
Comparing U.S. States’ Mathematics Results in PISA and Other International and National Student Assessments

Maria Stephens and Anindita Sen

For the first time in 2012, three U.S. states – Connecticut, Florida, and Massachusetts – participated in the OECD’s Program for International Student Assessment (PISA) as individual entities in order to obtain an international benchmark of student performance. PISA measures students’ reading, mathematics, and science literacy at 15 years of age, which is near the end of compulsory school in most of the participating countries. However, while this participation marked states’ debut into PISA, which has been on going on a three-yearly basis since 2000, it was not the states’ first foray into international student assessment. These three states, as well as others, have shown the same increasing interest in measuring their students against other students around the world that numerous countries themselves have shown (see Exhibit 1). For example, one of the three PISA-participant states was involved as early as 1995 in the IEA’s administration of the Trends in International Mathematics and Science Study (TIMSS). U.S. states are now involved in all three major international student assessments, including the Progress in International Reading Literacy (PIRLS) Study.

This subnational participation in international assessments provides value nationally by contributing to a better understanding of the variation in national statistics and, for states, by providing a sense of the global comparative health of their education systems. However, one of the challenges in using and interpreting the international data, among the numerous other sources of information to which states have access to, is in understanding differing results across programs. This article thus focuses on the question: *What specific factors might explain differences in the United States’ results on PISA 2012 and their results on other recent international and national assessments?* It describes

the results of a comparative analysis of four possible factors: (1) differences in overall content distribution of the items, (2) differences in relative strengths and weaknesses on content and cognitive subscales, (3) differences in sampling, and (4) differences in participating countries. It does not examine epistemological, ontological, or other methodological differences among the assessments. Data examined include the mathematics results from PISA 2012, TIMSS 2011 at the eighth grade, and the National Assessment of Educational Progress (NAEP) 2011 at the eighth grade.¹

The article describes: (1) the educational contexts of the three PISA-participant states (Connecticut, Florida, and Massachusetts) and the assessment programs whose results are examined (PISA 2012, TIMSS 2011, and NAEP 2011) and, (2) differences in the key mathematics results for the three PISA-participant states on the three assessments, and (3) the results of the comparative analysis of factors potentially contributing to those differences.

Table 1: Countries' and jurisdictions' participation in PISA and other international student assessments: 1995-2012

Program	Year	No. of countries	No. of benchmarking jurisdictions ¹	No. of U.S. states among the benchmarking jurisdictions
PISA ²	2012	64	4	3
	2009	70	5	0
	2006	57	0	0
	2003	41	0	0
	2000	43	0	0
TIMSS ¹	2011	63	14	9
	2007	55	8	2
	2003	49	4	1
	1999	39	27	13
	1995	43	6	6
PIRLS ⁴	2011	48	9	1
	2006	41	0	0
	2001	35	0	0

¹ “Benchmarking jurisdictions” refers to subnational entities that participate independently in an assessment – i.e., either representing an incomplete subset of a nation’s subnational jurisdictions or those that finance their own participation in addition to the nation’s participation. The OECD does not separately identify “benchmarking jurisdictions” because until 2009, no subnational jurisdictions participated independently. (China’s two autonomous states of Hong Kong and Macao have participated since 2000 and 2003, respectively and are instead included in the country

¹ Mathematics is examined because it was the focus of the 2012 PISA cycle.

count. Additionally, a number of federal countries have voluntarily oversampled in various years to provide for disaggregation within the national data and these cases are not counted as benchmarking jurisdictions.) For the purposes of this table, we have included the five subnational jurisdictions that were represented in PISA 2009 (one each from China, the United Arab Emirates [UAE], and Venezuela and two from India) and the four from 2012 (Shanghai-China and the three U.S. states). The IEA has historically treated the subnational jurisdictions of the Flemish and French communities of Belgium and the nations of the United Kingdom as individual education systems, on par with other national systems and these are included in the country count for TIMSS and PIRLS. However, they separately identify other subnational jurisdictions such as the various Emirates of the UAE, U.S. states, or Canadian provinces. This column does not include district or district consortia participation.

2 Counts include countries, jurisdictions, and states that administered a given year's assessment in the primary year or a follow-up wave (e.g., 2000 PISA in 2001/2 or 2009 PISA in 2010).

