	Russian Federation
Dubai (UAE)	Poland	Hong Kong-China	Republic of Serbia
Estonia	Portugal	Indonesia	Shanghai-China
Finland	Slovak Republic	Jordan	Singapore
France	Slovenia	Kazakhstan	Chinese Taipei
Germany	Spain	Kyrgyz Republic	Thailand
Greece	Sweden	Latvia	Tunisia
Hungary	Switzerland	Liechtenstein	Uruguay
Iceland	Turkey	Lithuania	
Ireland	Trinidad and Tobago		
Israel	United Kingdom		
Italy	United States		

Note: Adapted from OECD (2009).

Methods. PISA 2009 contained test items in both multiple-choice and constructed-response formats (OECD, 2011). Multiple-choice questions were organized based on passages or graphics that relate to real-life situations that students may encounter. The majority of the test consisted of pencil-and-paper tasks, and students from 20 countries were administered some sections electronically to assess their ability to read digital texts. For the 2009 assessment, paper-and-pencil item tasks were arranged in 13 clusters: seven for reading, three for math, and three for science. Clusters were arranged in 13 booklets using a rotated test design so that each booklet contained four clusters. Students were assigned to one booklet that took about two hours to complete. Students were also administered a questionnaire to collect information regarding background, learning habits, attitudes toward reading, and involvement and motivation. School principals were given questionnaires to collect information about school characteristics, including demographic characteristics and the quality of the learning environment. Optionally, countries could have parents of students complete a questionnaire that focused on the students' past learning experiences, parents' reading engagement, home reading resources and support, and the parents' perceptions of and involvement in the school (OECD, 2009).

PISA scores follow a normal distribution with a mean of 500 and a standard deviation of 100, indicating that two-thirds of students in OECD countries scored between 400 and 600 points (OECD, 2009). Student performance on each subtest is represented by proficiency levels on a scale created using item response theory (IRT). Higher levels represent the ability to solve more complex and difficult problems or tasks. There are five proficiency levels for reading, six for mathematics, and six for science. Table 3 compares proficiency levels to PISA scale scores.

Table 3.

PISA Scale Scores and Proficiency Levels

Reading		Mathematics		Science	
PISA Scale Score	Proficiency Level	PISA Scale Score	Proficiency Level	PISA Scale Score	Proficiency Level
Above 625	5	Above 669.2	6	Above 707.8	6
553–625	4	607.0–669.2	5	633.3–707.8	5
481–552	3	544.7–606.9	4	558.7–633.2	4
408–480	2	482.4–544.6	3	484.1–558.6	3
335–407	1	420.1–482.3	2	409.5–484.0	2
Below 335	Below level 1	357.8–420	1	334.9–409.4	1

Note: Adapted from OECD (2009).

Each 2009 subtest was composed of additional subscales. Reading subscales included retrieving information, interpreting texts, and reflection and evaluation. The four math subscales included shape and space, change and relationships, quantity, and uncertainty. Lastly, the science subtest had three subscales: identifying scientific issues, explaining phenomena scientifically, and using scientific evidence (OECD, 2009).

PISA 2009 results. The OECD (2010a; 2011) as well as Paine and Schleicher (2011), authors of a report that recommended certain reforms to U.S. policy based on PISA 2009 results, presented findings from the most recent PISA study, while specifically highlighting major differences between policies of the U.S. and those of high-performing countries. Table 4 presents the top 10 and bottom 10 performing countries/economies on the overall average reading scale, in addition to how these countries/economies scored on the mathematics and science scales. According to these data, the U.S. scored higher than the OECD average on

the overall reading and science scales, but not statistically significantly higher, and scored statistically significantly lower than the OECD average on the mathematics scale (OECD, 2010a). As Table 4 shows, higher scores on each scale tended to be associated with higher scores on the other scales.

Table 4.

Comparing Countries' and Economies' PISA 2009 Performance

	Reading Scale	Mathematics Scale	Science Scale
OECD Average	**493**	**496**	**501**
Top 10 Countries/Economies			
Shanghai-China	556	600	575
Korea	539	546	538
Finland	536	541	554
Hong Kong-China	533	555	549
Singapore	526	562	542
Canada	524	527	529
New Zealand	521	519	532
Japan	520	529	539
Australia	515	514	527
Netherlands	508	526	522
United States	500	487	502
Bottom 10 Countries/Economies			
Tunisia	404	371	401
Indonesia	402	371	383
Argentina	398	388	401
Kazakhstan	390	405	400
Albania	385	377	391
Qatar	372	368	379
Panama	371	360	376
Peru	370	365	369
Azerbaijan	362	431	373
Kyrgyzstan	314	331	330

Note: Adapted from OECD, PISA 2009 Database. OECD members are indicated in bold.

The researchers found that the difference in scores between the highest- and lowest-performing OECD countries is equivalent to more than two school years, and the gap between the highest and lowest partner country/economy is even larger — equivalent to more than six years of schooling. In addition, the countries/economies with the highest overall reading performance (i.e., Korea, Finland, Hong Kong-China, and Shanghai-China) had the least variation in individual students' scores (OECD, 2010a).

The OECD (2010a) also examined how social background related to performance on PISA 2009. The researchers found that the school systems that performed the highest on PISA 2009 provided equal education to all students, regardless of the socioeconomic status of the individual or of the school attended. Canada, Finland, Japan, Korea, Hong Kong-China, and Shanghai-China all performed higher than the OECD mean, and this high performance remained for individual students within each of these countries/economies. In addition, students who attended schools with more socioeconomically advantaged students tended

to perform better, regardless of individual background. Although socioeconomic background was associated with test performance, lower performance did not always imply that the student or school was disadvantaged. According to Paine and Schleicher (2011) and the OECD (2011), socioeconomic differences accounted for a larger proportion of student variation in performance in the U.S. than in high-performing countries. In Japan, only 9% of a student's score was explained by socioeconomic differences, while in the U.S., 17% was explained by these differences. The possible relationship between economy and educational achievement is discussed in the "International Assessments: Economic Value" section.

The OECD (2010a; 2011) and Paine and Schleicher (2011) further explored teacher quality among PISA participants, while highlighting an important distinction between U.S. policy and policies of top-performing OECD nations. OECD countries (with the exception of Israel, Slovenia, Turkey, and the U.S.) tended to place a larger number of teachers in schools with more socioeconomically disadvantaged students; however, PISA findings suggested that these teachers were not necessarily of better quality. In addition, the U.S. was one of the few OECD countries that did not follow this practice (OECD, 2011). Table 5 presents the correlations between socioeconomic background of schools and the quality of teachers and the student-teacher ratio for the top 10 and bottom 10 performing countries/economies by average 2009 reading score. The U.S. had a statistically significantly lower correlation (-0.17) between socioeconomic background of schools and the student-teacher ratio than the OECD average, indicating that in the U.S., on average, lower socioeconomic school background was associated with a higher student-teacher ratio (i.e., more students per teacher).

Socioeconomic differences accounted for a larger proportion of student variation in performance in the U.S. than in high-performing countries … in the U.S., on average, lower socioeconomic school background was associated with a higher student-teacher ratio.

Table 5.

PISA Measures of Educational Equity

	Overall 2009 Reading Score	Correlation between the socioeconomic school background and percentage of teachers with university-level (ISCED 5A) among all full-time teachers	Correlation between socioeconomic school background and student-teacher ratio
OECD Average	**493**	**0.15**	**-0.15**
Top 10 Countries/Economies			
Shanghai-China	556	**0.32**	**-0.13**
Korea	539	-0.03	0.30
Finland	536	-0.01	0.08
Hong Kong-China	533	0.12	0.02
Singapore	526	**0.22**	**-0.14**
Canada	524	0.03	0.09
New Zealand	521	0.07	0.11
Japan	520	0.20	**0.38**
Australia	515	0.02	-0.07
Netherlands	508	**0.62**	**0.38**
United States	500	0.10	**-0.17**
Bottom 10 Countries/Economies			
Tunisia	404	0.20	-0.02
Indonesia	402	0.16	**-0.16**
Argentina	398	0.22	-0.02
Kazakhstan	390	**0.34**	**0.44**
Albania	385	**0.38**	0.15
Qatar	372	**-0.07**	0.11
Panama	371	-0.13	0.03
Peru	370	0.48	-0.02
Azerbaijan	362	**0.44**	0.23
Kyrgyzstan	314	**0.35**	0.27

Note: Adapted from OECD (2010a). Values in bold indicate statistically significant differences from the OECD average.

Furthermore, Paine and Schleicher (2011) highlighted some additional differences between the education systems of high-performing countries and of the U.S. For example, the researchers noted that countries with the highest performance had higher teacher salaries, more valued education credentials, and more education spending devoted to instructional services. In countries such as Finland, Japan, and Singapore, teachers had a higher status than in the U.S., as Paine and Schleicher (2011) stated:

> It is noteworthy that countries that have succeeded in making teaching an attractive profession have often done so not just through pay, but by raising the status of teaching, offering real career prospects, and giving teachers responsibility as professionals and leaders of reform. (p. 5)

In other words, increasing teacher salaries alone will not make the teaching profession more attractive in the U.S.; rather, more efforts may be necessary to increase responsibility and career satisfaction.

The authors also pointed out that the U.S. had very different spending patterns than other high-performing countries, in that the U.S. tended to spend more. In addition, the U.S. ranked comparably with Estonia and Poland, each of which spent half of what the U.S. does on education, and Luxembourg spent more money than the U.S. and scored significantly lower (Paine & Schleicher, 2011).

Utility of PISA data. One benefit of using PISA data is having the ability to determine what constitutes a successful school. Based on PISA 2009 results, the OECD (2010a) concluded that a successful school is one that performs above average and has fewer socioeconomic inequalities. The OECD also found that successful school systems were those with similar opportunities for learning. These schools embraced diverse students and personalized education. In countries where students tend to repeat grades more often, socioeconomic performance gaps were wider. Also, greater gaps were found where tracking occurs at younger ages. Notably, successful school systems placed priority on paying teachers more for better quality work, rather than hiring more teachers (OECD, 2010a). This practice may be important for policymakers to be aware of when considering the use of teacher incentives.

> …countries with the highest performance had higher teacher salaries, more valued education credentials, and more education spending devoted to instructional services.

Another publication by the OECD reviewed PISA and its value in terms of education reform, specifically as it relates to what the U.S. can learn from the PISA results. The OECD (2011) provided a definition for a high-performing country:

> This volume defines countries as high performing if: almost all of their students are in high school at the appropriate age, average performance is high and the top quarter of performers place among the countries whose top quarter are among the best performers in the world (with respect to their mastery of the kinds of complex knowledge and skills needed in advanced economies as well as their ability to apply that knowledge and those skills to problems with which they are unfamiliar); student performance is only weakly related to their socio-economic background; and spending per pupil is not at the top of the league tables. Put another way, this volume defines superior performance as high participation, high equity and high efficiency. (p. 14)

The OECD (2011) also provided a section devoted to how PISA can be used to help improve education systems in addition to examining causal relationships between various factors and performance. The authors stated the following ways in which PISA data can be used to improve education systems:

- PISA scores provide information regarding attainable educational achievements. For example, Finland had little variation in performance between schools, as those students coming from disadvantaged socioeconomic backgrounds did not always perform as poorly as students from similar backgrounds do in the U.S.

- The U.S. can use PISA scores of high-performing countries to set specific, measurable goals that have been achieved by these systems. PISA can also be used to monitor progress.
- PISA can be linked to national assessments. If the U.S. links its national assessments to PISA, as Oregon, Delaware, and Hawaii have already done, schools can be provided progress reports. Phillips and Jiang (2011) described how PISA is used for internationally benchmarking state performance standards. Items from PISA are embedded into state assessments and calibrated to the state scale, and common-item linking matches the state scale to the PISA scale. The linking can then determine which state standards are considered internationally competitive (Phillips & Jiang, 2011).[5]
- PISA data help countries determine the pace of improvement by validating scores internationally.
- The extensive background information collected by PISA tells us about factors associated with higher performance (OECD, 2011).

Paine and Schleicher (2011) argued that to be economically competitive with other countries/economies, the U.S. must improve the teaching profession and maintain common standards that are similar to those of the most successful school systems in the world. Paine and Schleicher suggested that improving PISA scores in the U.S. can narrow the achievement gap between the U.S. and other nations, in addition to improving the economy and gross domestic product (GDP). The researchers also stated that making such an improvement is possible because other countries have done so (e.g., Poland, South Korea, and Canada). In addition, substantial gains have been seen in achievement among U.S. schools and districts in Miami; Boston; Long Beach, California; and Charlotte-Mecklenburg, North Carolina; by improving failing schools (Paine & Schleicher, 2011).

The highlighted differences between the education systems of the U.S. and high-performing countries can be helpful to U.S. policymakers in making decisions regarding education funding and the status of the teaching profession. In addition, such findings can be useful for individuals, parents, and stakeholders to consider when making education decisions. Similar findings regarding factors related to high performance are found with other international assessments discussed later in this review. The results presented by the OECD (2010a; 2011) and Paine and Schleicher (2011) for the PISA 2009 study are important for understanding how the U.S. compares to other nations regarding a number of factors. Specific trends among certain countries/economies were also highlighted, indicating practices that are potentially beneficial for other countries/economies to adopt for themselves. For example,

5. See the "Linking NAEP with International Assessments" subsection on page 46.

… better performance in reading was associated with a number of factors, including equality of education, teacher characteristics, funding allocation, specific student practices and strategies, and personal reading habits.

better performance in reading was associated with a number of factors, including equality of education, teacher characteristics, funding allocation, specific student practices and strategies, and personal reading habits. A number of additional studies have used PISA data for similar purposes. Because of the large amount of literature concerning PISA and its uses, much research is beyond the scope of this review and will not be discussed here.

Critique. Some researchers have criticized the reliance of countries upon international assessments, specifically PISA. In a journal article, Bracey (2009) argued that the use of test scores, specifically average test scores, for comparing education systems is a mistake. According to PISA results, the U.S. ranked around the middle compared to other countries, although, as Salzman and Lowell (2008) pointed out, looking at the number of people with high scores in each country could be more effective, as not examining the amount of high and low performers makes scores "irrelevant as a measure of economic potential" (as cited in Bracey, 2009, p. 450). Looking at the number of people who reached the highest level on the PISA science test shows that the U.S. ranked first compared to Japan and Finland, both high-performing countries. Korea, also a high performer, had a smaller proportion of high scorers than the U.S. (1.1% vs. 1.5%). However, if we are to base performance upon the number of high-scoring students, we may also have to consider the number of low-scoring students, and the U.S. was the second lowest among all other OECD nations. Bracey emphasized that most of the variation was within the countries, rather than between, so perhaps the better solution is for the U.S. to compare itself to specific states that are successful rather than other nations. In addition, Bracey thought that the recommendations based on PISA results might not be culturally relevant: "Sending children to classes six days a week, extra preparation courses nights and weekends, and having a single examination that decides their fate, as is done in Japan, is not a choice most U.S. parents would make" (p. 450). Based on this idea, some lessons previously mentioned in this review may not be applicable, as they would require the U.S. to make fundamental cultural changes in addition to policy changes.

In an essay review of the 2006 OECD publication, *Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in PISA 2003*, Cummins (2008) also argued against the use of international assessments for the case of comparing instruction methods for immigrant and minority students across countries. Specifically, recommendations have been made based on PISA 2003 results that bilingual education for minority students should involve immersion in only the host language at an early age. However, as Cummins pointed out, empirical evidence exists that education in both languages is also effective for promoting academic achievement. PISA data showed large variations between countries in terms of immigrant student achievement. Interestingly, in Canada, second-generation students showed higher average achievement than native-born students. However, in Europe and the U.S., immigrants tended to have lower achievement, often significantly. On the contrary, in Denmark and Germany, second-generation students who only went to school in the host country showed lower achievement than first-generation students. Cummins suggested that based on these findings, more exposure to the host language is associated with worse performance in these countries. Furthermore, Cummins found problems with the interpretation of PISA results that claim underachievement is caused by a lack of opportunity to learn the host language. Cummins claimed that this interpretation ignores the fact that the relationship between the two does not imply causation. In addition, the direction of the relationship was not clear, and it is possible that underachievement caused a lack of opportunity for students. Lastly, no relationship was found between the language spoken at home and achievement in Australia and Canada, where immigrant students were found to have the highest achievement (Cummins, 2008). The argument presented by Cummins tells readers and policymakers to use caution when making decisions based on PISA results, as

Various countries have shown improvement in their education systems, which may signal to U.S. policymakers that improvement is possible for our country as well.

some factors only apply in specific countries, and associations between factors do not always depict a causal relationship.

Despite the arguments presented by Bracey (2009) and Cummins (2008) that caution against forming education policies similar to those of high-achieving countries without considering other factors such as cultural values and causal relationships, many of the policies that have shown to be successful in other countries *are* applicable elsewhere. Various countries have shown improvement in their education systems, which may signal to U.S. policymakers that improvement is possible for our country as well. Notably, Peterson, Woessmann, Hanushek, and Lastra-Anadón (2011) presented the PISA 2009 results indicating that Massachusetts alone is consistently part of the top 10 performing areas worldwide in both reading and mathematics. This statistic provides further evidence that improvement is possible in the U.S.

Some U.S. states perform comparably to the rest of the world; however, Peterson et al. (2011) reminded us that only five additional states — Kansas, Minnesota, New Jersey, North Dakota, and Vermont — have shown achievement comparable to that of Massachusetts. In addition, some of the country's wealthiest states were found to be among the world's lowest performers, including California, Florida, Michigan, Missouri, and New York. This reinforces the previously discussed idea that although socioeconomic background was associated with achievement in the U.S., this did not cause the majority of the variation among scores in high-performing countries, and other factors may be considered. The use of PISA data as indicators for what is associated with high performance, while being cautious to avoid misinterpretations, can provide valuable information for policymakers regarding the improvement of education systems worldwide; however, as Bracey (2009) and Cummins (2008) recommended, practices should not be duplicated without considering cultural factors.

TIMSS

Purpose. The International Association for the Evaluation of Educational Achievement (IEA) website (2011) summarizes the IEA's various international assessments. The IEA Trends in International Mathematics and Science Study (TIMSS) 2011 is the fifth cycle of TIMSS, having had previous cycles in 1995, 1999, 2003, and 2007 (IEA, 2011). TIMSS has been successful in measuring trends in student achievement in the areas of mathematics and science for the purpose of providing information to countries to help improve the teaching and acquisition of mathematics and science content (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009a). TIMSS allows countries to compare progress internationally in mathematics and science, monitor the effectiveness of teaching and learning, understand the most ideal learning contexts, and address internal policy issues. In addition, TIMSS administers questionnaires to gather data from students, teachers, and principals regarding the various contexts for learning mathematics and science, as well as to gather data regarding the curriculum in each country (Mullis et al., 2009a).

