PIONEER
EDUCATION

Student Achievement: MCAS and International Exams

By Ken Ardon

WHITE PAPER
No. 275 | September 2024

PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

MISSION

Pioneer Institute develops and communicates dynamic ideas that advance prosperity and a vibrant civic life in Massachusetts and beyond.

Vision

Success for Pioneer is when the citizens of our state and nation prosper and our society thrives because we enjoy world-class options in education, healthcare, transportation, and economic opportunity, and when our government is limited, accountable, and transparent.

Values

Pioneer believes that America is at its best when our citizenry is well-educated, committed to liberty, personal responsibility, and free enterprise, and both willing and able to test their beliefs based on facts and the free exchange of ideas.

PIONEER EDUCATION

This paper is a publication of Pioneer Education, which seeks to increase the education options available to parents and students, drive system-wide reform, and ensure accountability in public education. The Center's work builds on Pioneer's legacy as a recognized leader in the charter public school movement, and as a champion of greater academic rigor in Massachusetts' elementary and secondary schools. Current initiatives promote choice and competition, school-based management, and enhanced academic performance in public schools.

PIONEER HEALTH

Pioneer Health seeks to refocus the Massachusetts conversation about healthcare costs away from government-imposed interventions, toward market-based reforms. Current initiatives include driving public discourse on Medicaid, presenting a strong consumer perspective as the state considers a dramatic overhaul of the healthcare payment process, and supporting thoughtful tort reforms.

PIONEER TRANSPORTATION

Pioneer Transportation seeks reforms that allow commuters to take full advantage of the coming mobility revolution — with access to a range of affordable and on-demand public and private transportation options, as well as transit-friendly real estate development.

PIONEER OPPORTUNITY

Pioneer Opportunity seeks to keep Massachusetts competitive by promoting a healthy business climate, transparent regulation, small business creation in urban areas, and sound environmental and development policy. Current initiatives promote market reforms to increase the supply of affordable housing, reduce the cost of doing business, and revitalize urban areas.

Table of Contents

Executive Summary

The Massachusetts Education Reform Act of 1993 made substantial changes to the Commonwealth's K–12 system. Among those changes was the creation of the Massachusetts Comprehensive Assessment System, or MCAS. MCAS scores across most exams increased during the 2000s, but the MCAS does not provide information about how Massachusetts students performed relative to their peers outside the Commonwealth. National and international exams can provide a helpful comparison to understand the rising MCAS scores.

The National Assessment of Educational Progress demonstrated that passage of Ed Reform contributed to strong student performance that gave Massachusetts the best schools in the country. From 1999 to 2015, Massachusetts students also participated in periodic international exams that allow a comparison of their performance to their peers overseas. The results from two international exams, Trends in Math and Science Study and the Program for International Student Assessment, demonstrated that not only was student achievement in Massachusetts improving on the MCAS, but it also improved relative to students in the rest of the US and in other countries. By 2015, Massachusetts had some of the best schools not only in the country but also in the world. This is truly a success story for the Commonwealth.

Since 2015, changes in the curriculum and the MCAS have made it more difficult to track student progress. Unfortunately, the Commonwealth has not participated in either international exam since that time, and we cannot evaluate how the changes affected the performance of Massachusetts students relative to students overseas. Rejoining one or both international exams would provide a useful measuring stick for Massachusetts schools, and it could also help set a high goal as we strive to be among the best in the world. Regular assessment on the MCAS played a part in improving Massachusetts schools, and international exams could serve the same purpose by providing feedback on the quality and rigor of the MCAS and curriculum.

Introduction

The Massachusetts Education Reform Act of 1993 (MERA) made substantial changes to the Commonwealth's K–12 system. It created a new state funding formula to ensure that poor districts have adequate funds, established curriculum standards, mandated assessments, and developed an accountability system.[1] Passage of MERA contributed to strong student performance that gave Massachusetts some of the best schools in the country and made its students internationally competitive in math and science.[2]

High-quality schools lead to increased student equity and a stronger economy. The most obvious benefit for students is that better academic performance is likely to lead to higher wages, but there are other effects as well: students with higher test scores are less likely to drop out, more likely to complete college, less likely to commit crimes or need government assistance, more likely to have a successful marriage, and they lead healthier lives. There is even evidence that better education increases happiness.