3 The counts include participants in 4th and/or 8th grade TIMSS.

4 Only more recently (2011) has the Progress in Reading Literacy Study (PIRLS) been opened for subnational participation. Florida was the U.S. state that participated in PIRLS 2011.

Background on State Education Systems and Assessment

Education in the United States is decentralized, with each state having responsibility for governing its own education system. These responsibilities including distributing federal and state funding, establishing policies (such as the duration of compulsory education, requirements for graduation, and minimum teacher qualifications), providing guidance regarding curriculum, conducting student assessments, and ensuring equal access to education for all eligible students. Often, some of these responsibilities — particularly those related to instruction — are further delegated to localities, which manage the operation of schools in their districts. While some aspects of education are very similar across states (e.g., the organization of schools), other characteristics (e.g., policies for compulsory education, demographics, teacher salaries) vary (see Exhibit 2, which provides a brief overview of education in the PISA-participant states).

The three PISA-participant states, as well as the other U.S. states, have access to a number of different macro-measures of student performance, and for the purposes of this article, we focus on those that can currently be compared across states, including the National Assessment of Educational Progress (NAEP) and the international assessments, PISA and TIMSS (see textbox for information and context on other macro-measures of student performance).

NAEP is the longest-standing measure of student performance for most states. The NAEP 4th- and 8th-grade assessments in reading and

mathematics, which are given every two years, are effectively required, as participation is a condition of receiving Title I funding, which is a primary financial resource (over \$14 billion in 2014) for school districts and schools with high percentages of disadvantaged students (Federal Education Budget Project, 2014). The other NAEP assessments, including those at 12th grade and in other subjects, are voluntary. NAEP is designed to measure the knowledge and skills students have acquired in school on content determined through the collaborative input of a wide range of experts and participants from government, education, business, and public sectors in the United States. As the “nation’s report card,” NAEP is supposed to reflect what U.S. students should know and be able to do. For states, the benefit is the long trend line and the applicability to the U.S. context.

Table 2: Overview of Selected Education System Characteristics in 3 U.S. States: 2011-12

	Connecticut	Florida	Massachusetts
<i>Governance</i>			
Appoints the State Superintendent	State Board	Governor	State Board
Appoints the State Board	Governor	Governor	Governor
<i>Structure</i>			
Typical organization	Elementary education (Kindergarten through grade 5) Middle school (grade 6-8) High school (grade 9-12)		
Entrance age	Must be 5 by January 1 (of school year)	Must be 5 by September 1	Localities determine
Compulsory education	5-18	6-16	6-16
<i>Demographics</i>			
No. of districts	200	76	401
No. of schools	1,150	4,212	1,835
No. of students	554,437	2,668,156	953,369
No. of teachers	43,805	175,006	69,342
Student-teacher ratio	12.7	15.2	13.7
Percent of students FRPL ¹	34.5	56.0	34.2
<i>Finance</i>			
Total expenditure on public elementary and secondary education ²	\$9,094,036,286	\$23,870,090,268	\$13,649,965,365
Average annual salary of public elementary and secondary teachers 2011-12	\$70,821	\$46,232	\$72,000

1 Reference year is 2010-11. FRPL is free and reduced price lunch, indicating students with lower socioeconomic resources.

2 Reference year is 2010-11.

Sources: NCES, 2014; ECS, 2014a; ECS, 2014b.

The international student assessments, PISA and TIMSS, are not required, though the Common Core initiative—described in the text box—has underscored the value of states' engaging in assessment in an international context. The Common Core initiative, which began around 2008 to increase rigor across state education systems, is both a result of and a driver of states' participation in international assessments. Since the 1995 administration of TIMSS, a total of 18 states have participated in at least one cycle of TIMSS, with 9 participating in multiple cycles and 9 participating in the most recent 2011 cycle. Additionally, as an indirect measure, states have looked to the estimates produced by the NAEP-TIMSS Linking Studies, the most recent of which used improved methodology to estimate TIMSS scores for each of the 50 states based on their NAEP scores and the NAEP and TIMSS results from the 9 states participating in both assessments in 2012 and 2011, respectively (NCES, 2013). This has been an important – if less reliable – source of information and significantly less costly than actual participation in international assessments.