Participants. TIMSS is generally administered to students in the fourth and eighth grades, although certain countries administer the assessment to sixth- and ninth-grade students (IEA, 2011). Almost 70 countries now participate in TIMSS (IEA, 2011; Mullis et al., 2009a). In 2011, TIMSS had nine benchmarking states in the United States: Alabama, California, Colorado, Connecticut, Florida, Indiana, Massachusetts, Minnesota, and North Carolina (IEA, 2011). These states are able to compare student performance on a state level to other national participants. See Table 10 for the 2011 TIMSS and PIRLS participants.

Methods. In *TIMSS 2011 Assessment Frameworks*, Mullis et al. (2009a) described the content frameworks and the assessment design of TIMSS 2011, including major content and cognitive domains of mathematics and science that are covered by the assessments. According to these authors, TIMSS used curricula as the organizing model to best provide students with opportunities and to determine what factors influence the use of these opportunities. There are three aspects of the TIMSS curriculum: the intended curriculum, the implemented curriculum, and the achieved curriculum. Data regarding these aspects of learning and curriculum were gathered via the questionnaires that were administered to the National Research Coordinator in each country. In addition, teachers provided information regarding their preparation, experience, and attitudes; the mathematics and science content taught to TIMSS students; the instructional approaches used in teaching mathematics and science; and the resources available in classrooms. School principals provided information about school characteristics, resources, instructional time, and school climate. Finally, student questionnaires collected information concerning home lives, school lives, demographic information, school climate, and attitudes toward math and science (Mullis et al., 2009a).

TIMSS Advanced

TIMSS Advanced is an assessment that is administered to students in the final year of secondary school (usually 12th grade) to assess students' knowledge in advanced mathematics and physics. Having been administered in 1995 and most recently in 2008, TIMSS Advanced is meant for students who have engaged in studies to further prepare for the rigors of tertiary education. In 2008, 10 countries participated in TIMSS Advanced. The assessments will be administered again in 2015 and will include an optional population of first-year tertiary students (TIMSS & PIRLS International Study Center, 2012).

According to Mullis et al. (2009a), questionnaire data collected from TIMSS 2011 contained information about what improves teaching and learning in mathematics and science within four types of contexts: national and community contexts, school contexts, classroom contexts, and student characteristics and attitude. Table 11 presents the types of information collected by students, teachers, and principals via the PIRLS questionnaires so researchers can examine factors that affect students' learning of reading. The same information is collected for TIMSS, while being specific to mathematics and science learning (Mullis et al., 2009a).

The TIMSS 2011 assessment contained 28 item blocks; half for science, half for math (Mullis et al., 2009a). There were 10–14 items in each block for fourth grade, and 12–18 for eighth grade. Fourth-grade students were given 72 minutes of testing time, and eighth-graders were given 90 minutes. At least half of the total points were represented by multiple-choice questions, with the rest represented by constructed-response questions. The score distribution had a mean of 500 and a standard deviation of 100 (Mullis et al., 2009a). Scores were reported according to proficiency levels at TIMSS International Benchmarks that were established using item response theory (IRT). Benchmarks categorize student achievement as Advanced (625), High (550), Intermediate (475), or Low (400) (Olson, Martin, & Mullis, 2008). These international benchmarks will be used for future cycles of TIMSS.

In describing the mathematics assessment framework for TIMSS 2011 (which is similar to that of TIMSS 2007), Mullis et al. (2009a) described its organization around two dimensions. The first is a content dimension, which specifies the subject matter that is assessed (i.e., number, algebra, geometry, or data and chance), and the second is a cognitive dimension that specifies the thinking processes that are assessed (i.e., knowing, applying, or reasoning). Similarly, the TIMSS 2011 science assessment framework was organized around content and cognitive dimensions. The fourth-grade content domain included life science, physical science, and earth science, while the eighth-grade content domain included biology, chemistry, physics, and earth science. See Table 6 for target percentages of the TIMSS 2011 assessments to content areas.

Table 6.

Target Percentages of the TIMSS 2011 Assessments to Content Domains at Fourth and Eighth Grades

Grade	Content Domains	Percentages
	Mathematics Assessment	
4th Grade	Number	50%
	Geometric Shapes and Measures	35%
	Data Display	15%
8h Grade	Number	30%
	Algebra	30%
	Geometry	20%
	Data and Chance	20%
	Science Assessment	
4th Grade	Life Science	45%
	Physical Science	35%
	Earth Science	20%
8th Grade	Biology	35%
	Chemistry	20%
	Physics	25%
	Earth Science	20%

Note: Adapted from Mullis et al. (2009a).

TIMSS 2011 recognized the importance of scientific inquiry in teaching and learning, and stressed that the construct is best assessed in the context of one of the content domains and drawn-upon skills of the cognitive domains, rather than assessed in isolation. Therefore, related items assessed these aspects within the two dimensions (Mullis et al., 2009a).

TIMSS Results. Tables 7 and 8 present the top 10 and bottom 10 performing systems in mathematics and science, respectively, from the 2011 administration. The top and bottom performers are in descending order of the pooled average scale score that includes both fourth- and eighth-grade scale scores. As such, the tables only include systems in which both fourth- and eighth-grade student populations participated in the TIMSS 2011 assessment.

Table 7.

TIMSS 2011 Mathematics Scores

	Pooled Average Scale Score	4th-Grade Scale Score	8th-Grade Scale Score
Top 10 Countries			
Korea, Rep. of	609	605	613
Singapore	608.5	606	611
Chinese Taipei	600	591	609
Hong Kong SAR	594	602	586
Japan	577.5	585	570
Russian Federation	540.5	542	539
Finland	529.5	545	514
United States	525	541	509
England	524.5	542	507
Lithuania	518	534	502
Bottom 10 Countries			
Thailand	442.5	458	427
Georgia	440.5	450	431
Chile	439	462	416
Iran, Islamic Rep. of	423	431	415
Bahrain	422.5	436	409
Qatar	411.5	413	410
Saudi Arabia	402	410	394
Tunisia	392	359	425
Oman	375.5	385	366
Morocco	353	335	371

Note: Adapted from Mullis, Martin, Foy, & Arora (2012a). Only includes systems in which both fourth- and eighth-grade student populations participated.

Table 8.

TIMSS 2011 Science Scores

	Pooled Average Scale Score	4th-Grade Scale Score	8th-Grade Scale Score
Top 10 Countries			
Singapore	586.5	583	590
Korea, Rep. of	573.5	587	560
Finland	561	570	552
Japan	558.5	559	558
Chinese Taipei	558	552	564
Russian Federation	547	552	542
Hong Kong SAR	535	535	535
United States	534.5	544	525
Slovenia	531.5	520	543
England	531	529	533
Bottom 10 Countries			
Thailand	461.5	472	451
Bahrain	450.5	449	452
United Arab Emirates	446.5	428	465
Georgia	437.5	455	420
Saudi Arabia	432.5	429	436
Armenia	426.5	416	437
Qatar	406.5	394	419
Oman	398.5	377	420
Tunisia	392.5	346	439
Morocco	320	264	376

Note: Adapted from Martin, Mullis, Foy, & Stanco (2012). Only includes systems in which both fourth- and eighth-grade student populations participated.

Utility of TIMSS data. TIMSS data can be used in a variety of contexts, including studies that conclude with suggestions for education policies all over the world. Schütz, Ursprung, and Woessmann (2008) used the TIMSS 1995 and 2001 assessment data sets for their study on the effects of family background on students' educational performance. The results of Schütz et al.'s study imply suggestions for school systems worldwide. TIMSS data sets provide information gleaned from both score and questionnaire data, including educational performance, family background, and relevant control variables for students in all participating systems. By formulating an index of the inequality of educational opportunity in 54 countries, the authors found that educational tracking is associated with lower equality of opportunity in terms of family background, but extensive early childhood education increased the equality of educational opportunity for children from varied family backgrounds. In addition, the results showed that equality of opportunity varied across countries. Educational performance was measured by a pooled average score of the two TIMSS tests for countries from both studies, and family background was measured by the number of books students had in their homes, as indicated by the student questionnaire. The researchers found that generally, students in higher-performing systems tended to have more books per household than students in lower-performing systems (Schütz et al., 2008).

Furthermore, Schütz et al. (2008) found that in all countries, educational performance was statistically significantly influenced by the family background variable. Because students' educational performance was measured with standardized test scores and an international standard deviation of 100, these statistics can be interpreted as percentages of the international standard deviations that educational performance increased when raising the number of books at home by one category[6] (Table 9) (Schütz et al., 2008). The international standard deviation for TIMSS allows for easier interpretation of statistics.

Family background was found to have impacted student performance most in the following OECD member countries: England, Germany, Hungary, and Scotland, while students from Canada, Flemish Belgium, France, and Portugal were affected the least. The U.S. fell in the top 25% of OECD countries with the most unequal opportunity. OECD countries also exclusively showed a statistically significant association between equality of opportunity and mean test score for a country (Schütz et al., 2008).

Table 9.

Family Background Effects as an Index of Inequality of Educational Opportunity

	Family Background Effect
Top 10 Countries	
England	28.81
Taiwan (Chinese Taipei)	27.91
Scotland	26.95
Hungary	25.84
Germany	25.57
Korea	24.75
Macedonia	24.05
Slovak Rep.	24.01
Bulgaria	23.32
United States	23.13
Bottom 10 Countries	
Belgium (Flemish)	10.95
Hong Kong	10.82
Portugal	10.40
Canada	9.76
France	8.32
Colombia	7.55
Morocco	6.84
Tunisia	6.32
Indonesia	4.83
Kuwait	2.49

Notes: Adapted from Schütz et al. (2008). The coefficient estimate was on books at home. The dependent variable was TIMSS 1995 and 2001 international test score. Regressions controlled for age, gender, family status, whether the student was born in the country, whether the mother and father were born in the country, interactions between immigration variables and books, and a TIMSS 2001 dummy and a constant. Regression was weighted by students' sampling probabilities. OECD members are marked in bold.

The authors of the study also examined the interaction of variation across countries in education policies and family background at the individual student level to determine what

6. For example, category 4 (101–200 books) would increase to category 5 (more than 200 books).

impact education policies have on equality of opportunity. Schütz et al. (2008) found that for all TIMSS participants, the effect of family background was larger and equality of opportunity was lower when a country tracked its students into schools by ability earlier. Educational inequality was shown to increase preschool enrollment up to 60%, and decrease thereafter. The authors examined school characteristics to find that neither school starting age nor half-day versus whole-day schooling were associated with significant differences in equality of opportunity. In addition, neither average educational spending nor the country's level of economic development was associated with equality of opportunity (Schütz et al., 2008).

The results of the study by Schütz et al. (2008) can inspire reforms for education policies for schools worldwide, such as establishing comprehensive school systems and extensive early childhood education to increase the equality of educational opportunity for students from a variety of family backgrounds. The results also suggest what will not improve equality of opportunity for students; educational spending and length of the school day were not associated with an increase in equality of educational opportunity. In addition to providing suggestions for countries concerning education policies, the authors demonstrated that when analyzing TIMSS data, especially across countries, confounding variables can be controlled. For example, Schütz et al. (2008) stated that the varying immigrant populations in countries could cause a bias in cross-country estimates of family background effects when immigration status and family background are correlated and when family background effects are the same between native and immigrant families. The authors were able to control for these confounding variables in the construction of the family background effects measure (Schütz et al., 2008). In addition, the use of an international standard deviation allowed for easier interpretation. These facts demonstrate adequate use of statistical methods to provide for increased validity when analyzing TIMSS data.

Similarly, Stanco (2012) used TIMSS data to provide recommendations to countries regarding education policies, in addition to providing a model for future studies related to school effectiveness across various world contexts. Stanco used TIMSS 2007 data, which showed a gap in mathematics and science achievement between students in the U.S. and top-performing countries, to investigate how factors related to school effectiveness that were associated with greater science, technology, engineering, and mathematics (STEM) achievement in the United States compared to those factors in Chinese Taipei, the Czech Republic, Singapore, and Slovenia. STEM achievement was measured using TIMSS 2007 scores, and was examined in relation to factors of school effectiveness that are associated with school resources, fidelity of curriculum implementation, and school climate, while controlling for students' home resources. The results indicated that there were differences in how these factors operated across the countries. Strong predictors of STEM achievement included the absence of discipline problems, no attendance problems, and a supportive school climate. In addition, teacher preparation, teaching the curriculum, and the use of instructional strategies that involve scientific inquiry were found to be important in relation to STEM achievement (Stanco, 2012). Considering the results of this study in addition to the study by Schütz et al. (2008) can allow for improved policies for education across countries and provide a strong basis for further analyses.

Critique. Although studies have demonstrated the benefits of TIMSS 2011 (and earlier) data, researchers have also negatively critiqued the study. In Bracey's (2000) critique of TIMSS Advanced 1995 data, he stated, "…the systems and cultures of the nations involved differ to an extent that renders the scores uninterpretable" (p. 4). The author presented the popular interpretation of the 12th-grade TIMSS "final year" exam — that the U.S. is falling way behind the rest of the world with respect to mathematics and science achievement. However, because the average age of participating 12th-graders varied across countries, Bracey (2000) did not find

these interpretations to be valid. In addition, 12th-grade course work tended to vary among students, even within the U.S. itself, and this was identified by TIMSS staff in the opening pages of the "final year" report. In addition, certain U.S. states had higher average scores than that of the overall U.S. average as well as of the highest-scoring country.

To expand on Bracey's (2000) arguments, Wang (2001) presented concerns regarding TIMSS 1995 and 1999 primary and middle school data, and expressed uncertainty about country rankings released from the TIMSS as well as technical concerns regarding instrument construction, curricular inequivalency, and statistical outliers. Upon inspecting the data released by TIMSS, Wang found that the percentage of countries changing at least one position in rank ranged from 17% to 59% because of an ignored imputation error that could create inconsistency in reporting.

Furthermore, Wang (2001) presented critiques from various researchers that argue that TIMSS did not measure the effectiveness of one teaching method versus another, and pointed to an inconsistency of the emphasis on problem solving related to question format. Despite U.S. initiatives for greater emphasis on problem solving, the exam was predominately in multiple-choice question format, and did not focus on higher-order thinking. Wang (2001) finally discussed grade and content level differences across countries for primary and secondary school, discussed earlier by Bracey (2000) for secondary students, as well as age outliers, both of which have the potential to drastically affect the interpretation of scores.

Although the utility of TIMSS data exampled by Schütz et al. (2008) and Stanco (2012), one may want to approach interpreting scores and analyses with caution, based on the critiques by Bracey (2000) and Wang (2001). Inconsistencies in ranking and differences in education systems could cause biased interpretations. Regardless of the negative aspects of TIMSS as identified by researchers, TIMSS continues to be commonly used worldwide as an indicator of mathematics and science achievement and curricula, and is regarded as a good measure of this achievement. Additionally, more recent TIMSS assessment frameworks have remedied some of the problems the researchers have found (Mullis et al., 2009a).

... TIMSS continues to be commonly used worldwide as an indicator of mathematics and science achievement and curricula....

PIRLS

Purpose. The IEA's (2011) Progress in International Reading Literacy Study (PIRLS) allows for measuring reading achievement in various contexts. Most recently updated in 2011, PIRLS measures trends in reading literacy achievement in primary school to help strengthen the teaching and learning of reading skills worldwide (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009b). PIRLS is updated every five years, and PIRLS 2011 combined newly developed reading assessment passages and questions with relevant passages and questions from PIRLS 2006, and now allows for measuring change since 2001. PIRLS 2011 also investigated experiences that young children have both at home and in school when learning to read, by examining national policies and practices related to literacy and administering questionnaires to students, parents/caregivers, teachers, and school principals (IEA, 2011).

Participants. The international population for PIRLS 2011 consisted of students from approximately 55 countries (including the U.S. and Florida as its benchmarking state) and included students in the grade that is equivalent to four years of schooling. The mean age of test-takers was at least 9.5 years (IEA, 2011). See Table 10 for the 2011 PIRLS participants.

Table 10.

TIMSS and PIRLS Participating Education Systems, 2011

TIMSS & PIRLS		TIMSS Only	PIRLS Only
Australia	Lithuania	Armenia	Bulgaria
Austria	Malta	Bahrain	Colombia
Azerbaijan	Morocco	Chile	France
Belgium	Netherlands	Ghana	Honduras
Botswana	New Zealand	Honduras	Trinidad and Tobago
Canada	Northern Ireland	Japan	
Chinese Taipei	Norway	Jordan	
Croatia	Oman	Kazakhstan	
Czech Republic	Poland	Korea	
Denmark	Portugal	Lebanon	
England	Qatar	Macedonia	
Finland	Romania	Malaysia	
Georgia	Russian Federation	Palestinian National Authority	
Germany	Saudi Arabia	Serbia	
Hong Kong SAR	Singapore	Syria	
Hungary	Slovak Republic	Thailand	
Indonesia	Slovenia	Tunisia	
Iran	South Africa	Turkey	
Ireland	Spain	Ukraine	
Israel	Sweden	Yemen	
Italy	United Arab Emirates		
Kuwait	United States		

Source: IEA (2011).

Methods. Mullis et al. (2009b) provided the framework for PIRLS 2011, and explained the reason for choosing the fourth year of schooling as the focal point for PIRLS, similar to TIMSS. The fourth year is an important transition point in developing reading skills. PIRLS 2011 focused on three aspects of reading literacy: purposes for reading, processes of comprehension, and reading behaviors and attitudes. PIRLS included two purposes for reading, each of which made up half of the test: reading for literacy experience and reading to acquire and use information. There were four types of comprehension processes: focus on and retrieve explicitly stated information; make straightforward inferences; interpret and integrate ideas and information; and examine and evaluate content, language, and textual elements. Overall, the PIRLS booklets contained five literary and five informational passages, and the prePIRLS booklets contained three literary and three informational passages. Each booklet contained two passages with about 12 questions, half of which were multiple choice, and half of which were constructed response. Students were given 80 minutes to complete the test. Lastly, the questionnaires were given to students, parents, teachers, and principals to gather data on their experiences in developing reading literacy in various contexts (Table 11), and countries completed

PrePIRLS

Because there are countries where most fourth-grade children are still developing fundamental reading skills, the PIRLS assessment was extended to different grade levels by developing a less difficult reading assessment called prePIRLS. PrePIRLS is meant for students who are still learning how to read, and contains less difficult items while still measuring the same constructs. Scores can be validly compared to general PIRLS scores (IEA, 2011).

questionnaires about their education systems and reading curricula. The score distribution for PIRLS has a mean of 500 and a standard deviation of 100 (Mullis et al., 2009b). Similar to TIMSS, PIRLS proficiency levels are reported as Advanced (625), High (550), Intermediate (475), and Low (400) (Martin, Mullis, & Kennedy, 2007).

Table 11.