Education also affects economic growth for several reasons. Businesses are more likely to move to an area with good schools, and education can contribute to innovation. Research shows that the quality of schools is particularly important; economic growth depends more on the amount students learn rather than simply the number of years they spend in school.

This paper will review overall student performance as well as the performance of student subgroups on the assessment system developed in response to MERA, the Massachusetts Comprehensive Assessment System (MCAS). To give a broader context and better understanding of student performance, it will also explore how the Commonwealth's students compare to those in the rest of the United States and in other countries.

This paper will review overall student performance as well as the performance of student subgroups on the assessment system developed in response to MERA, the Massachusetts Comprehensive Assessment System (MCAS).

Measuring Student Performance

MCAS provides annual information on the performance of both students and schools. The exams focus primarily on English and math, as well as science and technology/engineering (Sci/Tech)[3]. The exam is closely tied to the state's curriculum standards, and as standards change the MCAS is adjusted. The most notable change is that from 2017–2019, most exams were updated to the Next Generation MCAS which is partially based on the Common Core standards and tests.

While the MCAS measures performance compared to state standards, it may not provide a full indication of how much Massachusetts students learn or how they compare to students in other states or countries. This could happen for several reasons. One is that students learn material that is not on the MCAS, so that exam scores do not capture all learning. A more worrying issue is that that students in Massachusetts could meet state standards, but if the standards are too easy, then students may not learn as much as students elsewhere.[4] Evaluations of the accuracy or rigor of the MCAS are particularly important when the curriculum or the MCAS exams change, as in the switch to the Next Generation MCAS.

Researchers generally agree that Massachusetts' standards have been rigorous, but it can be helpful to evaluate the performance of students against an outside standard. Comparing students in Massachusetts to students in the rest of the United States or against students in other countries can not only confirm the rigor of the MCAS, but the comparison can also provide meaning to MCAS scores and ensure that they accurately measure student performance. For example, if the average MCAS scores change by five points, it may be hard to interpret the increase; knowing how performance changed relative to students outside Massachusetts could provide context for the change.

Comparing students in Massachusetts to students in the rest of the United States or against students in other countries can not only confirm the rigor of the MCAS, but the comparison can also provide meaning to MCAS scores and ensure that they accurately measure student performance.

Exams such as the SAT or the National Assessment of Educational Progress (NAEP) allow a comparison across the United States. A recent paper published by Pioneer Institute, "MCAS, NAEP, and Educational Accountability," discussed whether high scores on the MCAS translated into other exams. As the paper showed, Massachusetts students consistently rank at or near the top on the NAEP, and SAT scores rose through 2019.

Another way to evaluate the performance of students in Massachusetts is to compare them to their peers overseas. Many businesses in Massachusetts compete with foreign companies, either against imports to the United States or by exporting products and services. The workers in these Massachusetts companies must in some sense compete with workers from other countries. If Massachusetts schools lag behind schools overseas, our economy may suffer.

There are two primary international exams given at regular intervals, the Trends in International Math and Science Study (TIMSS), and the Program for International Student Assessment (PISA). As its name implies, TIMSS tests students on math and science, while PISA tests students on math, reading, and science. This paper evaluates the MCAS performance of Massachusetts students from 1999 until just before the COVID pandemic in 2019, and also on TIMSS and PISA from 1999 to 2015.[5]

Performance During the First 20 Years of MCAS

MERA created MCAS in 1993, and the first exams were administered by the end of the decade. MCAS started with 3 subjects being tested in 3 grades: English language arts (ELA), math, and science and technology/engineering (Sci/Tech) in grades 4, 8, and 10. Overhauls in 2001–2003 and 2006–2008 first added and then removed a history exam and also adjusted the grade levels for other exams (see Table 1). Since 2008, the MCAS has been relatively stable, with ELA and math tested in grades 3–8 and 10, and Sci/Tech in grades 5, 8, and 10.