Table 3: Overview of Selected Characteristics of Assessment Programs

	PISA 2012	TIMSS 2011	NAEP 2011
Frequency	Every 3 years	Every 4 years	Every 2 years
Target population	15 years old ¹	Grades 4 and 8	Grades 4, 8, and 12
No. of schools sampled ²	>300	>1,000	7,610
No. of students sampled ²	~11,000	~30,000	175,200

1 In the United States, PISA's age-based national sample included students mostly in grade 10 (71 percent in 2012), though some were in grade 11 (17 percent), grade 9 (12 percent), or other grades (less than 1 percent).

2 For TIMSS and NAEP, the numbers are for grade 8 only. For all three assessments, the numbers include state participants.

Sources: Provasnik et al., 2013; Mullis et al., 2012; NCES, 2012.

Evolving State Assessments and the Context for International Participation

All U.S. states also have state assessments. Some states have had assessment systems for decades, others initiated them in the 1990s with the passage of state accountability laws, and the rest developed or expanded them under the requirements of the No Child Left Behind Act (NCLB) 2001 (Chingos, 2012). NCLB required that states test all students in grades 3-8 and in one grade in high school in mathematics and reading. Prior to NCLB, only 13 states had assessment systems this extensive (Danitz, 2001, as cited in Chingos, 2012). State assessments, however, are in the midst of another major change, as most states – with a boost from incentives from the federal level – have adopted the Common Core State Standards, which is an initiative that developed common standards in core academic subjects, and most are collaborating on the development of assessments of those standards that will replace their existing

systems. This will mean that, for the first time, there will be comparability in learning standards across states and in performance measures among at least some states. The lack of comparability and variable quality across states has been an often-cited weakness of NCLB in the past (e.g., Linn, Baker, and Betebenner, 2002).

The main purpose of the Common Core is to increase the rigor of standards and align them with the expectations of education institutions and employers so that students meeting the standards will be ready for college or a career. A major driver of the Common Core was the states themselves – the initiative is managed by the Council of Chief State School Officers and the National Governors Association – and their expressed need for improved benchmarking – namely, “comparing outcomes to identify top performers or fast improvers, learning how they achieve great results and applying those lessons to improve...performance” (NGA, 2008, p. 9), with an explicit acknowledgement that the standards and benchmarks should have an international component. Thus not only should the standards be rigorous enough to allow U.S. students to compete in the global economy, states should measure performance in an international context (with implicit favour being given to PISA as the assessment of choice [Schneider, 2009]).

Forty-five states and the District of Columbia (including the three PISA-participant states) have adopted the Common Core standards in both English language arts and mathematics and an additional state in mathematics only. Two consortia, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment (SBAC) Consortium, are the primary groups working on the new assessments that will roll out in the 2014-15 school year. Connecticut is signed on with SBAC, Massachusetts with PARCC, and Florida with another private provider.² One analysis has suggested that with a quality implementation of the Common Core in mathematics and well-designed assessment tasks particularly at the secondary level, U.S. students would be learning the kind of mathematics that would make them potentially more competitive in PISA (OECD, 2013).³

States have participated in PISA because of its targeting of students nearing the end of compulsory school and its focus on students’ ability to apply the knowledge and skills they have learned cumulatively during their schooling, as well as in other contexts, for solving problems in a real-world context. States have participated in TIMSS, on the other hand,

2 Florida has contracted with American Institutes for Research (AIR) to develop its state assessments. AIR is the home organization of the authors of this article; the authors are in a separate division and independent of that project.

3 The referenced analysis classified the PISA 2012 items against the Common Core progression according to where they sit in the progression of standards up to high school level, the degree to which they represent attributes of modeling, and their modeling level. The analysis found a degree of commonality between the PISA and Common Core constructs leading the authors to conclude that “the high school curriculum in the United States will attend to modeling to a greater degree than has happened in the past...[and if] more students work on more and better modeling tasks than [they do] today, then one could reasonably expect PISA performance to improve” (p. 90).

because its grade-based target populations are similar to NAEP (grades 4 and 8) and it also similarly focuses on school achievement. Both the national and two international assessments collect data on mathematics and science performance, though the sampling requirements and other features differ (see Exhibit 3).