Types of Questionnaire Data Collected for PIRLS 2011, According to Context

Context	Data
National and Community Contexts	Languages and Emphasis on Literacy
	Demographics and Resources
	Organization and Structure of the Education System
	The Reading Curriculum in the Primary Grades
Home Contexts	Economic, Social, and Educational Resources
	Parental Emphasis on Literacy Development
	Parental Reading Behaviors and Attitudes
School Contexts	School Characteristics
	School Organization for Instruction
	School Climate for Learning
	School Resources
	Parental Involvement
Classroom Contexts	Teacher Education and Development
	Teacher Characteristics and Attitudes
	Classroom Characteristics
	Instructional Materials and Technology
	Instructional Strategies and Activities
	Assessment
Student Characteristics and Attitudes	Student Reading Literacy Behaviors
	Positive Attitudes Toward Reading
	Student Attitudes Toward Learning to Read

PIRLS results. Table 12 presents the top 10 and bottom 10 performing systems from the 2011 administration of PIRLS. In 2011, the United States ranked among the top 10 performers and scored significantly higher than the PIRLS scale average.

Table 12.

PIRLS 2011 Scores

	Reading Scale Score
Top 10 Systems	
Hong Kong SAR	571
Russian Federation	568
Finland	568
Singapore	567
Northern Ireland	558
United States	556
Denmark	554
Croatia	553
Chinese Taipei	553
Ireland	552
Bottom 10 Systems	
Malta	477
Trinidad and Tobago	471
Azerbaijan	462
Iran, Islamic Rep. of	457
Colombia	448
United Arab Emirates	439
Saudi Arabia	430
Indonesia	428
Qatar	425
Oman	391

Source: Mullis, Martin, Foy, & Drucker (2012b).

Utility of PIRLS data. Research shows that examining PIRLS data to compare countries can be crucial for improving achievement or closing achievement gaps in a country. Similar to a study in 2004 that used PIRLS 2001 data, Tunmer et al. (2008) used PIRLS 2006 data to test the prediction that unless fundamental changes were made to New Zealand's literacy strategy, there would be no substantial reduction in the achievement gap between "good and poor" readers. Tunmer et al. found that no significant changes in reading achievement had occurred over the past five years. International benchmarks that were based on the type of questions students were able to answer showed that New Zealand had large proportions of students performing at the highest and lowest levels. The authors stated that the large gap in proficiency was due to the consistent discrepancy between high and low socioeconomic schools in the country (Tunmer et al., 2008).

The 2008 study by Tunmer et al. used two measures of literacy to assess learning contexts, both of which were measured on high, medium, and low categories. The Early Home Literacy Activities (EHLA) Index was based on parents' responses to questions regarding the frequency of literacy-related activities parents practiced with their children before the children started school. The Parents' Attitudes Toward Reading (PATR) Index was based on the degree to which parents agreed or disagreed with statements about reading (e.g., "I only read if I have to").

Overall, the results from the two measures were very similar. For each index, the percentage of New Zealand students in the high category was high compared to other countries, whereas the difference between students in the high and medium categories of the index was much larger than that of most of the other countries (Tunmer et al., 2008). The use of PIRLS data in this case demonstrated that countries can compare specific aspects of literacy to those of other countries, as well as examine gaps in measures of literacy based on context.

Tunmer et al. (2008) demonstrated a few of the many uses for PIRLS data. Their study used the data to examine trends in scores of one country over time to examine possible growth and to compare multiple countries in terms of single scores and individual constructs. Tunmer et al.'s study can inform policy at the country level by applying lessons learned from PIRLS results to close gaps related to reading literacy achievement. PIRLS gave researchers information regarding teaching practices, parents' attitudes, and early home literacy activities that were compared among countries that participated. These data revealed that although New Zealand ranked high in reading achievement, the large differences between the high and medium categories were significant and represented a substantial gap in reading skills. Because no changes were made between 2001 and 2006 to New Zealand's literacy strategy, no changes were seen.

Critique. Although the IEA presents PIRLS as a comprehensive and formative assessment, others have argued against the construction of the test and interpretation of PIRLS data. Hilton (2006) focused on PIRLS 2001 and its validity for indicating an increase in literacy attainment in England. In England, there was a dearth of evidence about whether curriculum standards were rising or falling, and assessing the validity of PIRLS data allowed the researcher to conduct such an examination. Hilton argued that PIRLS research was methodologically "weak," and therefore England using the research to rank itself third out of 25 in reading achievement showed low validity. She stated that the use of a single indicator in various contexts and experiences of the population creates the potential for cultural and linguistic bias in PIRLS. Consistent with findings presented earlier in this review, Hilton suggested that an increase in economy may have been the actual cause for the increase in England's reading attainment. This realization was based on the existence of a causal relationship between socioeconomic status and reading attainment indicators found in PIRLS 2001. Table 13 compares 2002 wealth data for the top 10 and bottom 10 countries in reading achievement to PIRLS 2001 average reading scores. From the table, a general trend is evident that as wealth increased, so did PIRLS reading achievement scores. Further discussion of this topic can be found in the "International Assessments: Economic Value" section of this review.

… the use of a single indicator in various contexts and experiences of the population creates the potential for cultural and linguistic bias ….

Table 13.

Contrasting Wealth and 2001 PIRLS Reading Scores of Participating Nations

	US$ GDP per capital	PIRLS reading score
Top 10 Nations*		
Sweden	26,125	561
Netherlands	26,538	554
England	29,400	553
Bulgaria	2,037	550
Latvia	3,516	545
Canada (O&Q)	23,395	544
Lithuania	5,313	543
Hungary	6,450	543
United States	36,184	542
Italy	20,664	541
Bottom 10 Nations**		
Cyprus	13,134	494
Moldova Republic	353	492
Turkey	2,904	449
Macedonia	1,831	442
Colombia	1,887	422
Argentina	2,240	420
Iran	7,166	414
Kuwait	15,764	396
Morocco	1,324	350
Belize	3,324	327

Note: Adapted from Hilton (2006). Based on world records for 2002 GDP rankings — current exchange rate method; countrywatch.com.
*Average GDP per capita US $17,553
**Average GDP per capita US $5,011

Hilton (2006) referred to the details of the PIRLS methodology and stated that:

> Although the PIRLS researchers went to considerable trouble to make comparability as culturally fair as possible through the design of the test items and the careful piloting of the items in different countries, the methodology, based on what appear to be sound psychometric rules, by its nature ignores deep cultural differences both between nations and between different groups in each nation. (p. 822)

Hilton also pointed to cultural aspects of countries that may limit the interpretation of PIRLS scores. For example, comprehending the sentence "Stephanie likes to play soccer with Tim and go to ballet with Tiffany" requires having cultural understandings of what soccer and ballet are, and these hobbies are not shared by children worldwide. Therefore, this question may be easily understood by some students and not others. In addition, the test scores did not control for economic, cultural, or linguistic data that represent the culture and educational experience of students. According to Hilton, this factor contributed to weak cultural validity of the PIRLS international assessment. To fix this problem, PIRLS began to administer the questionnaires to students, teachers, parents, and principals in an attempt to understand cultural factors that might be controlled for, but this information was variable and did not necessarily underlay differential success (Hilton, 2006).

Hilton (2006) also criticized the sampling method of PIRLS. She said that the U.S. is similar to Russia, in that both countries consist of a large variety of school systems with ethnic and linguistic minorities, and therefore comparing them to countries such as Belize, a country with a small population, can be very misleading. In addition, comparing countries with a different number of languages spoken across the country can be misleading. As Hilton stated, it is almost impossible to account for the large number of students in other nations who speak different languages at home and at school. More important, the method of creating one test and then translating it into several languages is faulty, as going from one language to another is not mere language translation; rather, it requires the knowledge of certain embedded cultural meanings in the language (Hilton, 2006). Hilton argued that PIRLS is in fact not a valid measure of reading attainment because of the presence of cultural, linguistic, and economic bias. Thus, test results of a single measure may be best interpreted with caution. Comparison of scores across nations can be misleading, as nations have extremely variable population sizes, native and second languages, and cultural values. These arguments presented by Hilton may be true for any international comparative assessment, and further support the argument that considering cultural and geographic characteristics of countries is crucial for interpreting comparative data. This is especially important for readers and interpreters to note when looking for trends in the data. However, Hilton's study adequately demonstrated using PIRLS data to explore various implications and trends in achievement. Despite Hilton's warnings, PIRLS data are still useful for comparing reading achievement among various countries and learning contexts, as demonstrated by Tunmer et al. The ideas raised by Hilton can merely provide caution for interpreting results of any assessment.

Comparison of scores across nations can be misleading, as nations have extremely variable population sizes, native and second languages, and cultural values.

TIMSS and PIRLS Aligned

In 2011, PIRLS and TIMSS aligned their cycles to allow for a comprehensive reading, mathematics, and science assessment of fourth-graders, as well as to collect a variety of contextual background information, to allow for an in-depth examination of school environments, instructional resources, and teaching strategies. TIMSS and PIRLS are both coordinated by the International Study Center at Boston College, and international reports for both the 2011 TIMSS and PIRLS assessments were released in December 2012 (Mullis et al., 2009b; IEA, 2011).

CIVED

Purpose. The IEA's Civic Education Study (CIVED) covered the content domains of democracy and citizenship, national identity, and social cohesion and diversity (IEA, 2011). According to the IEA, the study was carried out in two phases. The first phase involved conducting case studies to examine the context and meaning of civic education in several countries, followed by a second phase that consisted of developing instruments based on the case studies to gather information about civic knowledge, attitudes, and engagement of students. The CIVED assessment contained items that measure the following in students: knowledge of fundamental principles of democracy; skills in interpreting political communication; knowledge of concepts of democracy and citizenship; attitudes related to students' nations, trust in institutions, opportunities for immigrants, and the political rights of women; and expectations for future participation in civic-related activities. In addition, students, teachers, and principals completed questionnaires about the learning contexts (IEA, 2011).

Participants. Fewer countries participated in CIVED than participated in PISA, TIMSS, and PIRLS. Twenty-four countries participated in Phase 1 of CIVED, and 28 countries participated in Phase 2; the U.S. participated in both (IEA, 2011). Phase 1 targeted mostly full-time eighth-grade students (or the grade that had the most 14-year-old students), and an optional survey was conducted in some countries for upper-secondary students ages 16.6 to 19.4. According to Baldi, Perie, Skidmore, Greenberg, and Hahn (2001), the participating countries of Phase 2 included countries with a tradition of a democratic government and some that have experienced recent transitions. Table 14 lists the CIVED participants.

Table 14.

CIVED Participating Education Systems

Phases 1 & 2		Phase 1 Only	Phase 2 Only
Australia	Hungary	Canada	Chile
Belgium	Italy	Netherlands	Denmark
Bulgaria	Lithuania		Estonia
Colombia	Poland		Latvia
Cyprus	Portugal		Norway
Czech Republic	Romania		Slovak Republic
England	Russian Federation		Sweden
Finland	Slovenia		
Germany	Switzerland		
Greece	United States		
Hong Kong SAR			

Source: IEA (2011). Note: Upper-secondary students from Israel also participated.

Methods. The first phase of the CIVED was conducted in 1996–1997, and data for the second phase were collected in 1999 for the standard population and in 2000 for upper-secondary students (IEA, 2011). The assessment items in Phase 2 were designed to measure knowledge and understanding of key principles that are universal across all countries (Baldi et al., 2001). Civic knowledge (civic content and civic skills) was measured with 38 multiple-choice cognitive items, which used just under half of the entire test time. CIVED also included student, teacher, and principal questionnaires, which captured information similar to the information captured by the questionnaires for the other international assessments, while being specific to civic knowledge and attitudes (Baldi et al., 2001).

CIVED results. Concerning students' civic knowledge and understanding, the CIVED researchers found that the high-performing group of 14-year-old students lived in countries with long-standing democracies or countries that were building democracy and experiencing massive political transitions in the 1990s (IEA, 2011). Students in Poland performed the best, followed by Finland, Cyprus, Greece, Hong Kong SAR, and the U.S. Overall, most students had an adequate understanding of fundamental democratic values and institutions. As may be expected, older students (upper-secondary students) had higher levels of civic knowledge than did 14-year-olds. Males performed better than females, particularly in the area of economic knowledge. Among 14-year-old students, there were minimal gender differences with regard to civic knowledge, but there were substantial differences with regard to attitudes. For example, females were found to be more supportive of women's political rights and immigrants' political rights than were males. Regarding students' attitudes, the IEA (2011) presented results that indicated that students were skeptical about some traditional forms of political engagement (one exception being voting). The

CIVED researchers also found that older students felt less positive about their countries than did younger students (IEA, 2011).

Lastly, and concerning the impact of school and home environment on performance, the IEA (2011) stated that upper-secondary students felt more comfortable expressing ideas and opinions while in the classroom. In addition, older students felt particularly strong about the idea that participating in student government and other similar activities provided a positive solution to problems in school. Upper-secondary female students tended to be more engaging and comfortable than males within the school and community. Most 14-year-olds reported television as being their most frequent news source, which was found by the CIVED researchers to be positively associated with students' level of civic knowledge and intention to vote. Among upper-secondary students, using television as a news source was also significantly and positively associated with students' intention to vote. Schools that modeled democratic practice also effectively promoted civic knowledge and engagement (IEA, 2011).

Utility of CIVED data. CIVED data have been used in a variety of analytical contexts, both nationally and internationally. Amadeo, Torney-Purta, Lehmann, Husfeldt, and Nikolova (2002) discussed and demonstrated how comparing data across countries can contribute to the "educational debate." Cross-country comparisons can highlight similarities and differences among students in various countries. These comparisons can also allow for comparing and contrasting practices, policies, and goals of different countries. In particular, Amadeo et al.'s study aimed to understand how students were involved in their countries politically, both in and out of school. The researchers highlighted some important elements involved in being part of a democracy; tolerance, willingness to participate, and understanding responsibilities are just as important as civic knowledge (Amadeo et al., 2002).

Other studies focused on policies and practices within the U.S. Baldi et al. (2001) presented the results from the national CIVED analyses, and demonstrated the use of CIVED data for the purpose of further understanding civic knowledge and attitudes among students, as well as informing educators, policymakers, and parents of the status of civic education. They particularly highlighted how the U.S. compared to the other 27 countries that participated in Phase 2 of CIVED (Baldi et al., 2001). This comparison could inform readers about the policies of other countries that could be applied in the U.S. to improve civic knowledge, or vice versa.

CIVED in the United States. Baldi et al. (2001) began their review by presenting results concerning the civic achievement of students, particularly U.S. students relative to those of other countries. Table 15 presents the average CIVED assessment scores for the top 10 and bottom 10 performing countries. U.S. students performed statistically significantly better than the international mean (100), and no other country scored statistically significantly higher than the U.S.

Table 15.

Average Civic Knowledge Achievement (CIVED Scores), by Nation

Nation	Average CIVED Score
Top 10 Nations	
Poland	111
Finland	109
Cyprus	108
Greece	108
Hong Kong SAR	107
United States	106
Italy	105
Slovak Republic	105
Norway	103
Czech Republic	103
Bottom 10 Nations	
Switzerland	98
Bulgaria	98
Portugal	96
Belgium (French)	95
Estonia	94
Lithuania	94
Romania	92
Latvia	92
Chile	88
Colombia	86

Note: Adapted from Baldi et al. (2001).

Similarly, U.S. students scored significantly higher than the international mean on the civic skills subscale — higher, in fact, than every other participating country — but did not significantly differ from the international mean on the civic content subscale (Baldi et al., 2001).

Baldi et al. (2001) used CIVED data to examine civic knowledge in the context of the school and classroom. The authors presented descriptive information on school environment, such as how civic subjects were studied and the views of school personnel regarding civic education. In addition, relationships were examined between school and classroom characteristics and CIVED civic achievement scores. At the time of the CIVED, 70% of U.S. schools with ninth-graders had civic-related subject requirements. Similarly, 55% of U.S. schools required students to take five or six periods of civic-related subjects per week, while only 19.6% required less than one period. Regarding attitudes of U.S. principals, 95% agreed that civic content should be integrated into human and social science subject content, while 78% agreed it should be integrated into all subject content. The majority (64%) of U.S. principals reported agreeing that civic education should be its own course, while 29% felt it should just be an extracurricular activity. Lastly, schools with a lower percentage of free and reduced-price lunch programs had higher civic achievement scores. Table 16 presents CIVED scale scores by a variety of school characteristics (Baldi et al., 2001). Based on the data provided in the table, class size and school size do not appear to be related to civic content knowledge in the U.S.

Table 16.

Ninth-Grade U.S. Students' Average CIVED Achievement Scale Scores, by School Characteristics

	Total Civic Knowledge	Civic Content	Civic Skills
Total	106.5	101.9	113.6
Civic-Related Subject Required			
Yes	108.2	103.6	114.9
No	104.0	99.4	111.9
School Participation in Civic Education–Related Programs			
Yes	105.9	101.3	113.3
No	103.8	100.0	110.2
School Type			
Public	106.1	101.6	113.1
Private	109.9	104.7	118.9
School Size			
500 or less	101.3	97.6	108.2
501–1,000	110.7	105.8	117.2
1,001–1,500	109.2	104.7	115.2
1,501–2,000	109.0	104.2	115.2
More than 2,000	104.5	99.4	113.1
Percent of Students Eligible for Free or Reduced-Price Lunch			
1st Quartile (0–13)	111.8	106.6	119.0
2nd Quartile (14–25)	110.7	106.0	116.5
3rd Quartile (26–48)	100.8	96.1	110.2
4th Quartile (49–100)	95.5	92.2	103.0
Class Size			
20 or less	102.8	97.9	112.1
21–25	109.6	105.1	115.6
26–29	106.8	102.2	113.5
More than 29	102.2	97.9	109.8

Note: Adapted from Baldi et al. (2001).

In addition to examining instructional variables, Baldi et al. (2001) presented the results concerning the impact of demographic, socioeconomic, and out-of-school variables that were previously shown to be related to civic knowledge of U.S. students. The researchers reported that white and multiracial students scored higher than black and Hispanic students on all three scales. Asian students scored higher than black students on all three scales as well, although Asian students did not score higher than Hispanic students on the content subscale. Female students performed better than male students on the skills subscale. CIVED assessment scores were positively related to the number of books in a student's home, whether students received a newspaper, parents' educational attainment, and having higher expectations for continued education. Additionally, the following student characteristics were associated with higher scores: being born in the U.S., having had fewer absences during the month before the assessment, and participating in extracurricular activities or any other organization. Table 17 presents civic achievement scores according to various demographic, socioeconomic, and out-of-school factors.

Table 17.

Ninth-Grade U.S. Students' Average Overall CIVED Achievement Scores by Various Demographic, Socioeconomic, and Out-of-School Contexts

Factors	Total Civic Knowledge	Factors	Total Civic Knowledge
Sex		**Frequency of English Spoken in the Home**	
Male	105.6	Sometimes	96.2
Female	107.5	Always or almost always	108.0
Race/Ethnicity		**Number of Books in the Home**	
White	111.6	0–10	90.7
Black	92.7	11–50	99.0
Hispanic	97.1	51–100	104.9
Asian	109.4	101–200	111.5
Multiracial	109.1	More than 200	115.3
Country of Birth		**Receives a Daily Newspaper**	
U.S.	107.6	Yes	109.7
Foreign born	97.9	No	102.5
Region		**Frequency of Participation in Organized Extracurricular Activities**	
Northeast	109.7	Never or almost never	98.6
Southeast	102.7	A few times each month	108.0
Central	109.3	Several days a week	109.2
West	104.2	Almost every day	109.2
Frequency of Changing Schools in Past 2 Years as a Result of Moving		**Number of Parents in the Home**	
Never	108.8	Two	109.2
Once	102.5	One	99.3
Twice or more	99.4	None	96.1
Parents' Highest Level of Education		**Expected Years of Further Education**	
Elementary or less	91.0	0–2	89.0
Some high school	94.5	3–4	91.3
Finish high school	101.4	5–6	98.5
Some vocational/ technical education	107.4	7–8	110.5
Some college	108.9	8–10	117.0
Completed a bachelor's	118.7	More than 11	113.2
Number of Days Absent from School Last Month		**Time Spent Each Day on Homework**	
0	109.2	Not assigned	95.9
1–2	107.3	Doesn't complete	97.1
3–4	100.5	30 min. or less	102.7
5–9	100.0	1 hour	106.7
More than 10	93.2	More than 1 hour	111.9

Note: Adapted from Baldi et al. (2001).