Table 1: MCAS Exams by Year

Year	Exams[6]	Changes
1998	9	Added G4, G8, and G10 ELA, Math, and Sci/Tech
1999	10	Added G8 History
2000	10	No change
2001	10	Removed G4, G8, and G10 Sci/Tech Added G3 Reading, G6 Math, G7 ELA
2002	8	Removed G8 History and ELA
2003	10	Added G5 and G8 Sci/Tech
2004–2005	10	No change
2006	16	Added G3, G5, and G7 Math; G5, G6, and G8 ELA
2007	16	No change
2008	17	Added G10 Sci/Tech
2009–2015	17	No change
2016	5	Many students took PARCC rather than MCAS
2017	17	NextGen MCAS for ELA and Math in grades 3–8, legacy MCAS for Sci/Tech and G10
2018	17	No change
2019	17	NextGen MCAS for all exams except G10 Sci/Tech

There are two primary international exams given at regular intervals, the Trends in International Math and Science Study (TIMSS), and the Program for International Student Assessment (PISA). As its name implies, TIMSS tests students on math and science, while PISA tests students on math, reading, and science.

There are different ways to measure student performance, but this paper will focus on the percentage of students in the top two categories (out of four) on each year's exam[7]. The names of the top 2 categories changed during this period from "proficient" and "advanced" on the original MCAS to "meeting" or "exceeding" expectations on the Next Generation MCAS. The accompanying graphs show the percentage of students in the top two categories on ELA, math, and Sci/Tech for various grades each year. Not every test was given in each year, as explained above, and the first year of the Next Generation exams for each subject/grade is indicated by the arrows. The circles indicate the years when the scoring was most comparable, between 2001, when there was a large jump in performance that will be discussed below, through 2015, the last year of the legacy MCAS on most exams.