U.S. States' Mathematics Results from PISA, TIMSS, and NAEP

While each of the sources of student performance data provides valuable input for U.S. states, interpretation of results across the multiple measures requires careful consideration. (Again, since the state assessments aligned with the Common Core are not fully in place yet, we focus here on the data available from PISA and how that aligns with data available from TIMSS and NAEP.)

On average, students in the United States performed below the OECD average in mathematics literacy, scoring 481 points compared to the OECD mean of 494 in 2012 (see Exhibit 4 and Kelly et al., 2013). This masks variation among the states, however. Students in Connecticut scored an average of 506 points, which was above the U.S. mean though statistically comparable to the OECD mean. Just 12 of the 68 total participating education systems scored higher than Connecticut and its scores were comparable to those of students in 15 other systems, including the United States' partners in the G-8 Canada, France, Germany, and the United Kingdom. Students in Massachusetts scored an average of 514 points, which was statistically comparable to Connecticut's mean but above both the U.S. and OECD means. Nine education systems outperformed Massachusetts and its scores bested an additional six education systems than did Connecticut's. In contrast, students in Florida scored 467 points on average, which was lower than both the U.S. and OECD means. Florida's mean score was below that of 38 education systems and statistically comparable to a set of five education systems—Lithuania, Sweden, Hungary, Croatia, and Israel—that were outperformed by the other two PISA-participant states. These findings were not necessarily surprising as the two Northeastern states are typically above average in NAEP and Florida is typically below-average, as they were in 2011.

Looking across the assessments highlights some differences and generates interesting questions. Connecticut performed above the U.S. mean in mathematics in PISA 2012 and eighth-grade NAEP 2011, and above the international mean in TIMSS 2011 though similar to the OECD mean in PISA 2012. (It should be noted that the OECD mean in PISA is based only on the scores of the participating OECD education systems whereas

the international mean in TIMSS is based on all the participating TIMSS education systems, which is a much more diverse group in terms of student outcomes.) However, despite the differences in performance relative to the international means, Connecticut appears to have a greater advantage in PISA than in TIMSS (and NAEP), based just on distance from the U.S. mean. What might account for this advantage?

Table 4: Mathematics performance of U.S. 15-year-olds in PISA and eighth-grade students in TIMSS and NAEP: 2011 and 2012

	PISA 2012 (15-year-olds)			TIMSS 2011 (Grade 8)			NAEP 2011 (Grade 8)	
	Mean score	Relative to U.S.	Relative to OECD	Mean score	Relative to U.S.	Relative to Int'l	Mean score	Relative to U.S.
Connecticut	506	+	=	518	=	+	287	+
Florida	467	-	-	513	=	+	278	-
Massachusetts	514	+	+	561	+	+	299	+
United States	481	†	-	509	†	+	283	†
Across countries	494	+	†	500	-	†	†	†
Range across 50 states	†			466 (Ala.) – 561 (Mass.) ¹			260 (DC) -299 (Mass.)	
Range across countries	368 (Peru) – 613 (Shanghai-China)			331 (Ghana) – 613 (Korea, Rep. of)			†	

Note: PISA measures mathematics literacy, or the application of mathematics for solving real-world problems. TIMSS and NAEP focus more exclusively on school-based mathematics.

¹ The range is based on scores estimated in the NAEP-TIMSS Linking Study; results for the three PISA states, however, are actual TIMSS results as they also participated in TIMSS 2011.

† Not applicable

+ Significantly higher than reference at the .05 level.

- Significantly lower than reference at the .05 level.

= Not significantly different than reference at the .05 level.

Sources: Kelly et al., 2013; Mullis et al., 2012; NCES, 2012; and NCES, 2013.

Massachusetts also performed above the U.S. mean in mathematics on all three assessments, as well as above the respective international means for PISA 2012 and TIMSS 2011. Based on distance from the U.S. mean, however, Massachusetts appears to have a greater advantage in TIMSS (and NAEP) than PISA. Again, what might account for this particular advantage?