The results Baldi et al. (2001) presented in their study can be useful for education decision-makers, as well as for parents. Education policymakers can consider those aspects of the curriculum and school environment that are associated with high performance on the CIVED assessments to incorporate into their current policy. Parents can consider out-of-school factors that are associated with high performance to create an environment that promotes the highest

achievement. The authors compared the U.S. with the rest of the CIVED participating countries to create a context for associations found within the U.S. These practices could also be applied in countries around the world to promote the highest civic knowledge for students, and to promote active citizens within democratic systems. However, based on the opinions of researchers reviewed in this paper, researchers and policymakers should consider cultural, economic, and geographic characteristics before applying specific practices of one country to another.

Within-country examination of CIVED data. CIVED data can be used to explore gaps between ethnic groups related to academic and political outcomes specific to one country, in addition to providing suggestions for education policy. Torney-Purta, Barber, and Wilkenfeld (2007) examined factors associated with the gaps between Latino and non-Latino students in the U.S., and presented possible explanations on individual and school levels. The researchers also provided implications for education policy and alternative ways to use the CIVED data set. After controlling for language, country of birth, and "political discussions with parents," Torney-Purta et al. found that Latino students had lower civic knowledge scores than non-Latino students, and Latino students reported lower ratings for "perceiving an open classroom climate" and "studying political topics." Although the factors of "discussing politics with parents," "reading the newspaper," "studying political topics" in the classroom, and "experiencing an open classroom climate" positively and significantly related to higher civic knowledge scores, this did not explain why non-Latino students scored higher. Some school factors were found to partially explain the gap between the performance of Latino students and that of non-Latino students: an open classroom climate; time devoted to studying democratic ideals; and time devoted to studying political topics (Torney-Purta et al., 2007).

Overall, Torney-Purta et al. (2007) demonstrated the use of CIVED study data for the purpose of comparing specific ethnic groups at both individual and school levels to examine differences in civic knowledge, perceived expectations in a democratic system, and attitudes. The findings can contribute to suggestions for education policy, and also demonstrate how CIVED data can be used to compare measures in more than one context. Because many school-related characteristics were able to predict the outcomes of civic knowledge, education policy may be encouraged to make use of interactive classroom activities, maintain an open climate for discussion, and include political topics in study.

Critique. Despite the beneficial uses of CIVED data that Torney-Purta et al. (2007) demonstrated, the authors also discussed weaknesses in the data set. For example, the student questionnaires did not inquire about the immigration status of their parents. In addition, ninth-grade students were found to have difficulty accurately reporting the educational level of their parents, so this variable could only be analyzed at a school level. Analyzing this factor at the individual level could allow for further understanding of the effects of socioeconomic status on Latino student development (Torney-Purta et al., 2007).

Baldi et al. (2001) also presented some limitations to the CIVED data, the major one being that the assessment items were not tied to the school curricula of the respective nations. Rather, the questions exclusively covered concepts that were vital to democracies worldwide and may exclude key knowledge relative to particular countries. In addition, the CIVED scales regarding student attitudes did not have identical means and standard deviations, so results could not be compared across item scales. Baldi et al. used the example that the mean score on the "trust in government-related institutions" scale cannot be compared with the mean score on the "positive attitude toward one's nation" scale. The scales had no common items; hence, the comparisons are not meaningful. The authors also warned the reader against using the results to make causal inferences. Some of the differences in scores could be contributed to by other factors that are not included in CIVED. Lastly, Baldi et al. cautioned that when

interpreting results, one should note that the students were tested in October 1999, close to the beginning of the school year. Associations involving school and classroom factors may not have been applicable to the aspects of schooling of the current year, or may only have been applicable to the short time students have spent in that school year (Baldi et al., 2001).

Despite these limitations in the data set, many of which were common across studies, CIVED results have been used in a variety of contexts and are not discounted as a good measure of civic knowledge and attitudes. Overall, the Civic Education Study can positively contribute to the world's understanding of the civic knowledge students hold in addition to the most beneficial environments to promote the learning of this knowledge. Although mathematics and science test scores showed to be important indicators of how well one country is doing in comparison to others, they are not exclusive in contributing to educational achievement. Assessing a combination of skill sets can provide a more comprehensive assessment for countries.

Additional International Assessments

The international assessments discussed thus far comprise the more well-known and most often cited in comparison to what will be discussed in this section. The IEA (2011) is author to three international assessments in addition to TIMSS, PIRLS, and CIVED. These assessments, summarized below, are worth noting as they play a role in comparing student performance internationally.

- The International Civic and Citizenship Education Study (ICCS) was first conducted in 1971, followed by administrations in 1999 and 2009. It assessed student achievement in civics and citizenship related to knowledge, conceptual understanding, and competencies. The study provides information about contexts for learning about civics and citizenship, specifically the school and classroom climates, as well as factors associated with high performance in civics and citizenship. Three different modules of the assessment were created according to issues specific to regions for Asia, Europe, and Latin America.

- The International Computer and Information Literacy Study (ICILS) will be conducted in 2013 to examine outcomes related to student computer and information literacy (CIL) of various countries. According to the IEA (2011), "CIL refers to an individual's ability to use computers to investigate, create, and communicate in order to participate effectively at home, at school, in the workplace, and in the community." The study looks at how CIL varies both within and between countries, examines factors that influence CIL, and provides suggestions for education systems and schools to improve CIL among students based on the data. Questionnaires will also be administered to students, teachers, and school administrators to gather information about the attitudes and background characteristics of students, classroom practices, and the use of computers and technology within the schools.

- The Teacher Education and Development Study in Mathematics (TEDS-M) compared how different countries prepare primary and secondary mathematics teachers for teaching. Data were collected in 2007 and 2008. The assessment was administered to teacher education institutions, education professors, and future teachers. The study examined the national policy context, salient characteristics of mathematics teacher education programs, and the level of knowledge of both mathematics and teaching acquired by teachers in training. A linked study looked at the relationship between the salaries of

mathematics teachers and the performance of their students on international mathematics tests. The following countries participated in the TEDS-M: Botswana, Canada, Chile, Chinese Taipei, Georgia, Germany, Malaysia, Norway, Oman, Philippines, Poland, Russian Federation, Singapore, Spain, Switzerland, Thailand, and the United States.

National Assessments

The U.S. education systems differs from those of other countries in terms of standardized assessments and college entrance examinations. While the U.S. has privately funded organizations creating and administering college entrance exams, such as the SAT® or ACT, and each state uses a different standardized assessment to examine achievement within schools and states, other countries have one exam administered by a single organization or ministry of education. Karp (n.d.) compared the Baccalauréat used in France to the A-levels used in the United Kingdom. Both tests are used for students to obtain a standardized qualification at the end of high school. The major difference is that A-levels are attained in single subjects, and the Baccalauréat is one nationally recognized qualification. Students in the U.K. complete various A-level subject exams according to their interests and university requirements. The Baccalauréat is one examination that encompasses the core subjects of French, philosophy, mathematics, and two foreign languages. Like the SAT, if a student does not perform adequately on even one subject of the Baccalauréat, the student must take the entire examination again, rather than just specific subjects, which is the case for the A-levels (Karp, n.d.). Exams similar to A-levels are used in Hong Kong (Hong Kong Examinations and Assessments Authority, 2010).

According to Finland's Matriculation Examination Board (n.d.), Finland uses one exam to determine whether students have obtained adequate knowledge required by secondary curriculum and for universities to determine whether students are qualified to attend the institution. Similar exams are administered for the same purposes to students in Germany and Estonia. These countries' assessments differ from those in the U.S. in that the same test is administered to all students for the same purpose, while the U.S. does not have one common exam for high school completion or college entrance. Because each state in the U.S. has its own education system, the U.S. looks at student achievement at the state, district, and school levels. This approach allows for a comparison of achievement across states, schools, and districts, in addition to examining how the nation is performing as a whole. Trends and progress can also be measured over time, and a comparison of how various states are progressing in relation to one another can elicit further inquiry into the practices and policies of states and districts that are associated with achievement growth. Hence, a national assessment that uses this methodology was established for the U.S. to measure growth trends and to determine common characteristics of high-performing districts and states.

Trends and progress can also be measured over time, and a comparison of how various states are progressing in relation to one another can elicit further inquiry into the practices and policies of states and districts that are associated with achievement growth.

NAEP

The National Assessment of Educational Progress (NAEP) is run by the National Center for Education Statistics (NCES), and is a nationally representative and continuing assessment of academic achievement for American students (NCES, n.d.a). The assessments cover the subject areas of mathematics, reading, science, writing, the arts, civics, economics, geography, and U.S. history, and are administered to students in grades 4, 8, and 12. In addition, long-term trends are measured by administering NAEP assessments in mathematics and reading to students at ages 9, 13, and 17. Because NAEP is uniformly administered across the U.S., the results of the assessments provide information about trends of student progress over time. NAEP provides results for students in all states as well as in large urban school districts. The assessment also provides results for various groups within the population regarding school environment, instructional experiences, and subject achievement, using Item Response Theory (IRT) models and estimating scale score distributions. NAEP is administered nationally for all subjects, but state- and district-level data are only available for public schools in the subjects of mathematics, reading, science, and writing (NCES, n.d.a)

The Nation's Report Card (NRC) (NCES, n.d.a) used NAEP scores to compare performance across various groups within the population and presented its findings online. The NRC website, nationsreportcard.gov, presents results for each 2011 subject assessment as well as for the 2008 long-term assessment. This review presents summaries of major findings for the 2011 NAEP assessments in mathematics, reading, science, and civics for the purpose of considering how policy making may differ based on national assessment scores versus international assessment scores.[7]

NAEP math. The NRC (NCES, n.d.b) provided summaries of findings for the 2011 NAEP mathematics assessment at the national, state, and district levels. The NAEP mathematics assessment measures students' knowledge of mathematical content and their ability to apply this knowledge to solve problems, and was most recently administered in 2011 and 2009. On the national and state levels, the 2011 assessment was administered to 209,000 fourth-graders from 8,500 schools, and 175,200 eighth-graders from 7,610 schools. On the district level, the assessment was administered to 21 districts nationwide. National results indicate that for both fourth- and eighth-grade students, the average score increased between 1990 and 2011, and more students' scores reached proficient or advanced levels in 2011 compared to previous years. Among fourth-grade students, Hispanic, white, and black students performed better in 2011 compared to 2009, and eighth-grade Hispanic students performed better in 2011 compared to 2009 (NCES, n.d.b).

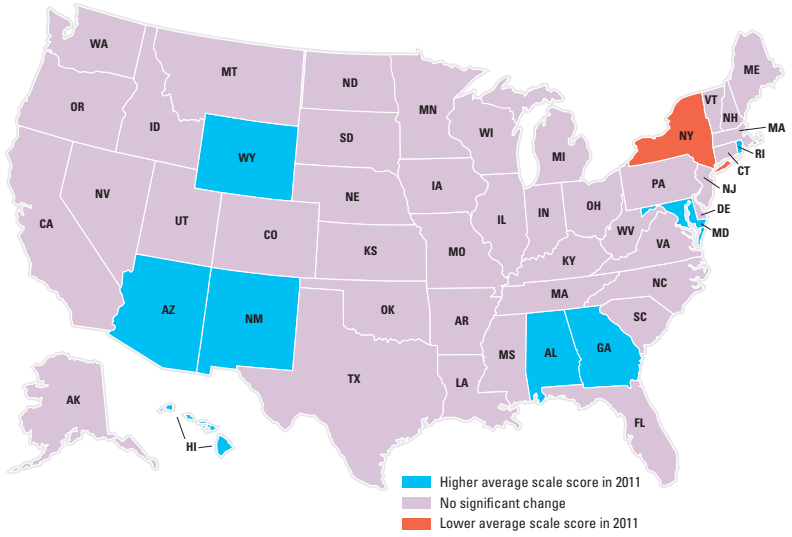
For the state level, the NRC (NCES, n.d.b) presented statistics for how both fourth- and eighth-grade students performed on the 2011 NAEP mathematics assessment. Figures 2 and 3 show how each state's average score changed from 2009 to 2011 for both fourth and eighth grades, respectively.[8] Among fourth-graders, the majority of states saw no significant differences, but lower performance was seen in New York, and higher performance was seen in Alabama, Arizona, the District of Columbia, Georgia, Hawaii, Maryland, New Mexico, Rhode Island, and Wyoming. Among eighth-graders, the majority of states saw no significant differences, but lower performance was seen in Missouri, and higher performance was seen in Arkansas, Colorado, the District of Columbia, Hawaii, Maine, Mississippi, Nevada, New Mexico, Ohio, Oklahoma, Rhode Island, Texas, and West Virginia (NCES, n.d.b).

7. See the "Suggestions for Using International Assessments" section.

8. Reminder: Students assessed on the national level come from both public and private schools, but those assessed on the state and district levels come from only public schools.

Figure 2.

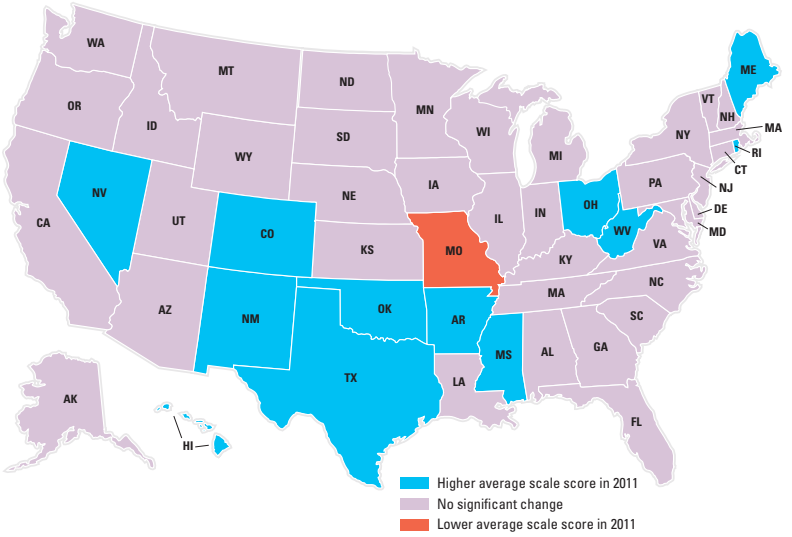
Change in average fourth-grade mathematics scores from 2009 to 2011.



Source: NRC (NCES, n.d.b)

Figure 3.

Change in average eighth-grade mathematics scores from 2009 to 2011.



Source: NRC (NCES, n.d.b)

The 21 public school districts that were assessed in 2011 were compared to 18 districts that were assessed in 2009. For both fourth- and eighth-grade students, only districts in Atlanta scored higher in 2011 than in 2009. For fourth-grade students, districts in Austin, Baltimore City, and Philadelphia scored higher in 2011 compared to 2009, and eighth-graders from Charlotte (NC), Chicago, Detroit, the District of Columbia, and Jefferson County (KY) scored higher in 2011 compared to 2009.[9] Notably, performance gaps between high- and low-income students remained between 2009 and 2011 for most school districts. Among fourth-grade students, gaps were smaller in Boston and Detroit than in large cities overall, and among eighth-grade students, gaps were smaller in Dallas, Detroit, Houston, Miami-Dade, and New York City than in large cities overall (NCES, n.d.b).

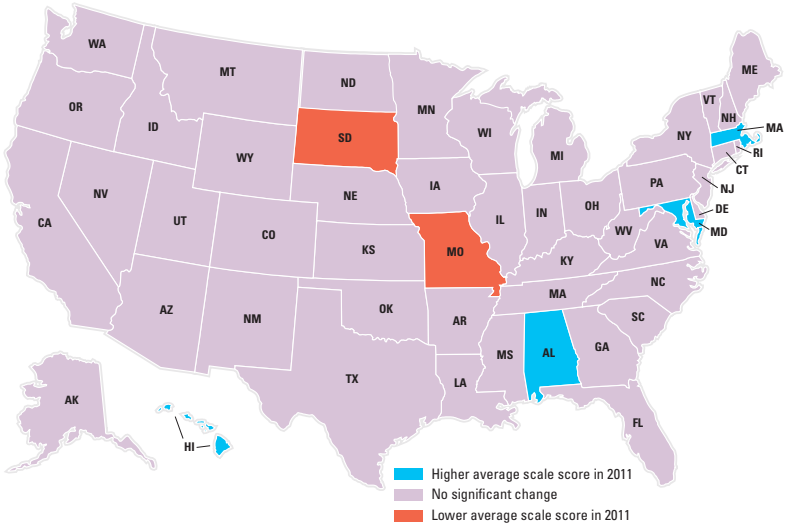
NAEP reading. The NAEP reading assessment measures students' ability and knowledge of reading both literary and informational texts. In 2011, 213,100 fourth-grade students from 8,540 public and private schools, and 168,200 eighth-grade students from 7,670 public and private schools nationwide participated in the assessment. For the district level, the 2011 NAEP reading assessment was administered to students from 21 districts, and scores for this year were compared to those of 18 districts that were tested in 2009. On the national level, performance did not change for fourth-grade students from 2009 to 2011, but improved for eighth-grade students. More eighth-grade students' scores fell at or above the proficient level in 2011 compared to 2009, and scores increased from 2009 to 2011 for white, black, and Hispanic eighth-grade students (NCES, n.d.b).

For the state level, the NRC (NCES, n.d.b) presented statistics for how both fourth- and eighth-grade students performed on the 2011 NAEP reading assessment. Figures 4 and 5 show how each state's average score changed from 2009 to 2011 for both fourth and eighth grades, respectively. For fourth-grade students, the majority of states did not show any significant differences; however, scores decreased in Missouri and South Dakota and increased in Alabama, Hawaii, Maryland, and Massachusetts. For eighth-grade students, the majority of states did not see any significant differences, and no states saw decreases in scores from 2009 to 2011. Increases were seen among eighth-grade scores in Colorado, Connecticut, Hawaii, Idaho, Maryland, Michigan, Montana, Nevada, North Carolina, and Rhode Island (NCES, n.d.b).

9. Reminder: Students assessed on the national level come from both public and private schools, but those assessed on the state and district levels come from only public schools.

Figure 4.