Figure 1 - ELA performance over time, selected grades

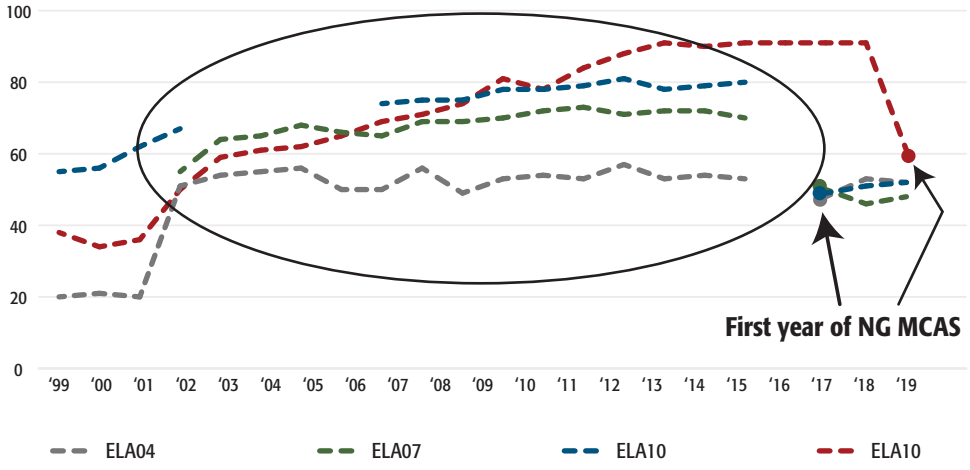


Figure 2 - Math performance over time, selected grades

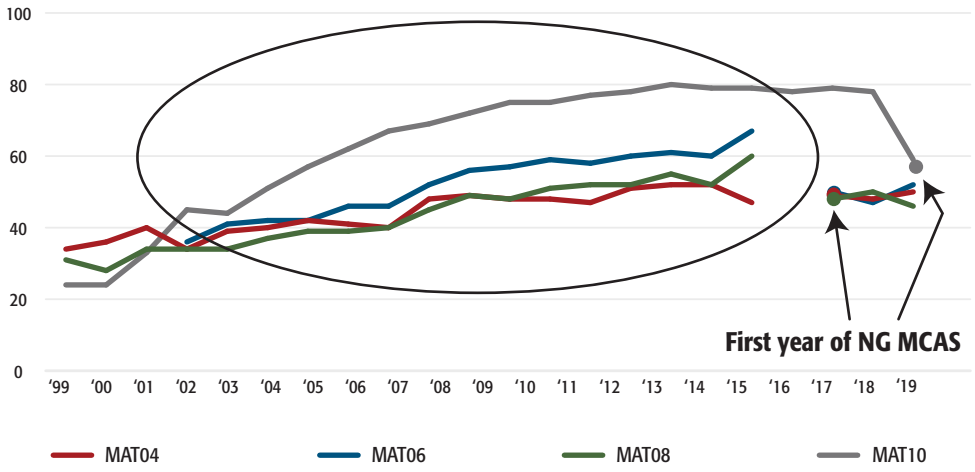
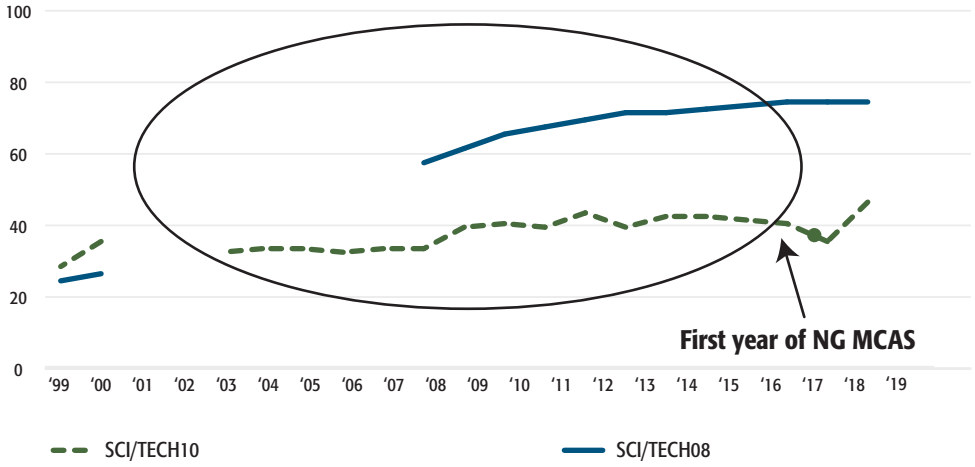


Figure 3 – Sci/Tech performance over time, selected grades



Scores on statewide exams (meaning the percentage of students in the top two categories) will generally change slowly if exams are consistent and accurate because we should not expect a substantial change in the performance of tens of thousands of students in one year. There are two discontinuities visible in the scores that contradict this expectation. The first is that there were large jumps in scores on three exams in 2001: 31 percentage points on ELA04, 14 percentage points on ELA 10, and 12 percentage points on MAT10.

There are two likely explanations for these large increases. One is that the 10th graders who took the test in 2001 were the first to face passing the MCAS as a graduation requirement. Before 2001, students may not have taken the test seriously[8]. In grade 4, when the graduation requirement is not likely to provide motivation, the magnitude of the increase suggests that grading standards must also have changed because it strains credulity to believe that student performance jumped this quickly[9].

The second and more pronounced discontinuity is that the switch to the Next Generation MCAS led to large declines in the percentage of students scoring well on almost every exam. These changes are summarized in Table 2. Most of the exams show the change from 2015 to 2017, but for the 8th grade Sci/Tech exam and the 10th grade ELA and math exams the changes are from 2018 to 2019. The 10th grade Sci/Tech exam had not switched as of 2019.