Finally, Florida performed lower than the U.S. mean in mathematics in PISA 2012 and eighth-grade NAEP 2011, but similar to the U.S. mean in TIMSS 2011. On the international assessments, despite being lower than the OECD mean in PISA 2012, Florida is above the TIMSS 2011 international mean. How would Florida's relative standing change if the groups of education systems participating in PISA and TIMSS were comparable? What are some possible explanations for Florida's weaker-than-average performance?

Analysis of Differences in Results

A first analysis to explore these questions is to examine the similarities and differences in terms of item content, which has been collected through studies comparing the various international assessments with each other and with NAEP.⁴ Generally speaking, these studies have shown that, overall, there are more similarities between NAEP and TIMSS than between NAEP and PISA, as might be expected given the former two programs' focus on curriculum-based achievement and the latter's on literacy (Provasnik et al., 2013; AIR, 2013). For example, PISA differs from TIMSS and NAEP in terms of the distribution of test items across content areas: PISA 2012 had a larger percentage of items that would be considered data analysis, probability and statistics items on the NAEP framework than did NAEP 2011/2013 or TIMSS 2011, whereas it had a smaller percentage of items classified as algebra (see Exhibit 5).⁵ Additionally, the most recent comparison study identified several topics covered by the NAEP 2013 item pool that were not covered by the PISA 2012 item pool – i.e., that were unique to NAEP – including: estimation; mathematical reasoning using numbers; position, direction, and coordinate geometry; mathematical reasoning in geometry; measurement in triangles; experiments and samples; mathematical reasoning with data; and mathematical reasoning in algebra (AIR, 2013). In terms of item complexity, PISA 2012 had a greater percentage of items classified as “moderate” on the NAEP framework than did NAEP 2013, and a smaller percentage classified as “low” (data not shown, AIR, 2013).

4 See <http://nces.ed.gov/surveys/international/cross-study-comparisons.asp> for a listing of these studies through 2013.

5 This is based on results from two studies: one (Lin, Darling, and Dodson, 2013) that compared the NAEP 2011 and TIMSS 2011 grade 8 mathematics items (among other elements) and another (AIR, 2013) that compared the NAEP 2013 grade 8 and PISA 2012 items (among other elements). Though different expert panels undertook the studies, the distribution of NAEP grade 8 mathematics items across content areas was assessed similarly by the two groups.

Table 5: Distribution of items across NAEP mathematics content areas

Content Areas in the NAEP Framework	PISA 2012 ¹	NAEP 2013 Grade 8	NAEP 2011 Grade 8	TIMSS 2011
Number properties and operations	33%	19%	17%	28%
Geometry	14%	17%	17%	9%
Measurement	16%	19%	19%	12%
Data analysis, probability and statistics	27%	15%	14%	18%
Algebra	11%	30%	32%	34%

¹ This is based on the 64 (of 85) PISA items that were classified to the NAEP grade 8 framework.

Sources: Provasnik et al., 2013; AIR, 2013.

So, theoretically if students in Connecticut – where there appears to be a relative advantage in PISA – have had greater exposure to data analysis, probability, and statistics items or items of similar nature or complexity to PISA items, this might contribute to their relatively strong performance in PISA. On the other hand, if students in Massachusetts – where there is a relative advantage in TIMSS – have had a strong focus on algebra this could partly explain the excellence in TIMSS and NAEP. This could be explored by examining the state standards and assessments in place.

A second analysis examines states' scores on the mathematics subscales, which in PISA 2012 included three processes (employ, formulate, and interpret) and four content categories (space and shape, change and relationships, quantity, and uncertainty) to determine if states' relative strengths and weaknesses align with relative areas of emphasis or de-emphasis in the various assessments. For example, Connecticut was comparatively strong in items requiring interpretation, of which there were a larger percentage in PISA 2012 than in NAEP 2013 (see Exhibit 6). Items in the interpretation category were a relative strength for all states, however. In terms of the content subscales, there were again similar patterns among the PISA-participant states, with change and relationships (i.e., algebra) and uncertainty (i.e., probability and statistics) as relative strengths and quantity and space and shape as relative weaknesses. It is difficult to relate these results to item distributions in NAEP, however, because in the comparison study on which the data are based, a high percentage of NAEP 2013 items were found not to fit the PISA framework.