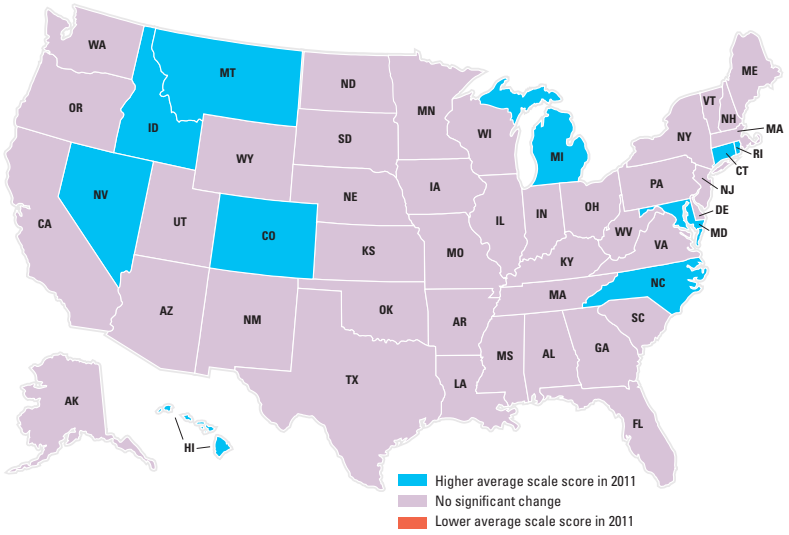
Change in average fourth-grade reading scores from 2009 to 2011.



Source: NRC (NCES, n.d.b.)

Figure 5.

Change in average eighth-grade reading scores from 2009 to 2011.



Source: NRC (NCES, n.d.b.)

When comparing the 21 school districts that participated in the 2011 NAEP reading assessment to the 18 districts that participated in 2009, the NRC (NCES, n.d.b) stated that only one district, Charlotte (NC), had higher scores in 2011; however, national average reading scores were higher. Scores were higher for both fourth- and eighth-grade students in Austin, Charlotte, Hillsborough County (FL), Jefferson County (KY), and Miami-Dade in 2011 than for large cities nationally. Scores were also higher than for large cities nationally for fourth-graders in Boston,

New York City, and San Diego school districts. Performance gaps between high- and low-income students remained from 2009 to 2011. Performance gaps among eighth-graders were smaller for Baltimore City and Miami-Dade than for large cities overall for fourth-graders, and gaps were smaller for Dallas, Detroit, Houston, and New York City than for large cities overall (NCES, n.d.b).

NAEP science. The NAEP 2011 science assessment was administered to 122,000 eighth-grade students from 7,290 schools, and measured content knowledge in physical science, life science, and Earth and space sciences, in addition to the practices of identifying science principles, using science principles, using scientific inquiry, and using technological design (NCES, n.d.b). The NRC presented results for eighth-grade students on the national and state levels. Nationally, students scored higher, and more students performed at or above basic and proficient levels in 2011 than in 2009. Performance gaps were smaller in 2011 between white and black students and white and Hispanic students than in 2009. Male students performed higher than females in both assessment years, and both genders increased scores overall. On the state level, the majority of states saw no significant differences in scores between 2009 and 2011; however, increases were seen in student scores from Arkansas, Colorado, Georgia, Hawaii, Maine, Maryland, Michigan, Mississippi, Nevada, North Carolina, Rhode Island, South Carolina, Utah, Virginia, West Virginia, and Wyoming (NCES, n.d.b).

NAEP civics. The NAEP Civics Assessment of 2010 was administered to students in grades 4, 8, and 12, and measured students' civic knowledge, intellectual and participatory skills, and civic dispositions. In 2010, the civics assessment was administered to 7,100 fourth-grade students from 540 schools, 9,600 eighth-grade students from 470 schools, and 9,900 12th-grade students from 460 schools. The NRC (NCES, n.d.b.) compared the results of the 2010 civics assessment to the results of the 1998 and 2006 assessments to examine how the civic knowledge and skills of students have changed over time. The main findings are as follows:

- The average civic score increased for fourth-graders from 1998 and 2006 to 2010, and decreased for 12th-graders from 2006 to 2010.
- A larger proportion of fourth-grade students scored at or above the proficient level in 2010 than in 2006 and 1998, and a smaller proportion of 12th-grade students scored at or above the proficient level in 2010 than in 2006.
- The proportions of students in each grade at the advanced level did not significantly change in 2010 compared to 2006 and 1996.
- Eighth-grade Hispanic students scored higher in 2010 than in 2006, and Hispanic students in all grades scored higher in 2010 than in 1998.
- No significant differences were found between male and female 12th-graders; however, the average score for female 12th-graders was lower in 2010 than in 2006 and 1998. No changes were seen for males.

The NCES coordinated NAEP scores with some international assessments to gain a better understanding of how U.S. students compare with the rest of the world. This is further discussed in the following section.

Suggestions for Using International Assessments

These recommendations, based on results from the various international assessments regarding how U.S. students perform in relation to students from other nations and on critiques and recommendations from researchers, may offer insight into how U.S. policy can change to promote performance improvements among students.

Examine Top Performers

Although some may believe that the U.S. is a leader in academics, the nation actually has ranked around the middle for achievement in reading, mathematics, and science. Of bigger concern is not the number of high-performing students in the U.S., as some U.S. states were among the top performers in the world, but rather that the U.S. had the second-highest number of low-performing students among OECD countries. The large variation in performance seen in the U.S. differed from that of high-performing countries (i.e., Korea, Finland, Hong Kong-China, and Shanghai-China) all of which showed the least variation in individual scores on PISA (OECD, 2010a). Despite the OECD's recommendation that disadvantaged schools should receive more funding, the districts in the U.S. continue to provide more funding to high-performing schools according to students' scores on standardized tests, and less funding for disadvantaged schools whose students score low on standardized tests. The OECD also recommended that more teachers be placed in disadvantaged schools. In the U.S., lower socioeconomic background of schools was associated with higher student-teacher ratio (OECD, 2010a; 2011). These schools had high proportions of students from low socioeconomic backgrounds. According to the OECD (2010a; 2011), the rest of the world does not show these socioeconomic performance gaps. As it relates to PISA scores, only 6% of the differences in average performance worldwide were attributed to GDP per capita, and in the U.S. alone, 17% of the variation in scores was attributed to socioeconomic differences (OECD, 2010a; 2011). Research suggests that in addition to adopting the policies recommended by the OECD, the U.S. should consider adopting common achievement standards as a way of conforming to policies of high-performing systems (Paine & Schleicher, 2011).

Looking Beyond Top-Performing Systems

U.S. policymakers should also consider looking at systems that have shown improvement in education, or countries that have similar geographical, economical, or cultural characteristics to the U.S. For example, the Russian Federation, like the U.S., has a large immigrant population. Although many Russian immigrants are native Russian speakers, while U.S. immigrants mostly are non-native English speakers (Matthews, 2009), the cultural implications may be similar. In addition, although some nations are not high-performing, reforms made

… only 6% of the differences in average performance worldwide were attributed to GDP per capita, and in the U.S. alone, 17% of the variation in scores was attributed to socioeconomic differences.

based on the results of international assessments have preceded positive growth. These ideas are discussed in the "International Assessments and Common Core in a Decentralized System" section of this paper.

Examine states. Although international assessments can provide valuable information for the U.S. regarding how it compares internationally on student achievement to the rest of the world, and considering adopting those policies could be beneficial, it may not be useful for the U.S. to consider policies of other nations, but rather those policies that have been successful within the U.S. The U.S. is home to some of the world's top performers in educational achievement. In addition, some of the policies of other systems may not be culturally relevant in the U.S. (Bracey, 2009; Biddle, 2012). U.S. policymakers may consider looking at what has been successful within the nation at the state level and adopt these practices nationwide. Unfortunately, without a common core curriculum, it can be difficult to assess how districts and states are performing relative to one another and whether there is improvement. Despite the U.S. lacking a standard curriculum and state assessments, NAEP can still be used to measure trends in growth on the national, state, and district levels, in addition to determining common practices of high-performing districts and states.

Paine and Schleicher (2011) stated that improving PISA performance among U.S. students will narrow the achievement gaps between the U.S. and high-performing countries, as well as improve the economy and GDP. Notably, higher educational achievement is possible for students in all regions of the U.S. Substantial gains have been seen in countries such as Canada, Poland, and South Korea, and even for areas within the U.S., such as Boston; California; Charlotte-Mecklenburg, NC; Long Beach; and Miami (Paine & Schleicher, 2011). In addition, Peterson et al. (2011) examined the PISA 2009 results, which indicated that Massachusetts alone is consistently one of the top 10 performing areas worldwide in both reading and mathematics. Examining specific policies and practices within Massachusetts schools can be extremely valuable to other U.S. states for improving scores on international assessments. These practices are likely to be more easily adopted than others from different countries, as the practices are more likely culturally relevant, and substantial improvement has been seen.

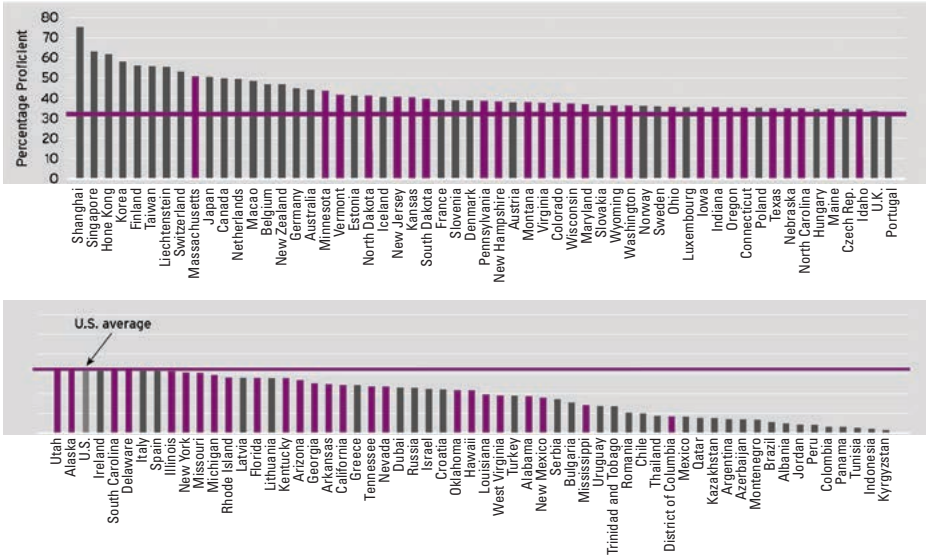
Linking NAEP with International Assessments

International assessments provide information about factors associated with high performance, and NAEP does the same between districts and states. Therefore, it seems that examining information collected by the collaboration of the two would be the most beneficial and have the highest probability of showing improvement. It may be beneficial to determine which successful practices and policies in other systems have been successful in the U.S. as well. To do this, NAEP scales can be linked to those of international assessments; this has been done in a large number of studies. As Fleischman, Hopstock, Pelczar, and Shelley (2010) stated, "While PISA and NAEP may appear to have substantial similarities, each test was designed to serve a different purpose, assesses different target populations, and are based on separate and unique frameworks and items. As such, PISA and NAEP provide different, and complementary, information about student performance" (p. 55). In this sense, analyzing results of both national and international assessments would provide additional beneficial information.

Results of linking studies. Peterson et al. (2011) demonstrated how NAEP and PISA administrators have collaborated to obtain the type of information that Fleischman et al. (2010) described, and highlighted similarities and differences between the assessments. Their study reported the percentage of both public and private school students in the U.S. that scored at or above the proficient level using PISA and NAEP scores. The researchers found that in Massachusetts — the top-performing U.S. state — slightly over half (51%) of students were proficient in mathematics. Only five additional states had over 40% proficiency: Kansas, Minnesota, New Jersey, North Dakota, and Vermont. As previously mentioned, some of the wealthiest U.S. states scored below average for the U.S. overall (e.g., California, Florida, Michigan, Missouri, and New York). To put these findings into perspective, Shanghai had a 75% math proficiency rate and Finland had a 56% proficiency rate. For reading, Massachusetts had a 43% proficiency rate, while Shanghai's rate was 55% and Finland's rate was 46%. Peterson et al. described a "crosswalk" that is necessary to provide these estimates. Administering the assessments in the same year to the class of 2011 at around ages 14 and 15 will best accomplish this. Peterson et al. (2011) stated, "Given that NAEP identified 32 percent of U.S. eighth-grade students as proficient in math, the PISA equivalent was estimated by calculating the minimum score reached by the top-performing 32 percent of U.S. students participating in the 2009 PISA test" (p. vii). Figure 6 shows the mathematics proficiency rates of various PISA participating systems, in addition to the coordinated NAEP scores for some high-performing U.S. states.

Figure 6.

Percentage of students in the class of 2011 at the proficient level in math in the U.S. states and foreign jurisdictions in PISA 2009.



Source: Peterson et al. (2011)

The study also compared performance gaps between the systems of the U.S. and other countries to those between U.S. states, and Peterson et al. (2011) discussed outcomes in terms of various demographic factors.

A study from the American Institutes for Research (Phillips, 2007) linked NAEP 2000 scales for mathematics and science to those of TIMSS for 1999 as well as 2003, using similar methods to Peterson et al. (2011) to calculate percentages of students who fall into each

proficiency level. Phillips (2007) also demonstrated how to project TIMSS international benchmarks for NAEP achievement in mathematics and science. Table 18 compares the international benchmarks with the projected achievement level for NAEP.

Table 18.

TIMSS International Benchmarks Compared to Projected NAEP Achievement Levels

TIMSS	TIMSS International Benchmarks	NAEP	Projected NAEP Achievement Level in Math	Projected NAEP Achievement Level in Science	Projected NAEP Achievement Level Minus TIMSS International Benchmark in Mathematics	Projected NAEP Achievement Level Minus TIMSS International Benchmark in Science
Advanced	625	Advanced	637	670	12	45
High	550	Proficient	556	567	6	17
Intermediate	475	Basic	469	494	-6	19
Low	400					

Source: Phillips (2007)

Phillips (2007) used the projected values to find percentages similar to those found by Peterson et al. (2011) regarding where the U.S. stands relative to other countries. For mathematics, more countries had percentages of students significantly above that of the U.S., but for science, more countries had percentages of students significantly below that of the U.S. (Phillips, 2007). Although this shows high performance of U.S. students in science, the U.S. is still not among the high-performing countries for reading and mathematics. More important, focusing on improving achievement for disadvantaged students and schools may improve the overall rankings of the U.S.

> . . . examining the results of both national and international assessments enables us to view performance and trends through a wider lens.

The studies presented here demonstrate ways to link NAEP scores to those of international assessments to provide further information than the assessments themselves for the purposes of policy reform. In addition, examining the results of both national and international assessments enables us to view performance and trends through a wider lens.

International Assessments: Economic Value

At the 2009 NAACP Centennial Convention, President Barack Obama declared: "a world-class education is a prerequisite for success" (Obama, 2009). Numerous indicators depict a positive correlation between economic success and educational achievement. First, researchers from the World Bank emphasized the existence of a direct link between the quality of education and individual earnings, especially in developing countries (Hanushek & Woessmann, 2007). In addition, Peterson et al. (2011) suggested that the U.S. could see economic gains by improving student performance in mathematics on international assessments. Specifically, Peterson et al. asserted that if U.S. students were to reach PISA levels of proficiency similar to those of students in Canada and South Korea, the annual U.S. growth rate would increase

by 0.9 percentage points and 1.3 percentage points, respectively. In equivalent U.S. dollars (US$), gains in the U.S. could reach $75 trillion over the next 80 years (Peterson et al., 2011).

While evidence for the existence of a relationship between educational achievement and economic growth and success seems strong, other researchers assert there is little to no link between the two. Tienken (2008) suggested that the correlation between education, specifically math scores, and economic growth is actually negative. In this regard, "the education system needs the economy more than the economy needs the education system" (Bils & Klenow, 1998, as cited in Tienken, 2008, p. 9). Tienken pointed to other factors that may determine economic growth, such as tax and trade policy, public housing, health policies, legal issues, market conditions, and government reliability.

Alternatively, Wolf (2004) denied the existence of a relationship between educational achievement and economic growth. Although she agreed that an economy could not function without educated people, and that innovative companies use university-based research, she said "[i]t does not follow that education policy is therefore an effective tool for ensuring economic prosperity, let alone that it can guarantee specific levels of growth or national income" (p. 315). Wolf pointed to reasons why countries with stronger economies have better education. Wealthier countries have more educated citizens, and educated people are paid more. These wealthy countries do not just have better education; they also have more motorways and hospitals. Countries with more money have more people who can afford higher levels of education. Clear education effects are lacking in the examination of rising economies (Wolf, 2004).

Wolf (2004) provided some counterexamples to support her argument against the existence of such a relationship. First, no relationships are seen between university enrollment rates and income per head among OECD countries. For example, Switzerland has not seen increases in enrollment rates; however, the nation continues to keep its position as the wealthiest of nonoil states. In addition, Robinson (1999) demonstrated that no correlation exists between individual student performance on international assessments and economic performance (as cited in Wolf, 2004, p. 322). Wolf also pointed to the idea that the relationship would not be one way. In a growing economy, citizens tend to get educated to compete, as education is less expensive. Wolf (2004) found that "Growth generates education, whether or not education generates growth" (p. 323). When jobs that require more educated workers become in higher demand because of a growing economy, more workers must become educated to fill the positions.

Further, the OECD (2010a) found that countries with similar economies can perform very differently. The researchers found a correlation between GDP per capita and educational performance, although this measure only predicted 6% of the differences in average performance. This finding suggests that factors other than GDP influence educational achievement (OECD, 2010a).

However, in addition to the findings from PISA 2009 that showed that students who attend schools with more socioeconomically advantaged students tended to perform better, and that students within socioeconomically disadvantaged schools had larger student-teacher ratios (OECD, 2010a), in a publication highlighting the relationships between PISA score improvement and economic growth, the OECD (2010b) asserted that boosting scores on PISA even modestly could dramatically improve a country's GDP. The researchers stated that if each OECD country were to raise their average PISA score by 25 points over the next 20 years, aggregated GDP gains of US$115 trillion could be seen over the lifetime of the generation born in 2010. Poland achieved this gain between 2000 and 2006 by raising its average PISA score in reading by 29 points. Further, the researchers attested that aggregated GDP gains of up to

Conflicting arguments exist in the debate over education contributing to the growth of an economy.

US$200 trillion could be seen if all students were to raise their scores to a minimum level of proficiency (score of 400 for PISA). The OECD (2010b) finally concluded that an increased emphasis on cognitive skills in school would promote higher PISA scores, which, in turn, could increase a nation's GDP.