Table 2 - Change in Scores on Introduction of Next Generation MCAS

Exam	Grade	Last Year - Legacy MCAS	First Year - Next Gen MCAS	Change
Reading/ELA	3	60	47	–13
ELA	4	53	48	–5
ELA	5	71	49	-22
ELA	6	71	51	–20
ELA	7	70	50	–20
ELA	8	80	49	–31
ELA	10	91	61	–30
Math	3	70	49	–21
Math	4	47	49	2
Math	5	67	46	–21
Math	6	62	50	–12
Math	7	51	47	–4
Math	8	60	48	–12
Math	10	78	59	–19
Sci/Tech	8	35	46	11

In most grades, the changes are quite large, with declines of 10, 20, or even 30 percentage points. The odd exceptions are for the 4th grade ELA and math exams and the 7th grade math exam, where scores barely changed, and the Sci/Tech exam, where scores improved (grades with small or positive changes are shaded).

The drop in scores does not mean that student performance declined — a drop in student learning is probably the least likely explanation of the decline in test scores. Instead, the decline is likely caused by the change in the test, which could affect scores in several ways.

One potential reason for the decline is that the Next Gen MCAS uses different designations for the four performance levels[10]. However, that explanation does not hold up if the terms are read literally — if students were "proficient" or better in 2015, one would think that they would also have been "meeting expectations," and if a student is "meeting expectations" then one would hope that they are also "proficient" or better.

The more likely explanations for the decline in scores are that the new exam tested different material, and the grading could be more rigorous. The Next Gen MCAS contains new questions, including on areas such as critical thinking, and we would expect scores to change somewhat. At the same time, the state may have raised expectations – i.e. the test could have gotten harder. Whatever the explanation, the result is that it became more difficult for students to score highly on the Next Generation MCAS, which makes comparisons across time difficult.

While scores on most exams changed substantially when the state transitioned to the Next Gen MCAS, not all of them did. Scrutinizing exams with little change in scores is particularly interesting. In both 3rd and 5th grade, the percentage of students with high scores fell substantially, while there were only small changes in 4th grade. Taken at face value, it would be hard to understand how the performance of students in both 3rd and 5th grade fell, while performance for students in 4th grade did not. A simple explanation is that some exams were made more difficult by adjusting the grading scale to correct inconsistencies on the legacy MCAS, when the 4th grade exams and the 7th grade math exam appeared to be more difficult than other exams.

This anomaly shows up on the old MCAS if we look at performance across grades; students in 4th grade performed much worse than students in both 3rd and 5th grade. For example, in 2015, the percentage of students in the top two categories on the ELA/reading exams in 3rd through 5th grade went from 60 percent to 53 percent to 71 percent, while on the math exams they went

Students in both 3rd and 5th grade fell, while performance for students in 4th grade did not. A simple explanation is that some exams were made more difficult by adjusting the grading scale to correct inconsistencies on the legacy MCAS, when the 4th grade exams and the 7th grade math exam appeared to be more difficult than other exams.

from 70 percent to 47 percent to 67 percent. It strains credulity to believe that these patterns were accurate representations of student performance. On the Next Generation MCAS, the comparable scores reflect much more consistent and believable performance; on ELA/reading the 3rd through 5th grade scores are now 47 percent-48 percent-49 percent while on math they are 49 percent-49 percent-46 percent.

Math scores in middle school show a similar pattern. In 2015, the percentage of students scoring in the top two categories on 6th, 7th, and 8th grade math exams went from 62 percent to 51 percent to 60 percent. On the Next Generation MCAS, the corresponding percentages are 50 percent-47 percent-48 percent. In other words, the Next Generation MCAS appears to have more consistent scoring across grades than the legacy MCAS.

The unusual pattern of scores across grades on the old MCAS would have had limited importance when evaluating an entire school or district. Even with a more difficult exam in 4th grade than 5th grade, one could look at whether scores improved or fell in each grade over time. However, the inconsistent pattern may have had an impact on 4th or 7th grade students or teachers who were being evaluated for the absolute level of performance rather than changes in scores.

What Happened to Overall Scores – Did Performance Improve?