A third analysis relates to sampling. As PISA uses an age-based sample, sampled students may come from various grades, which is a distinction from TIMSS and NAEP. This feature of PISA is in keeping with

Table 6: Mathematics performance and percentage distribution of items by PISA process and content subscales

Employ		Process subscales			Content subscales			
		Formulate	Interpret	Space and shape	Change and relationships	Quantity	Uncertainty	
Mean score	Connecticut	502	504	515	487	515	502	512
	Florida	466	458	475	446	476	458	475
	Massachusetts	509	512	524	498	518	506	523
	United States	480	476	490	463	488	478	488
	OECD	493	492	497	490	493	495	493
Percentage	PISA 2012	4.4%	3.2%	2.5%	2.5%	2.5%	2.6%	2.5%
	NAEP 2013 ¹	66%	23%	9%	7%	14%	6%	7%

1 The percentages for NAEP items in the content categories will not sum to 100 because 66 percent of the NAEP eighth-grade items were found not to fit the PISA framework.

Source: AIR, 2013 and PISA International Data Explorer (<http://nces.ed.gov/surveys/international/ide/>).

its goal to measure the outcomes of learning, rather than schooling *per se* and provides a neutral comparison point internationally. Intra-nationally, in the case of federal systems with variation in local education policy, this can be a source of some differences. For example, analysing the grade distribution of the students who took PISA in 2012 shows that Connecticut had a larger percentage of students in the 11th grade and smaller percentages in the 9th and 10th grades than Florida, Massachusetts, or the United States overall (see Exhibit 7). Conversely, Florida had a larger percentage of students in the 9th grade and smaller percentages in the upper grades than the other systems. In other words, a larger percentage of Connecticut's students were exposed to an additional year of schooling than were U.S. students on average or in Massachusetts or Florida. And a larger percentage of Florida's students had not yet been exposed to 10th- or 11th-grade mathematics than had students in the other systems. This is due to differences in policies on school entry and in grade retention practices. For example, Connecticut has one of the youngest kindergarten entry ages in the United States, allowing students to enrol at 4 years old as long as they will be 5 years old by mid-school year (e.g., January 1) and requiring enrolment at 5 (ECS, 2014; see also Exhibit 2). Other states more typically have cut-offs early in the school year, requiring that students be 5 years old, e.g., by September 1 and not requiring enrolment until 6, as in Florida and Massachusetts. What may then account for Florida's higher

rate of 9th-grade PISA participants are generally higher early grade retention rates than in the other two states (Warren and Sariba, 2012). Of the analyses described, the sampling explanations appear to have the strongest explanatory potential.

Table 7: Distribution of PISA participants by grade: 2012

	7 th grade	8 th grade	9 th grade	10 th grade	11 th grade	12 th grade
Connecticut	*	*	7	59	34	*
Florida	*	*	21	67	12	*
Massachusetts	*	*	1	82	17	*
United States	*	*	12	71	17	*

* Reporting standards not met.

Note: Results for Connecticut, Florida, and Massachusetts are for public school students only. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.

Source: PISA International Data Explorer (<http://nces.ed.gov/surveys/international/ide/>).

A final analysis questions the differing country populations in PISA and TIMSS: how would states’ standings relative to the OECD/international means change if the assessments included the same group of countries? Restricting the countries in the analyses to only those that participated in both PISA 2012 and TIMSS 2011 at eighth grade,⁶ both the OECD and international averages drop. So, while this brings Florida’s mean score closer to the PISA OECD mean score (though still statistically significantly below it), it further distances the state’s mean score in a positive direction from the TIMSS international mean – essentially, leaving the relative standings unchanged.

Conclusion

U.S. states’ participation in international assessments shows one source of variation in national statistics and also allows states to benchmark themselves to international standards, as has been shown to be an increasing interest over at least the last decade. However, given that states also have access to national assessment data, as well as their own state data and, in some cases, two sources of international data, making sense of the results can be challenging. Analyses described in this paper suggest that opportunity to learn may be an important factor in differing results among assessments – with the amount of schooling related to states’ PISA per-

6 This represents 28 countries, with the only difference being the participation of all nations of the United Kingdom in PISA versus only England in TIMSS. The referenced analysis is based on data (not shown) obtained from the PISA International Data Explorer.

formance. A next frontier for state participants in international student assessments will be in how they, and their localities, may try to extend the use of data beyond the core benchmarking function to absorb lessons from international partners and inform education policy.