Conflicting arguments exist in the debate over education contributing to the growth of an economy. Although gains in education have been seen along with economic gains, one cannot conclude that the economy improved as a result of an improvement in educational achievement. Using PISA scores from 2003, 2006, and 2009 (OECD, 2004; OECD, 2007; Fleischman et al., 2011) and economic data for these years (World Bank Group, 2012), figures were constructed to determine whether a relationship in either direction may exist. PISA countries/economies with the 10 top economic gains between 2003 and 2009 include: Hong Kong-China, Indonesia, Macao-China, Poland, Russian Federation, Slovak Republic, Thailand, Tunisia, Turkey, and Uruguay. Figure A1 shows how countries'/economies' performance in reading have changed from 2003 to 2009, while indicating the countries/economies with the top economic growth rates. In addition, trend lines for the countries/economies with the top 10 economic growth rates are highlighted. Figures A2 and A3 show the same figures for math and science, respectively.

As seen in Figure A1, not all countries/economies with the top 10 economic gains showed improvements in reading; however, the majority did. Hong Kong and Macao improved from 2003 to 2006; however, declines were seen after 2006 for both regions, suggesting that if a relationship were to exist, the entire nation of China could have been affected. The Slovak Republic and Uruguay saw improvements in reading scores only from 2006 to 2009. The Russian Federation, Indonesia, and Tunisia saw improvements in both 2006 and 2009.

Fewer countries/economies with the top economic gains saw improvement in math than in reading (Figure A2). Hong Kong and the Slovak Republic saw improvements in math scores from 2006 to 2009. Indonesia and the Russian Federation improved math scores between 2003 and 2006, but scores dropped in 2009. Poland and Uruguay increased scores from 2003 to 2006 and performance did not change in 2009. Turkey and Brazil saw improvements in math scores in both 2006 and 2009. As seen in Figure A3, while Turkey and Thailand saw improvement in science scores in 2009, Hong Kong and Tunisia were the only countries with top economic gains to see improvements in both 2006 and 2009.

According to the figures, although most countries/economies showed some improvements in PISA subtest scores over certain years, the majority of countries/economies with top economic gains did not see consistent improvement. No trends appear to be consistent across PISA subtests or countries/economies. Thus, more evidence is necessary to determine whether a relationship exists between educational achievement and economic growth in either direction. One may also consider that the 15-year-old students who produced scores for these years will not affect the economy until joining the workforce in future years. This suggests to policymakers and researchers that policy action may not show its effects for many years.

Summary: U.S. and State Performance on a Global Level

The United States has typically ranked near the middle among nations that participate in international assessments.

The United States has typically ranked near the middle among nations that participate in international assessments. Of the 34 OECD countries that participated in the 2009 administration of PISA, the U.S. ranked 14th in reading, 25th in math, and 17th in science; however, of all 65 participating systems, the U.S. graduating class of 2011 ranked 32nd overall on average (OECD, 2010c). No significant changes were seen in U.S. student performance averages between 2006 and 2009. Top-performing systems for the 2009 administration included Canada, China, Finland, Hong Kong, Korea, Shanghai, and Singapore. According to Peterson et al. (2011), only 32% of U.S. students reached the proficient level in mathematics, compared to 75% of students in Shanghai, 56% of students in Finland, and 45% of students in Germany.

The U.S. performed slightly better on TIMSS in 2011 than on PISA in 2009 for math and science.[10] U.S. fourth- and eighth-grade students scored significantly higher than average on TIMSS 2011. Among 52 participating countries and seven benchmark participants, fourth-grade U.S. students ranked 11th in math and seventh in science. Similarly, among 45 countries and 14 benchmarking participants, U.S. eighth-grade students ranked ninth in math and 10th in science. Top performers in both subjects included Chinese Taipei, Hong Kong, Japan, Korea, and Singapore (Mullis et al., 2012a; Martin et al., 2012).

Forty-five countries participated in the 2011 administration of PIRLS. On average, the U.S. ranked sixth on PIRLS 2011, and U.S. fourth-grade achievement in reading increased significantly from 2006. Top-performing systems on PIRLS 2011 included Finland, Hong Kong, Northern Ireland, Russian Federation, and Singapore (Mullis et al., 2012b).

State Performance

Although PISA, TIMSS, and PIRLS are different assessments that measure different subsets of populations and have different structures, one thing from the studies is apparent: The U.S. has consistently fallen around the middle of participating countries and jurisdictions. However, there are some states that stand far above the national average and are even among the ranks of the highest-performing systems. As previously mentioned, Massachusetts has consistently seen its students rank among the top performers on national and international assessments, having ranked first among U.S. states on PISA 2009, with 50% of its students scoring at the proficient level in math and 43% scoring at that level in reading (Peterson et al., 2011). Compared to other nations that participated in PISA 2009, Massachusetts was found to have the fifth-highest percentage of students scoring at the proficient level in reading, and had the ninth-highest percentage of students scoring at that level in math.

Massachusetts also performed to world-class standards on TIMSS and PIRLS for math and reading. According to Phillips (2010), fourth- and eighth-grade students in Massachusetts scored among students from the top-performing systems on TIMSS 2007 in math. The highest-achieving fourth-graders in math came from Japan, Hong Kong, Singapore, and

10. Among OECD nations.

Taiwan, while the highest-achieving eighth-grade students came from Hong Kong, Japan, Singapore, and South Korea; Massachusetts students' scores were comparable to those of these students. Fourth-graders from Massachusetts scored among the highest-achieving fourth-graders in the world in reading. On PIRLS 2006, Massachusetts would have ranked among Canada, Hong Kong, Hungary, Italy, Luxembourg, the Russian Federation, and Singapore, all of which were the highest-achieving nations on this assessment (Phillips, 2010).

International Assessments and Common Core in a Decentralized System

This section will review countries, states, and cities that have demonstrated high performance on international assessments, are of similar size to the U.S., or have made reforms to their education systems based on results of international assessments. Some countries or regions fall into more than one of these categories. Table A1 illustrates each country in terms of size, population, economy, and rankings on international assessments. In addition, each country will be discussed in terms of what it offers to promote education and high academic achievement. Lastly, we discuss additional aspects of education that have been shown to affect international assessment scores. The section begins with discussions about Finland, Japan, Singapore, and Hong Kong, all top performers on international assessments. Some of these nations have also demonstrated successful reforms based on results of international assessments. Only a few characteristics of and reforms made to the education systems of these nations will be discussed here, as many are beyond the scope of this review. Future research might choose to focus on specific aspects of education systems.

Top-Performing Nations

Finland. Finland is a top-performing nation, having ranked third in reading, sixth in math, and second in science on PISA 2009 (Fleischman et al., 2010). The teaching profession in Finland is the most desired career choice among students and is a highly competitive field to enter (Paine & Schleicher, 2011). According to researchers, only one out of every 10 applicants is accepted into teacher training programs. The OECD (2011) explained some of the attractive attributes of the teaching profession in Finland. First, once certified, teachers in Finland receive an amount of respect and trust comparable to that of physicians in the U.S., and are also given more autonomy than teachers in other nations. Teachers are given the freedom to create lesson plans, curricula, and assessments, and parents generally trust their decisions, rather than challenge them. In addition, no national assessments are administered to evaluate the level of knowledge students have attained from teachers, further demonstrating the trust the nation has in the teaching workforce. Teachers are expected to evaluate students regularly using guidelines from a national core curriculum and are trusted to do so (OECD, 2011).

Despite Finland having a successful education system, the research suggests that there are limitations to the applicability of the nation's policies and practices to the U.S. First, unlike the U.S., Finland has a small population, similar to that of Minnesota, and is scarcely diverse (Central Intelligence Agency [CIA], 2012). Teachers in Finland tend to earn competitive salaries compared to other professions, unlike in the United States; however, the salary is not significantly greater than the OECD average teacher salary (National Center on Education and the Economy [NCEE], n.d.). Also unlike the United States, schools in Finland are equally funded, regardless of wealth or location, and each school has a welfare team to ensure the contentment of each child. In addition, all education, from preschool to the university level, is free for any person living in the country (Strauss & Sahlberg, 2012).

Japan. Out of 65 participating countries/economies, Japan ranked eighth in reading, ninth in math, and fifth in science on PISA 2009 (Fleischman et al., 2010). Japan is ethnically homogeneous and smaller than the U.S.; however, the literature suggests that the nation's approaches to education may be transferable to the U.S. The OECD (2011) described various aspects of the Japanese education system that may contribute to its success on international assessments:

- First, Japan uses a national curriculum that each school and classroom uses uniformly. Teachers use the same materials and provide similar lesson plans, which make comparison among schools and among individual students simpler. All students, regardless of ability, are placed in large classes of their peers, and are all held to the same high expectations. Performance is considered a reflection of effort and a commitment to studying, rather than innate ability.

- Teachers in Japan receive a tremendous amount of support from other faculty members, as their performance is a reflection of their peers' performance. New teachers perfect their teaching skills and techniques by observing experienced teachers, applying new knowledge to their own classrooms, and receiving feedback on their performance. This consistent approach of observation, application and practice, and communication may continuously improve teaching skills and student achievement.

- Japanese students are constantly engaged in academics. How a student performs is a reflection of his or her family, teachers, and peers. Parents participate in school discussions and meetings and are in constant communication with their child's homeroom teachers. Students spend one hour each day in a homeroom class and remain in the class for the duration of high school. Homeroom is considered to be like a family within school, and homeroom teachers make home visits, speak to parents, and keep track of how students are performing academically. This enables students in Japan to feel part of a community while at school and have constant support. While in the classroom, teachers keep students engaged with experiments, observations, constructive reviews of mistakes, and problems to solve in groups.

- Students, parents, teachers, and school administrators are all accountable for student performance. High school and university entrance exam results are regularly printed in the newspaper, providing incentives for these parties to maintain high student achievement.

- Lastly, funding for education in Japan is allocated differently than in the U.S. Schools are visually plain and basic in structure, lacking cafeterias and other amenities. Textbooks are thin, concise, and printed in paperback. Most funding is placed into teacher development.

Although the OECD (2011) identified these factors as possibly contributing to Japan's academic success, the literature suggests that they are not completely transferable. For instance, Japanese students have a rigorous academic schedule. Students spend long hours in school and additional hours completing homework, while U.S. students spend time in extracurricular activities. (OECD, 2011).

Singapore. Singapore first participated in PISA in 2009, and ranked fifth in reading, second in math, and fourth in science (Fleischman et al., 2010). According to the OECD (2011), Singapore is dedicated to recruiting the highest-quality teachers. In response to a teacher shortage years earlier, the government began recruiting top students to the teaching profession by offering them monthly stipends while attending school. The stipend is comparable to the

monthly salary of first-year graduates in other competing fields (Paine & Schleicher, 2011). Singapore also looks to other nations for ways to improve its standards in regard to the structure of the education system and specific practices. For example, Singapore adapted Germany's dual system of education to fit its own education system (NCEE, 2012). Germany's dual education system combines apprenticeships in a company and vocational education at a vocational school into one program (OECD, 2011).

Although Singapore has been successful in building a strong body of teachers and adapting practices of other countries to its own education system, there may be limits to the functionality of the nation's practices in the U.S. Singapore is small and wealthy from decades of trade (CIA, 2012), so there are available funds to be allocated to teacher incentives. The research suggests that the U.S. would need to change the current method for recruiting teachers and improve teacher salaries in order to implement some of Singapore's practices.

South Korea. On PISA 2009, South Korea ranked second in reading, fourth in mathematics, and sixth in science, out of 65 participating systems (Fleischman et al., 2010). Based on the research, examining South Korea's education system may offer some insight into teacher professional development and collaboration to promote higher achievement among students. Mourshed, Chijioke, and Barber (2010) described South Korea's method of applying interschool learning to the education system, in which schools and teachers are given funds by districts to conduct research projects. Topics are chosen, research is conducted, reports are published, and other teachers are invited to peer-review the findings. Participation in these research projects is thought to enhance knowledge and collaboration between schools as well as increase teacher promotions. Teachers can also gain expertise in content area and pedagogy by observing classrooms and collaborating with others (Mourshed et al., 2010).

There are cultural limitations to what practices the U.S. can adapt from South Korea. South Korea performs well on international assessments; however, much of the knowledge that the exams assess is attained because of the studying practices of South Korean students. According to Ripley (2011), families in South Korea spend 2% of their GDP to pay for after-school tutoring academies known as *hagwons*. Despite the 10 p.m. curfew that authorities have begun to enforce for students in these academies, many students still attend school until 1 a.m. and return again at 8 a.m. In comparison to students in the U.S. who have time for extracurricular activities, social events, and leisure time, South Korean students study five days a week for up to 14 hours to compete for university admission slots (Ripley, 2011).

Countries Similar in Size to the United States

Next, countries of geographical size similar to that of the United States will be discussed. Brazil, Canada, China, Russian Federation, and the United States have been distinguished as the five largest countries in the world. Canada and some Chinese cities are also top performers on international assessments.

Brazil. Of the 65 countries participating in PISA 2009, Brazil ranked 53rd in reading, 57th in math, and 53rd in science (Fleischman et al., 2010). While Brazil is far from top-performing, the research shows that its reform strategies have improved the quality of education throughout the country. The nation's reading score increased by 16 points in nine years (OECD, 2010c). Because Brazil is one of the largest and most ethnically diverse countries, it may be worthwhile for the U.S. to consider what the nation has done to accomplish such improvements. Over the past 15 years, Brazil has recognized the importance of education for all students, despite the varying terrain and climate of its regions. Today, over 95% of Brazil's population has access to public education (OECD, 2011). A major reform made by Brazil was the implementation of an

international benchmark system. This system allows each school to track its own progress against baseline standards. For example, the Brazilian state of Minas Gerais increased the percentage of its students reading at the recommended level from 29% to 86% between 2006 and 2010 (Mourshed et al., 2010). This was accomplished by simply gathering testing feedback and employing an improved literacy program. By allowing states to consistently track student progress, Brazil was able to identify problems more easily and implement programs where necessary.

By allowing states to consistently track student progress, Brazil was able to identify problems more easily and implement programs where necessary.

Canada. Canada is a newcomer to the high-performing countries. On PISA 2009, Canada ranked sixth in reading, 10th in math, and eighth in science (Fleischman et al., 2010). Canada may be of interest to U.S. policymakers because it has a decentralized education system similar to that of the U.S., and has recently become one of the top-performing countries on international assessments (Paine & Schleicher, 2011). Specifically, Ontario has been a standout province in the country because of its cooperation among the ministries, unions, and government. According to Paine and Schleicher (2011), the teacher's union, the Ministry of Education, and the government came together in 2003 to discuss ways to improve teaching practices, with a focus on primary literacy. The parties agreed that teachers would work toward a goal of 75% of all students reaching a certain level of performance by graduation. In return, the government promised to supply an unlimited amount of professional development and leadership support for existing and prospective teachers. The result showed success, and Canada rose from the bottom of the PISA rankings to among the top-performing nations. By allowing teachers to express their ideas and come to terms with the government's expectations, Ontario was able to create a rapport between the parties — something vital to long-term change (OECD, 2011).

China. China's PISA scores strictly come from the country's most developed cities, such as Hong Kong and Shanghai, and therefore it may be unfair to compare them to test results from the U.S. or other countries. Many reform efforts have been made in China to recentralize the system in regard to funding as well as reforming curriculum. Hawkins (2000) described recentralization efforts since 1949 that attempted to promote equity among schools. A reduction in local school funding from the central government caused poorer schools to see negative results. Schools were forced to find alternative forms of funding through fundraising and private organizations, as well as through student tuition fees. As developments continued, the central government carefully monitored the progress (Hawkins, 2000). China's decentralization of education may provide a useful policy comparison for the U.S. to make structural changes to its system.

Because of the inequity of education in China, many reforms have been made in an attempt to ameliorate the performance gaps. The China Education Center (2012) described the implementation of "Compulsory Education Law of the People's Republic of China" in 1986, which involved nine years of compulsory schooling in primary and junior secondary schools. As of 2010, the net enrollment rate of primary-school-age children as well as children continuing their studies in junior secondary schools was found to be above 99%. The government specifically placed great importance on compulsory education in rural, poor,

and minority areas, as these areas have seen significantly lower achievement on international assessments compared to urban areas (Fleischman et al., 2010). The China Education Center stated that the development of rural education and the local economy has been promoted through efforts surrounding the integration of education development and the upgrading of the quality of the labor force. Similar performance gaps are seen in the U.S.; therefore, looking to China's reforms may provide some insight into how to reduce such gaps in the U.S.

Russian Federation. In addition to being of similar size to the U.S., the Russian Federation has a large immigrant population. According to Matthews (2009), Russia is the second-largest immigrant destination in the world, behind only the U.S. In 2008, almost seven million immigrants from countries such as the Ukraine, Uzbekistan, Moldova, and Kyrgyzstan entered Russia to fill the semiskilled job positions that the country was promoting. Immigrants in Russia differ from immigrants in the U.S., which may limit the comparisons between the effects of immigration between the two countries. The majority of Russia's immigrants originate from countries of the former Soviet Union; therefore, many are already familiar with the Russian language once entering the country (Matthews, 2009). Children may already be equipped with the ability to read and write in Russian and so will not struggle with assessment content in a "foreign" language. On the contrary, the majority of U.S. immigrants originates from Spanish-speaking countries and is unfamiliar with English. This disparity in immigrant language familiarity may prevent a transferable comparison between Russia and the United States. Nevertheless, the Russian Federation has a high educational achievement level that continues to increase (OECD, 2012). Almost 90% of Russian adults have attained at least upper-secondary education and 54% have attained tertiary education. Only three countries have a higher tertiary attainment rate among 25- to 34-year-olds than the Russian Federation. Further, only 43% of education expenditures are devoted to primary, secondary, and postsecondary, nontertiary education, which is the lowest proportion among OECD countries.

Systems with Responsive Policy

The literature has indicated that there are countries and cities that have made educational progress worth acknowledging. Although some of these countries are not top-performing, their education systems have been reformed to produce competent students who will be competitive in today's global economy. Many of these countries' efforts and determinations are reflected in their assessment scores. In this section, countries and regions that have made efforts to reform education systems are discussed, including Africa, Germany, Hungary, Poland, Shanghai, and finally the U.S. city of Boston.

Africa. There has been continuous participation across international assessments from northern African countries such as Morocco and Tunisia; however, there has been less assessment data from Africa's southern nations. Out of 45 participating countries, Ghana's and Botswana's eighth-grade students ranked 42nd and 43rd, respectively, in math on TIMSS 2011 (Mullis et al., 2012a). On PIRLS 2011, South Africa ranked third to last (Mullis et al., 2012b). Exploring the progress of African nations may be an area for future research, as no significant changes have been seen in scoring thus far despite recent reform efforts. According to Weber (2008), education reform efforts have been made in Africa in response to performance gaps similar to those in the U.S. Regarding South Africa specifically, Weber (2008) stated, "South Africa occupies the unenviable position where the divide between the rich and poor, which is also a racial gap between the white and black, is amongst the biggest in the world" (p. 3). Useful information could come from examining reform efforts and considering why improvement has yet to be seen.