Beyond the discontinuities, there are two important things to notice about overall MCAS scores. First, scores on the MCAS generally increased over time — i.e. student achievement improved. In particular, between the unusual jump in 2001 and the switch to the Next Generation MCAS, the percentage of students with high scores increased on many exams – i.e. from 2001 to roughly 2015 for most exams (the period circled on the graphs). As Table 3 shows, scores seven exams rose substantially while scores on three tests showed little change.[11]

Table 3 - Change in % Proficient or Advanced from 2001 to Last Year of Legacy MCAS[12]

	ELA04	ELA07	ELA08	ELA10	MAT06	MAT08	MAT10	Sci/Tech08	Sci/Tech08	Sci/Tech08
Initial Score (post 2001)	51	55	67	50	34	36	34	45	32	57
Last Score (2015 or 18)	48	50	80	91	47	67	60	78	35	74
Change	**–3**	**–5**	**13**	**41**	**13**	**31**	**26**	**33**	**3**	**17**

The gains in math were particularly impressive, as students in every grade saw substantial improvements. In addition to large improvements in math scores, the score on the 10th grade ELA exam also rose quite a bit. While there were fewer science exams, scores rose slightly in 8th grade and substantially in 10th grade. The overall pattern is clear, Massachusetts student performance improved, especially in math.

The second important fact about the MCAS scores is that the majority of 10th grade students scored proficient or advanced on all three exams (ELA, math, and science). Because of the rising scores, most students in the Commonwealth reached proficiency and graduated.[13]

Subgroups and the Distribution of Scores

While overall scores were relatively strong and improving from 2001 to 2015, different groups of students had different levels of success. Table 4 shows the gaps between Black and Hispanic students compared to white students. In 2001, the gap for Black students was slightly smaller than it was for Hispanic students, particularly in ELA. This is not surprising, as Hispanic students are more likely to have a first language other than English. In math, the gaps grow slightly larger among older students, suggesting that minority students were falling slightly farther behind in math in higher grades.

Table 4 - Racial Gaps, Various Years

	Black – White Gap			Hispanic – White Gap		
	4th grade	8th grade	10th grade	4th grade	8th grade	10th grade
ELA 2001	–34	–36	–31	–38	–41	–38
Math 2001	–28	–29	–35	–29	–31	–37
ELA 2015	–22	–19	–11	–29	–24	–16
Math 2015	–24	–29	–23	–25	–26	–29
ELA 2019	–26	–27	–32	–26	–29	–34
Math 2019	–28	–28	–31	–25	–28	–32

The table also contains data from both 2015 and 2019 to evaluate how gaps changed over time as well as whether the switch to the Next Generation MCAS affected the size of the gaps (recall that overall scores on the Next Generation MCAS are lower for most exams). In 2015, the gaps for both groups appeared to shrink substantially in ELA and more moderately in math, with particularly large improvements for older minority students on the ELA exam. The 2019 results confirm that the gap appears slightly smaller, but the size of the gains is much smaller for Black students. At the same time, the strong improvement for minority students on ELA exams for 10th grade students disappears. Overall, the state has made only modest progress closing racial gaps, although it is important to remember that overall scores for all students rose during this time.

Table 5 - Gap for Low-Income Students on ELA and Math Exams[14]

	4th grade	8th grade	10th grade
ELA 2001	–36.9	–38.7	–38.7
Math 2001	–28.2	–30.5	–32.3
ELA 2015	–29	–21	–11
Math 2015	–28	–28	–23
ELA 2019	–30	–33	–32
Math 2019	–30	–33	–34

Not surprisingly, table 5 shows that low-income students had lower scores. Because some low-income students are also immigrants who may not speak English as their first language, it might be expected that the gap in ELA would be larger than in math. This was true in 2001, but not in subsequent years. The performance gap was generally similar across 4th, 8th, and 10th grades.

The method of measuring income has changed over time, making it somewhat difficult to compare the size of the gap across years. However, we can still examine the patterns across grades and changes in those patterns. In 2015 the gap was smaller than in 2001, although this could have been caused by the changing definition of low-income. Importantly, this improvement came about while overall scores were rising, which means that scores for low-income students rose faster than scores for other students. The gaps that year also got smaller by 10th grade, which could be explained by the increased importance of 10th grade exams for graduation.