References

- American Institutes for Research (2013) *A comparison study of the Program for International Student Assessment (PISA) 2012 and the National Assessment of Educational Progress (NAEP) 2013 mathematics assessments*. Washington, D.C.: Author.
- Chingos, M.M. (2012) *Strength in numbers: State spending on K-12 assessment systems*. Washington, D.C.: Brown Center on Education Policy, The Brookings Institution. Available from: http://www.brookings.edu/~media/research/files/reports/2012/11/29%20cost%20of%20assessment%20chingos/11_assessment_chingos_final.pdf (Accessed 7th July 2014).
- Danitz, T. (2001, 27th February) Special report: States pay \$400 Million for tests in 2001. *Stateline*. Pew Center on the States.
- Education Commission on the States (2014) Early learning: Kindergarten online database [data set]. Available from: http://www.ecs.org/html/educationissues/kindergarten/kdb_intro_sf.asp (Accessed 7th July 2014).
- Education Commission on the States (2014) What Governors Need to Know: Highlights of State Education Systems. Available from: <http://www.ecs.org/clearinghouse/85/69/8569.pdf> (Accessed 7th July 2014).
- Federal Education Budget Project (2014) Available from: <http://feb.newamerica.net/background-analysis/no-child-left-behind-funding> (Accessed 7th July 2014).
- Kelly, D., Xie, H., Nord, C.W., Jenkins, F., Chan, J.Y., and Katsberg, D. (2013) *Performance of U.S. 15-year-old students in mathematics, science, and reading literacy in an international context: First look at PISA 2012* (NCES 2014-024). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Available from: <http://nces.ed.gov/pubs2014/2014024rev.pdf> (Accessed 7th July 2014).
- Linn, R.L., Baker, E.L., and Betebenner, D.W. (2002) Accountability systems: Implications of the requirements of the No Child Left Behind Act of 2001. *Education Researcher*. 31 (6), pp. 3-16.
- Provasnik, S., Lin, C., Darling, D., and Dodson, J. (2013) *A comparison of the 2011 Trends in International Mathematics and Science Study*

- (TIMSS) assessment items and the 2011 National Assessment of Educational Progress (NAEP) frameworks. Washington, D.C.: AVAR. Available from: http://nces.ed.gov/nationsreportcard/subject/about/pdf/naep_timss_comparison_items.pdf (Accessed 7th July 2014).
- Mullis, I.V.S, Martin, M.O., Foy, P., and Arora, A. (2012) *TIMSS 2011 international results in mathematics*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- National Centre for Education Statistics (2012) *Mathematics 2011: National Assessment of Educational Progress at grades 4 and 8, NCES 2012-458*. Washington, D.C.: NCES, Institute for Education Sciences, U.S. Department of Education.
- National Centre for Education Statistics (2014) Tables and Figures. Available from: <http://nces.ed.gov/quicktables/index.asp>. (Accessed 29th September 2014).
- National Centre for Education Statistics (NCES) (2013). *U.S. states in a global context: Results from the 2011 NAEP-TIMSS linking study, NCES 2013-460*. Washington, D.C.: NCES, Institute for Education Sciences, U.S. Department of Education. Available from: <http://nces.ed.gov/nationsreportcard/subject/publications/studies/pdf/2013460.pdf>.
- National Governors Association (2008) *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, D.C.: NGA. Available from: <http://www.corestandards.org/assets/0812BENCHMARKING.pdf> (Accessed 7th July 2014).
- OECD (2013) *Strong performers and successful reformers in education – Lessons from PISA 2012 for the United States*. Paris: OECD. Available from: <http://www.oecd.org/pisa/keyfindings/PISA2012-US-CHAP4.pdf> (Accessed 7th July 2014).
- Schneider, M. (2009) The international PISA test. *Education Next*, 9, 4. Available from: <http://educationnext.org/the-international-pisa-test/> (Accessed 19th June 2014).
- Warren, J.R., and Saliba, J. (2012, May). *Public school grade retention rates in the United States: Estimates by state, grade, year, and race/ethnicity*. Paper presented at the Population Association of America Annual Meeting, San Francisco, CA.