Germany. In addition to Germany being the largest economy in Europe, the nation has been responsive to the results of international assessments in sparking progressive action (OECD, 2011). For these reasons, it could be useful for the U.S. to examine how Germany has responded to such results. According to the OECD, in 2000, Germany found that its students ranked in the bottom half of PISA participants, and that almost one-quarter of its 15-year-olds could not read fluently. Further research found that many German students were following educational pathways based on those of their parents. Thus, the effect of socioeconomic background on test scores was extremely high. To ameliorate the situation, Germany implemented reforms with the hope of increasing education standards for students across the country. Although the reforms are under way and will take several years to produce full results, the current progress is noteworthy, and reforms may be transferable to the U.S. (OECD, 2011).

The implementation of reforms in Germany began by imposing a set of common curriculum standards for various grade levels to ensure that students across all states were held to the same expectations, and that teachers were fully aware of what students were expected to learn (OECD, 2011). By 2007, standards for primary school, lower secondary school, and secondary school were put in place at specific grade levels for various subjects. Based on the common standards, national assessments for grades 3, 8, and 9 were created to determine whether students were meeting the requirements and achievement levels established by the standards. This soon led to statewide assessments for grades 3 and 6. The Federal Ministry of Education in Germany was adamant about setting goals and meeting them, thus greater emphasis was placed on testing. The German Ministry of Education vowed to participate in every PISA, TIMSS, and PIRLS administration to determine where its system was succeeding in comparison with the systems of other countries (OECD, 2011).

Another example of reform made by Germany was related to prospective teacher development. The OECD (2011) explained that in order to produce students who could perform at the highest level, Germany found it crucial to have high-quality teachers. Similar to Finland and Singapore, Germany began accepting students for teacher training programs only from the top third of high school graduates. Germany also began requiring students to complete a two-year program that includes supervised teaching and related course work. Once in the classroom, an induction period is required that consists of supervision and mentoring as well as the passing of an examination (OECD, 2011).

Hungary. While Hungary is dissimilar from the U.S. in terms of population, government, economy, and language, its improvement in PISA performance is worth noting. According to Halász (2011), Hungary has made improvements in its education system since the first administration of PISA in 2000. Hungary's 2009 PISA scores were similar to those of the U.S.; however, between 2000 and 2009, the country's average reading score improved by a statistically significant amount (14 points) (Halász, 2011). The Ministry of Education and Culture (MEC, 2008) suggested that the rise in PISA reading scores among students in Hungary may be attributed to a few distinct reforms:

- Hungary increased the awareness of the importance of literacy development in higher grades. A rise in the number of students in Hungary reporting that they enjoy reading occurred after books considered to be read for pleasure (e.g., the Harry Potter series), were added to compulsory reading lists (Halász, 2011).

- The nation improved education for Hungary's most disadvantaged groups. The MEC (2008) reported the presence of a severe disparity in socioeconomic background and a lack of equal treatment in Hungary, particularly in regard to students of Roma origin. Schools are segregated and, like the U.S., those students and schools that need the

By applying knowledge garnered from a national assessment, Hungary was able to locate which areas could benefit most from extra resources.

most resources receive the fewest (MEC, 2008). In an attempt to amend the situation, policymakers began implementing social integration programs that restructured classroom activities and provided competence building for teachers (Halász, 2011).

- Hungary created an environment beneficial to learning by improving the infrastructures of school buildings.

- Hungary also established competence-based program packages, which combined the developments of curricula, organization, leadership, and teaching competencies (Halász, 2011).

- In 2000, the National Assessment of Basic Competencies was implemented. Prior to 2000, Hungary did not have a unified assessment system. Halász (2011) explained that the assessment framework was strongly influenced by PISA. Students are assessed in grades 6, 8, and 10 in literacy and numeracy. Each student is given an identification number, which allows schools to determine which students need the most attention. Also, similar to PISA, students provide background information to facilitate policymakers in determining variables that might contribute to academic performance (Halász, 2011).

Based on the literature, the reforms listed above may have contributed to the rise in PISA reading scores from the first cycle of PISA to 2009. By applying knowledge garnered from a national assessment, Hungary was able to locate which areas could benefit most from extra resources. That, in addition to encouraging reading and improving disadvantaged school performance, contributed to the improvement of education nationwide. This improvement could be reflected in future PISA scores. Because the large socioeconomic achievement gaps in the U.S. are similar to those in Hungary, it may be beneficial for the U.S. to examine how Hungary is improving its most disadvantaged schools.

Poland. While Poland is dissimilar to the U.S. in population, diversity, and economy, the nation has made notable progress in academic achievement in a short amount of time; something the U.S. has the ability to do based on its history of fast reform (OECD, 2011). In 1999, Poland reorganized the educational track by adding an additional year of general education to ensure that all students have a solid foundation of cognitive skills. Four thousand lower-secondary schools were built for students to attend after primary school and before secondary school. In addition, Poland decentralized the central government's administrative and financial control over schools. Responsibility was placed on schools, regions, districts, and municipalities directly. This allowed for every step of the reform to be monitored and for the staff to have a greater sense of autonomy. Poland also made reforms surrounding the teaching workforce. The nation understood the importance of high-quality educators, and promoted in-service training programs with salary and status incentives (OECD, 2011). Only years after the reform, Poland's PISA scores rose dramatically. Between 2000 and 2006, Poland's average PISA score increased by 29 points (OECD, 2010a).

Shanghai. In addition to being *the* top performer on PISA 2009 (Fleischman et al., 2010), Shanghai is the largest city in China, with about 20.7 million inhabitants — only about 13.8

million of whom are registered residents (OECD, 2011). Shanghai's population and land account for 1% and 0.06% of China, respectively, yet its economy accounts for one-eighth of the nation's total income. In addition, Shanghai's emphasis on education may be the greatest in the country.

Shanghai's concentration on education has given the nation preferential treatment for implementing new education reform. According to the OECD, the city underwent two waves of education reform in 1989 and 1998, which largely reflected the structure of PISA and its goals. The first wave of reforms gave students more freedom in course selection, and the second wave sought to integrate science with humanities, national curricula with school-based curricula, and knowledge acquisition with active inquiry. The second wave departed from the idea of content-based knowledge and memorization and moved the focus to creativity and complex cognitive skills to give students an understanding of core studies relevant to everyday life. Students were placed in more elective courses and extracurricular activities to enhance skills and prove that knowledge learned in school is capable of being applied to all aspects of society (OECD, 2011).

Boston. Massachusetts ranks among the top performers on international assessments worldwide. Although policies of certain states may not be comparable to others based on population, size, and demographics, the literature provides evidence that urban U.S. cities with diverse populations have already made progress in assessment and education by making adjustments to education systems. Prior to the implementation and development of the Common Core State Standards (CCSS), Boston improved its education system and raised the standards for students and educators (Mourshed et al., 2010).

After the launch of the 1998 Massachusetts Comprehensive Assessment System (MCAS), a rigorous statewide exam of 10th-graders, it was imperative that the state reevaluate education, as almost half of students failed (Mourshed et al., 2010). In 2001, the MCAS became a requirement for the entire state, causing state leaders to turn to the initial 1998 pilot data to make effective changes. Test results indicated which districts needed the most attention and resources. Boston, the state's largest and most urban city, received $5 million of the statewide $55 million in funding to implement system changes and programs such as double-block classes, summer programs, and after-school programs. The data also acted as motivation for a professional development program for 1,000 urban principals in the city. By 2003, 12th-graders who had taken the MCAS in 2001 had achieved a pass rate of 80%.

To ensure that teachers, principals, and administrators were up to date with current student achievement levels, Boston created the MyBPS data system. This system aided faculty in determining whether, and in which areas, outcomes were improving. Districts that performed well were given more flexibility from the state, and those performing at lower levels received more intervention. Annual targets were established to close achievement gaps between socioeconomic subgroups. District leaders also encouraged teachers whose students demonstrated the best outcomes to share their methods and ideas with teacher study groups. All teachers and principals were held accountable for their classes' and schools' outcomes, and many principals either retired or were replaced.

Mourshed et al. (2010) further explained how Boston's superintendent, Tom Payzant, encouraged parents to make education a part of their social values. The superintendent met with parents and communities to attend to concerns by visiting local churches and community centers. The idea was to engage parents in education so students would feel supported and held accountable at home, not just while in school.

Between 1998 and 2008, Massachusetts's MCAS scores rose dramatically. In mathematics, students went from a 23% passing rate to 84%. In reading, students went from a 43% passing rate to 91%. Within 10 years, the state standards were met by almost all students, regardless of socioeconomic background and community. Massachusetts also made national gains in NAEP scores between 1998 and 2007, earning the largest gains in mathematics and the third largest in reading (Mourshed et al., 2010).

Additional Components of Education

Certain aspects of education are commonly distinguished among top-performing education systems, including time spent in school and a standardized curriculum. In this section, we review both, as well as how the ideas apply in the U.S.

Time in school. The argument for a longer school day in the U.S. to increase learning has been a continuous debate, but research suggests that hours spent in school may not influence learning as heavily as once thought (Hull & Newport, 2011). While primary school students in the United States spent approximately 900 hours per year in the classroom in 2011, performance among these students was still average on international assessments. In addition, while U.S. students spent more time in middle school, it did not directly correlate with higher assessment scores. Students in some high-performing countries spent fewer hours in the classroom. For example, the OECD average time spent in school for primary school students was 759 hours in 2011, with Finland and South Korea requiring the fewest hours to be spent in the classroom, with 608 hours and 612 hours, respectively. Students in China spent even less time at school, attending only 35 weeks per year compared to 36 in the United States; however, some Chinese students attend school on Saturdays, which would increase their overall study time. U.S. middle school students spent an average of 990 hours in school per year, which was close to the OECD average of 886 hours. For lower-secondary school students, Finland again required the least amount of time in the classroom — 777 hours per year — while Italy, a lower-performing nation, required 1,001 hours in middle school. Lower-secondary school students from South Korea and Japan spent approximately 867 hours in school — not far from the U.S. average — but still performed better.

Hull and Newport (2011) further explained that there is no definitive correlation between time spent in school per year and assessment performance, despite top-performing countries (i.e., Finland and South Korea) requiring less time to be spent in school. Similar to students in Japan and China, South Korean students often attend private education academies after the school day ends, suggesting that these students may spend the most amount of time engaged in learning. This would create a major discrepancy between Finland's and South Korea's performance and time spent in school. In addition, Massachusetts, the highest-performing state on all assessments, did not require significantly more schooling than did other states. Further research is necessary to determine whether a relationship between performance and time spent in school exists (Hull & Newport, 2011).

Common Core. In 2010, the U.S. released its own set of national standards to prepare students for college and the workforce, which were established by members of the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO). The standards were intended for all U.S. states to adopt to create students capable of performing on a global level (NGA Center & CCSSO, 2012). Currently, 46 states, districts, and territories have adopted the Common Core State Standards (CCSS). The CCSS define the criteria of what all students should be learning and what all teachers should be teaching at each grade level, so that any student at a given level, regardless of region, will learn the same basic subject matter, lessons, and academic foundations as his or her peers across

the country. The uniformity is thought to ultimately aid in creating national and international benchmarking comparisons.

The NGA Center and CCSSO (2012) further explained that to ensure that the CCSS work as intended, the standards:

- are aligned with college and work expectations;
- are clear, understandable, and consistent;
- include rigorous content and application of knowledge through higher-order skills;
- build upon strengths and lessons of currently existing state standards;
- are informed by other top-performing countries to ensure that all students are prepared to succeed in the global economy and society; and
- are evidence-based.

These properties of the standards are also meant to ensure systematic implementation of the CCSS across the nation.

To formulate the standards, members of the NGA Center and CCSSO were informed by advanced and successful state standards, experienced teachers, content experts, states, leading thinkers, and parents and citizens for feedback. Researchers, teachers, and content specialists were also recruited to help write the standards and set evidence-based goals. Other organizations also provided guidance toward the completion of the standards, including the College Board; ACT; Achieve, Inc.; the National Association of State Boards of Education; and the State Higher Education Executive Officers.

The NGA Center and CCSSO hope that the CCSS will be a step toward centralizing education across the United States and will facilitate reforms by locating which areas of the country need the most attention. A centralized curriculum may help to close learning gaps related to socioeconomic background, as all students and teachers will be working with the same information and working toward the same curricular goals. Equity in teacher qualifications is also hoped to improve, as each teacher will have a curriculum that is explicit, understandable, and teachable.

Support for these standards is common among educators, educational organizations, private companies, state departments of education, and individuals with expertise in education (NGA Center & CCSSO, 2012). It has become more apparent that the need to educate students in terms of global readiness and college preparedness is crucial for success in the United States. The College Board (2009) released the following statement about the Common Core State Standards:

> If the U.S. is to return to a position of leadership in college completion and prepare students for high-skills jobs in a global economy, it is essential that states, schools, and higher education develop a consensus concerning the skills and knowledge required for success in college and beyond. The Common Core State Standards are an important first step in developing this consensus with rigorous and clear criteria that will provide a road map for success in rigorous college readiness programs.

The Bill & Melinda Gates Foundation (2010) also encouraged the establishment of the Common Core State Standards in regard to innovation and evidence-based instruction and instruction content. The Foundation (2010) stated:

> The new Common Core State Standards will bring consistency and clarity to American education. These college- and career-ready academic standards will provide a springboard for innovation in education. And, crucially, standards will help educators improve student achievement levels, an outcome that will benefit students personally while also fueling our nation's future economic success. Unlike most previous state standards, the Common Core State Standards are based on evidence, and not merely on what people thought was appropriate to include. The standards' developers drew from sources like incoming freshmen's college expectations, studies measuring the time required to teach core content, and the academic demands made on students in other countries. (p. 1)

The Bill & Melinda Gates Foundation placed emphasis on creating students to be prepared for college and careers and ultimately contribute to the success of the nation. With international assessments, it will be possible to monitor the impact of CCSS in comparison to other countries while balancing nation-specific needs with these benchmarks.

Conclusions

The research shows that examining national policy of states with relevant cultural, geographical, and economic features can provide further and beneficial insights for how the United States might reform its education systems. In this review, we presented the various international assessments — PISA, TIMSS, PIRLS, and CIVED, among others — and some national assessments, most notably NAEP, to provide approaches for how policymakers can make use of the results of such studies to inform education policy decisions. Reviewing studies that used results of international assessments allowed us to determine common features of high-performing education systems, such as rigorous national curriculum standards and a highly respected, high-quality teacher workforce. Assessments like NAEP can be linked to international assessments to broaden results and comparability for U.S. states and districts to other nations to help benchmark performance internationally and provide suggestions for best practices. More important, U.S. policymakers can look to the results of international assessments for ways to decrease performance gaps related to socioeconomic status. Although the relationship between economic prosperity and educational achievement is not clear, based on the research presented in this review, increasing educational achievement is a priority for the U.S. as well as for the nations discussed here, and attention must be paid to the number of low-performing students in our country without neglecting those that are high-performing. By identifying which states or districts perform well on international assessments, successful policies can be replicated on a larger scale.

References

Amadeo, J., Torney-Purta, J., Lehmann, R., Husfeldt, V., & Nikolova, R. (2002). *Civic knowledge and engagement: An IEA study of upper secondary students in sixteen countries*. Amsterdam: IEA.

Baker, K. (2007). Are international tests worth anything? *Phi Delta Kappan, 89*(2), 101–104.

Baldi, S., Perie, M., Skidmore, D., Greenberg, E., & Hahn, C. (2001). *What democracy means to ninth-graders: U.S. results from the international IEA Civic Education Study* (NCES 2001-096). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Biddle, B. J. (2012). Splendid contributions and flawed conclusions. *Educational Researcher, 41*(4), 179–181.

Bill & Melinda Gates Foundation. (2010). *Fewer, clearer, higher: Moving forward with consistent, rigorous standards for all students*. Retrieved from http://www.gatesfoundation.org/highschools/Documents/fewer-clearer-higher-standards.pdf

Bils, M., & Klenow, P. J. (1998). *Does schooling cause growth, or the other way around?* (Working Paper No. 6393). Cambridge, MA: National Bureau of Economic Research.

Bracey, G. W. (2000). The TIMSS "final year" study and report: A critique. *Educational Researcher, 29*(4), 4–10.

Bracey, G. W. (2009). PISA: Not leaning hard on U.S. economy. *The Phi Delta Kappan, 90*(6), 450–451.

Central Intelligence Agency. (2012). *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/

China Education Center. (2012). *Primary and secondary education*. Retrieved from http://www.chinaeducenter.com/en/cedu/psedu.php

College Board. (2009). *College Board statement in response to release of NGA center/ CCSSO Common Core State Standards initiative*. Retrieved from http://press.collegeboard.org/releases/2009/college-board-statement-response-release-nga-centerccsso-common-core-state-standards-initiat

Cummins, J. (2008). Essay review [Review of the book *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*, by P. Stanat & G. Christensen]. *Curriculum Inquiry, 38*(4), 493–497.

Egan, L., Beattie, K., Byrd, P., Chadwick, D., & DeCandia, M. (2011, June). *NAEP, PISA and TIMSS and PIRLS at a glance*. Handout for paper presented at the 2011 National Conference on Student Assessment, Orlando, FL. Retrieved from https://ccsso.confex.com/ccsso/2011/webprogram/Handout/Session2236/NAEP,%20PISA,%20and%20TIMSS,%20PIRLS%20At%20a%20Glance.pdf

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context* (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Halász, G. (2011, February 10). *The background of the improvement of PISA results in Hungary — The impact of the EU funded educational development programs*. Paper presented at the Trends in Performance Since 2000 — International Launch of PISA 2009 Report conference, Warsaw. Retrieved from http://halaszg.ofi.hu/download/Trendy.pdf

Hanushek, E. A., & Woessmann, L. (2007). *Education quality and economic growth*. Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Hawkins, J. N. (2000). Centralization, decentralization and recentralization: Educational reform in China. *Journal of Educational Administration, 38*(5), 442–454.

Hilton, M. (2006). Measuring standards in primary English: Issues of validity and accountability with respect to PIRLS and national curriculum test scores. *British Educational Research Journal, 32*(6), 817–837.