In 2015 the gap was smaller than in 2001, although this could have been caused by the changing definition of low-income.

Scores on the Next Gen MCAS in 2019 show a larger gap for all exams and all grades than on the legacy MCAS, but with the changes in the exam it is not clear how to interpret the changes. However, one clear difference between 2015 and 2019 is that low-income students were no longer catching up in higher grades.

Table 6 - Gender Gap (Female - Male), ELA and Math Exams

Female – Male	4th grade	8th grade	10th grade
ELA 2001	11	11	13
Math 2001	–2	–2	1
ELA 2019	8	19	12
Math 2019	–3	3	4

Table 6 shows the gender gap over time. Female students outperformed male students in ELA across all grades in both years, while in math their scores were similar. By 2019, female students were outperforming male students on every exam except 4th grade math. Both male and female students improved their performance, but women improved slightly faster. There is a hint that the gap between female and male students' performance may grow slightly larger as students age (i.e. that older female students may be pulling ahead of their male peers), but it is not conclusive.

As stated in "MCAS, NAEP, and Educational Accountability," the performance gaps indicate that the Commonwealth's schools are not serving all students well enough to help them reach a high standard.[15] The ability to measure the performance of these groups compared to their peers is a strong argument for the importance of the MCAS.

The performance gaps indicate that the Commonwealth's schools are not serving all students well enough to help them reach a high standard.[20]

Strong MA performance on NAEP

While MCAS scores generally indicate strong performance, there are a few reasons to compare Massachusetts students against an outside standard. As stated above, the relatively strong performance of students on the MCAS does not necessarily mean that they are performing well compared to students elsewhere — there is always the chance that the MCAS was and is an easy exam. The switch to the Next Generation MCAS also makes it difficult to measure or interpret long-term changes in scores because the newer exam appears to be harder in some grades. Finally, it can also be difficult to interpret the meaning of a rise in MCAS scores — e.g. if MCAS scores rise by five points, how does this translate into performance against students outside of Massachusetts?

The NAEP allows us to compare students in Massachusetts with their peers across the country. As "MCAS, NAEP, and Educational Accountability" showed, the strong performance on the MCAS translated to the NAEP. Not only did students in Massachusetts score very highly compared to their peers across the United States, but NAEP scores also rose during these years. These results confirm that the improvement in MCAS scores reflects real improvement in student achievement.

International Exams

The Trends in International Math and Science Study exam, or TIMSS, was established by an international association of research institutions and government agencies, the International Association for the Evaluation of Education Achievement. In 1995, TIMSS started as the *Third* International Mathematics and Science Study when students from approximately 40 countries took the exam. It has taken place every four years since and has expanded to more than 70 participants. Most participants are countries, but occasionally a smaller government entity such as a school district or state participates separately. TIMSS focuses on math and science and is usually given to students in 4th and 8th grades. Not all students in a country or state take the exam — TIMSS usually requires a sample of at least 5,000 students, which is used to estimate the performance of all students.

The Trends in International Math and Science Study exam, or TIMSS, was established by an international association of research institutions and government agencies, the International Association for the Evaluation of Education Achievement.

The Program for International Student Assessment, or PISA, is a regular study by the Organization for Economic Cooperation and Development (OECD) that tests 15-year-old students, regardless of the grade they are in.[16] The exams cover math, science, and reading. PISA began in

2000, takes place every three years, and focuses mostly on OECD countries although, just as with TIMSS, sometimes smaller areas such as a city or state participate on their own.

The United States has participated in TIMSS and PISA since they began. Massachusetts participated on its own in TIMSS three times, in 1999, 2007, and 2011, with 8th graders testing each year and 4th graders also participating in 2007. The Commonwealth participated in PISA twice, in 2012 and 2015, when roughly 1,700 students from 49 schools, mostly in 10th grade, took the exam.

The United States has participated in TIMSS and PISA since they began. Massachusetts participated on its own in TIMSS three times, in 1999, 2007, and 2011, with 8th graders testing each year and 4th graders also participating in 2007. The Commonwealth participated in PISA twice, in 2012 and 2015, when roughly 1,700 students from 49 schools, mostly in 10th grade, took the exam.