Hong Kong Examinations and Assessment Authority. (2010). *Hong Kong advanced level examination profile*. Retrieved from http://www.hkeaa.edu.hk/tc/hkale/Introduction/

Hull, J., & Newport, M. (2011). *Time in school: How does the U.S. compare?* Center for Public Education. Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Organizing-a-school/Time-in-school-How-does-the-US-compare

International Association for the Evaluation of Educational Achievement. (2011). *IEA studies*. Retrieved from http://www.iea.nl

Jerald, C. D. (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: National Governors Association. Retrieved from http://www.nga.org/files/live/sites/NGA/files/pdf/0812BENCHMARKING.PDF

Karp, J. (n.d.). *Education in France: The Baccalauréat vs. A-levels*. Retrieved from http://www.ofqual.gov.uk/qualifications-assessments/89-articles/13-a-and-as-levels

Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin M. O., Mullis, I. V. S., & Kennedy, A. M. (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timss.bc.edu/pirls2006/tech_rpt.html

Matriculation Examination Board. (n.d.). *The Finnish matriculation exam*. Retrieved from http://www.ylioppilastutkinto.fi/en/

Matthews, O. (2009). The Kremlin vigilantes. *The Daily Beast*. Retrieved from http://www.thedailybeast.com/newsweek/2009/02/13/the-kremlin-vigilantes.html

Ministry of Education and Culture. (2008). *Education in Hungary: Past, present, future — An overview*. Department for EU Relations. Retrieved from http://www.nefmi.gov.hu/letolt/english/education_in_hungary_080805.pdf

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. McKinsey & Company. Retrieved from http://mckinseyonsociety.com/how-the-worlds-most-improved-school-systems-keep-getting-better/

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012a). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012b). *PIRLS 2011 international results in reading.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009b). *PIRLS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009a). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Center on Education and the Economy. (n.d.). *Center on International Education Benchmarking: Learning from the world's high performing education systems*. Retrieved from http://www.ncee.org/programs-affiliates/center-on-international-education-benchmarking/

National Center on Education and the Economy. (2012). *Tucker's lens: Reflections on Singapore*. Center on International Benchmarking. Retrieved from http://www.ncee.org/2012/06/tuckers-lens-reflections-on-singapore/?utm_source=News and Updates from NCEE&utm_campaign=3ef7c56f7b-TopoftheClass_Volume1Issue6&utm_medium=email

National Center for Education Statistics. (n.d.a). *National Assessment of Educational Progress (NAEP): NAEP overview*. Retrieved from http://nces.ed.gov/nationsreportcard/

National Center for Education Statistics. (n.d.b). *The nation's report card*. Retrieved from http://nationsreportcard.gov/

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2012). *Common core state standards initiative*. Retrieved from http://www.corestandards.org/

Obama, B. (2009, July). *Remarks by the president*. Speech presented at the NAACP Centennial Convention, New York, NY.

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Organisation for Economic Co-operation and Development. (n.d.). *OECD Programme for International Student Assessment (PISA)*. Retrieved from http://www.oecd.org/pisa

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world* (Volume 1). Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2009). *PISA 2009 assessment framework — Key competencies in reading, mathematics and science*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2010a). *PISA 2009 results: Executive summary*. Retrieved from http://www.pisa.oecd.org/document/61/0,3746,en_32252351_32235731_46567613_1_1_1_1,00.html#Executive_summary

Organisation for Economic Co-Operation and Development. (2010b). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes*. Paris: OECD Publishing.

Organisation for Economic Co-Operation and Development. (2010c). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science* (Volume 1). Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2011). *Lessons from PISA for the United States: Strong performers and successful reformers in education*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2012). *Education at a glance: OECD indicators 2012.* Paris: OECD Publishing. Retrieved from http://www.oecd.org/edu/EAG%202012_e-book_EN_200912.pdf

Paine, S. L., & Schleicher, A. (2011). *What the U.S. can learn from successful education reform efforts*. Retrieved from http://mcgraw-hillresearchfoundation.org/2011/03/16/what-the-u-s-can-learn-from-successful-education-reform-efforts/

Peterson, P. E., Woessmann, L., Hanushek, E. A., & Lastra-Anadón, C. X. (2011). Globally challenged: Are U.S. students ready to compete? The latest on each state's international standing in math and reading. *Harvard University Program on Education Policy and Governance & Education Next, 11*(3), 1–34. Retrieved from http://hks.harvard.edu/pepgeducationnext.org

Phillips, G. W. (2007). *Expressing international educational achievement in terms of U.S. performance standards: Linking NAEP achievement levels to TIMSS*. American Institutes for Research. Retrieved from http://www.air.org/files/naep-timss.pdf

Phillips, G. W. (2010). *International benchmarking: State education performance standards*. Washington, DC: American Institutes for Research. Retrieved from http://www.air.org/files/AIR_Int_Benchmarking_State_Ed__Perf_Standards.pdf

Phillips, G. W., & Jiang, T. (2011, April). *Using PISA to internationally benchmark state performance standards*. Paper presented at the 2011 annual conference of the American Educational Research Association, New Orleans, LA. Retrieved from http://www.schoolinfosystem.org/pdf/2011/06/pisabenchmarkstates.pdf

Ripley, A. (2011). Teacher, leave those kids alone. *Time Magazine World*. Retrieved from http://www.time.com/time/magazine/article/0,9171,2094427,00.html

Robinson, P. (1999). The tyranny of league tables: International comparisons of educational attainment and economic performance. In R. Alexander, P. Broadfoot, & D. Phillips (Eds.), *Learning From Comparing: New Directions in Comparative Educational Research, Vol. 1*, Contexts, Classrooms and Outcomes. Oxford: Symposium.

Salzman, H., & Lowell, L. (2008). Making the grade. Nature, 453, 28-30. Retrieved from http://www.nature.com/nature/journal/v453/n7191/full/453028a.html

Schütz, G., Ursprung, H. W., & Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, 61(2), 279–308.

Stanco, G. (2012). *Using TIMSS 2007 data to examine STEM school effectiveness in an international context*. Retrieved from ProQuest Digital Dissertations (UMI 3503193).

Strauss, V., & Sahlberg, P. (2012). What the U.S. can't learn from Finland about ed reform. *The Washington Post*. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/post/what-the-us-cant-learn-from-finland-about-ed-reform/2012/04/16/gIQAGIvVMT_blog.html

Theisen, G. L., Achola, P. P., & Boakari, F. M. (1983). The underachievement of cross-national studies of achievement. *Comparative Education Review*, *27*(1), 46–68.

Tienken, C. H. (2008). Rankings of international achievement test performance and economic strength: Correlation or conjecture? *International Journal of Education Policy and Leadership, 3*(4), 1–15.

TIMSS & PIRLS International Study Center. (2012). *TIMSS advanced*. Retrieved from http://timssandpirls.bc.edu/timss_advanced/index.html

Torney-Purta, J., Barber, C. H., & Wilkenfeld, B. (2007). Latino adolescents' civic development in the United States: Research results from the IEA Civic Education Study. *J Youth Adolescence, 36*(2), 111–125.

Tucker, M. (Ed.). (2011). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Cambridge, MA: Harvard Educational Press.

Tunmer, W. E., Nicholson, T., Greaney, K. T., Prochnow, J. E., Chapman, J. W., & Arrow, A. W. (2008). PIRLS before swine: A critique of New Zealand. *New Zealand Journal of Education Studies, 43*(2), 106–119.

Wang, J. (2001). TIMSS primary and middle school data: Some technical concerns. *Educational Researcher, 30*(6), 17–21.

Weber, E. (Ed.) (2008). *Educational change in South Africa: Reflections on local realities, practices and reforms*. Rotterdam/Taipei: Sense Publishers.

Wolf, A. (2004). Education and economic performance: Simplistic theories and their policy consequences. *Oxford Review of Economic Policy, 20*(2), 315–333.

World Bank Group. (2012). *GDP growth (annual %)*. Retrieved from http://data.worldbank.org

Appendix

Table A1.

Population, Language, Government, Economy, and International Assessment Rankings, by Country

Country		Population (approximate) (2012)	Official Languages	Government	GDP per capita (USD) (2011)	PISA 2009 ranking			TIMSS 2011 Ranking		PIRLS 2011 Ranking
						Math	Reading	Science	Math	Science	Reading
China	Shanghai	1,343,239,923	Mandarin and Cantonese (among others)	Communist state	8,500, 121st highest	1st	1st	1st	N/A	N/A	N/A
	Hong Kong					3rd	4th	3rd	3rd (4th grade); 4th (8th grade)	9th (4th grade; 8th (8th grade)	1st
Singapore		5,353,494	Mandarin, English, Malay and Tamil	Parliamentary republic	60,500, 5th highest	2nd	5th	4th	1st (4th grade); 2nd (8th grade)	2nd (4th grade); 1st (8th grade)	4th
South Korea		48,860,500	Korean and English	Republic	32,100, 40th highest	4th	2nd	6th	2nd (4th grade); 1st (8th grade)	1st (4th grade); 2nd (8th grade)	N/A
Finland		5,262,930	Finnish and Swedish	Republic	36,700, 32nd highest	6th	3rd	2nd	8th (4th grade); 8th (8th grade)	3rd (4th grade); 5th (8th grade)	3rd
Japan		127,368,088	Japanese	Parliamentary government with a constitutional monarchy	35,200, 36th highest	9th	8th	5th	5th (4th grade); 5th (8th grade)	4th (4th grade); 4th (8th grade)	N/A
Canada		34,300,083	English and French	A parliamentary democracy, a federation, and a constitutional monarchy	41,100, 20th highest	10th	6th	8th	N/A	N/A	12th
Germany		81,305,856	German	Federal republic	38,400, 26th highest	16th	20th	13th	16th (4th grade)	17th (4th grade)	17th
Poland		38,415,284	Polish	Republic	20,600, 60th highest	25th	15th	19th	34th (4th grade)	30th (4th grade)	28th
Hungary		9,958,453	Hungarian	Parliamentary democracy	19,800, 63rd highest	29th	26th	22nd	20th (4th grade); 6th (8th grade)	10th (4th grade); 11th (8th grade)	20th
United States		313,847,465 3rd Largest	None (English majority)	Constitution-based federal republic; strong democratic tradition	49,000, 11th highest	31st	17th	23rd	11th (4th grade); 9th (8th grade)	7th (4th grade); 10th (8th grade)	6th
Russian Federation		138,082,178	Russian	Federation	17,000, 71st highest	38th	43rd	39th	10th (4th grade); 6th (8th grade)	5th (4th grade); 7th (8th grade)	2nd
Brazil		205,716,890	Portuguese	Federal republic	11,900, 101st highest	57th	53rd	53rd	N/A	N/A	N/A

Note: Information in this table is from the Central Intelligence Agency (2012), Fleischman et al. (2010), Mullis et al. (2012a; 2012b), and Martin et al. (2012).

Figure A1.

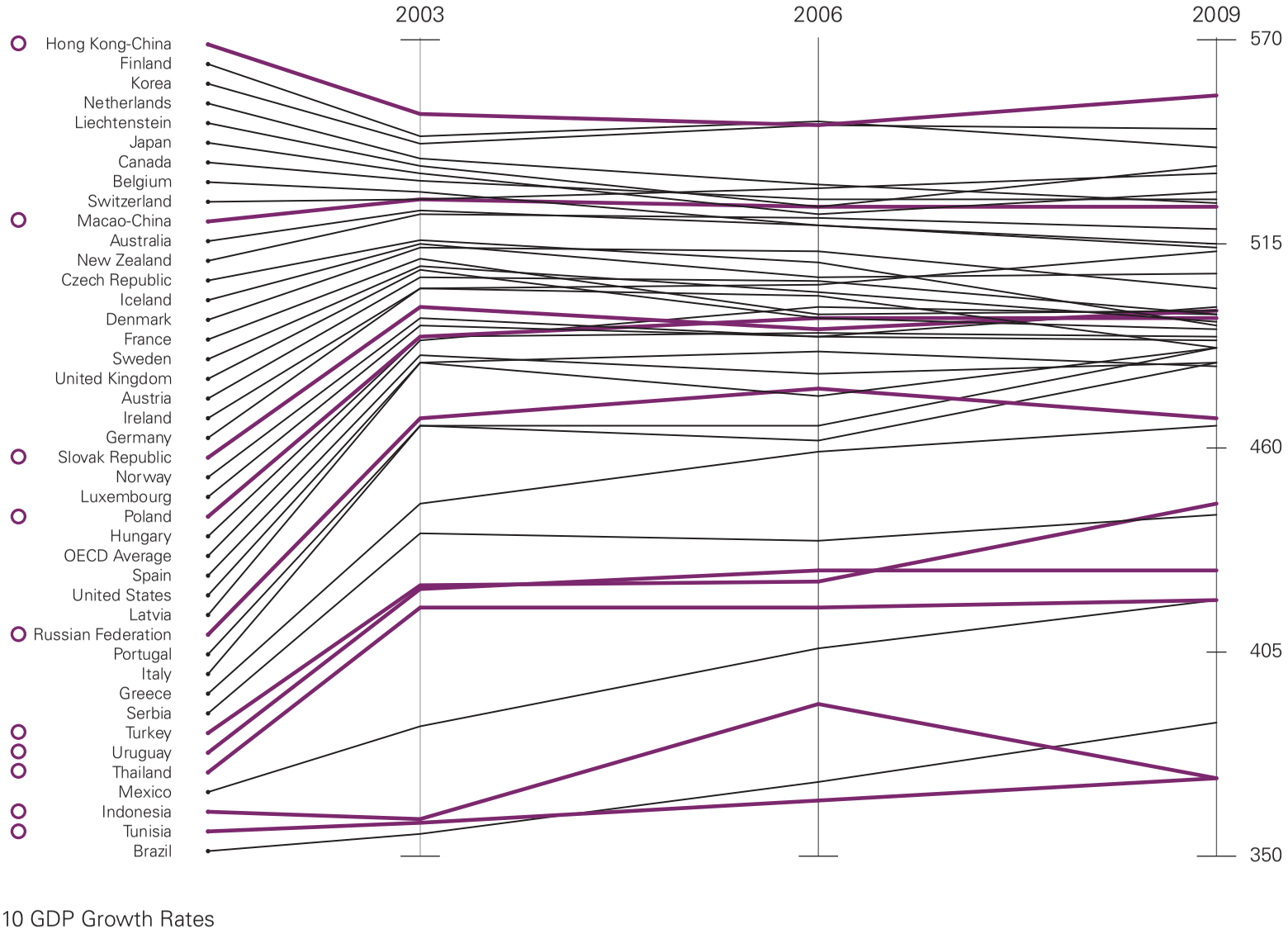
Economic growth and change in PISA reading performance, by country.

Top 10 GDP Growth Rates

Note: Trend lines for countries with the top 10 economic growth rates are in bold. Top 10 growth rates indicate the top 10 GDP growth rates among countries in the figure from 2003 until 2009. The OECD PISA performance averages and GDP growth rate averages include all OECD countries; however, the following OECD members are not displayed in the figure as these countries did not participate in all three PISA administrations: Chile, Estonia, Israel, and Slovenia. In 2003, the United Kingdom did not meet minimum response rates for schools, and therefore the OECD warns that the U.K. cannot be compared to other nations or regions. In addition, 2006 reading scores for the U.S. were never reported because of administration errors.

Figure A2.

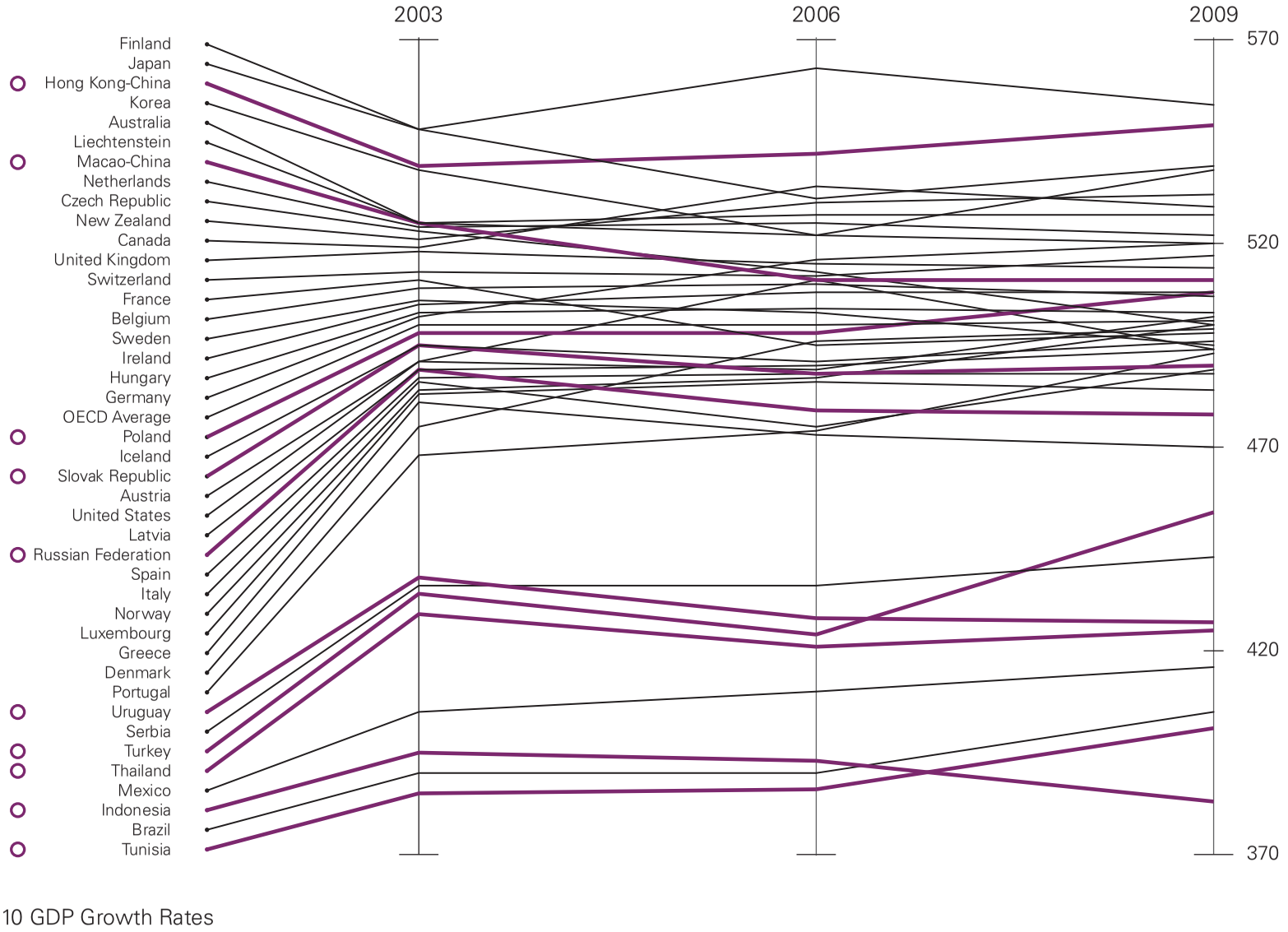
Economic growth and change in PISA mathematics performance, by country.



Note: Trend lines for countries with the top 10 economic growth rates are in bold. Top 10 growth rates indicate the top 10 GDP growth rates among countries in the figure from 2003 until 2009. The OECD PISA performance averages and GDP growth rate averages include all OECD countries; however, the following OECD members are not displayed in the figure as these countries did not participate in all three PISA administrations: Chile, Estonia, Israel, and Slovenia. In 2003, the United Kingdom did not meet minimum response rates for schools, and therefore the OECD warns that the U.K. cannot be compared to other nations or regions.

Figure A3.

Economic growth and change in PISA science performance, by country.



Note: Trend lines for countries with the top 10 economic growth rates are in bold. Top 10 growth rates indicate the top 10 GDP growth rates among countries in the figure from 2003 until 2009. The OECD PISA performance averages and GDP growth rate averages include all OECD countries; however, the following OECD members are not displayed in the figure as these countries did not participate in all three PISA administrations: Chile, Estonia, Israel, and Slovenia. In 2003, the United Kingdom did not meet minimum response rates for schools, and therefore the OECD warns that the U.K. cannot be compared to other nations or regions.

The Research department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: research.collegeboard.org

CollegeBoard

CollegeBoard

12b-6934