Table 7 – Massachusetts Participation in TIMSS and PISA

	TIMSS			PISA	
Year	**1999**	**2007**	**2011**	**2012**	**2015**
Grades	8	4, 8	8	10	10

These tests allow us to compare students and schools in Massachusetts to students and schools around the world. Had participation continued past 2015, they would have also provided an exterior reference to evaluate how the change in curriculum and the MCAS exam affected student performance, or how the learning loss from COVID compared.

TIMSS and PISA differ in several ways. TIMSS focuses on science and math, while PISA also tests reading. TIMSS primarily tests 4th and 8th grade students, while PISA tests 15-year-olds, roughly 70 percent of whom are in 10th grade with the rest split mostly between 9th and 11th grades.

Both exams report scores at different levels: advanced, high, intermediate and low for TIMSS, and high, medium, and low proficiency for PISA. They also scale their scores around a value of 500, but they do it in different ways that can lead to confusion. On the TIMSS, a score of 500 in any year is meant to represent the same performance as the average score of 500 in 1995. This means that the average on the TIMSS will usually vary from year to year. On the PISA, the test is scaled so that the average score each year is 500 and the standard deviation is 100.

Interpreting the scores can be confusing for several reasons, depending on how one wishes to evaluate the score. To avoid confusion about the meaning of the score itself, results are often reported as rankings — e.g. on the TIMSS 8th grade math exam in 1999, the US was the 19th ranked participant out of 38. Rankings are mostly clear and easy to understand, but there are some minor issues to consider before using them to evaluate performance. The first is that the scores are estimates generated from a sample of students, and as with any estimates, there is a margin of error. Small differences in scores could easily be due to random factors. For example, while the US ranked 19th, only 14 countries had scores that were *significantly* above the US score.[17] Another problem with using rankings is that the ranking illustrates only relative performance and not absolute performance, which means that the ranking can change if the participants or their scores change. If a very high performing country joined the exam in 2003, the ranking of the US students would fall even if their performance on the exam did not. Also, because the countries that take the TIMSS and PISA are different, a ranking of 10th on each exam would mean different things. Despite these issues, rankings on either exam can be used to illustrate performance if the ranking is interpreted carefully.

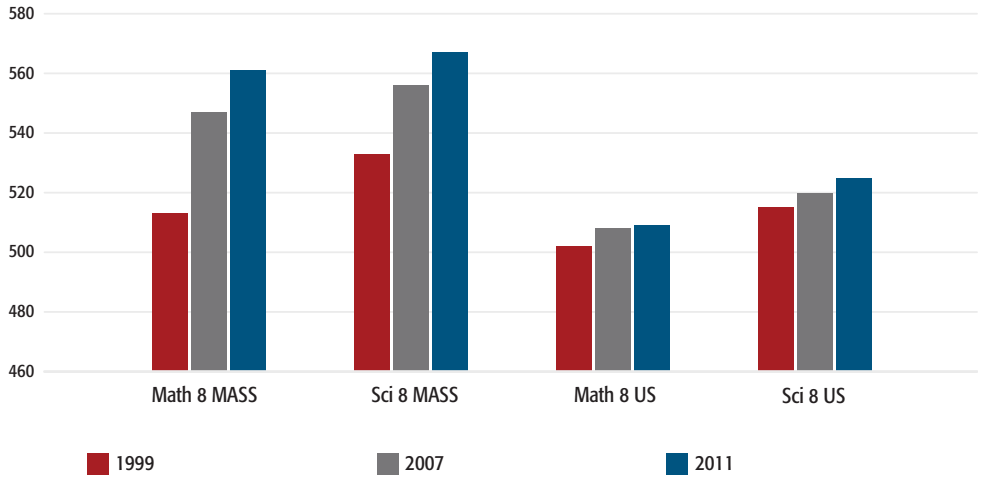
The situation becomes more complex when trying to evaluate the exam score rather than the ranking because the scores on the exams mean different things, as can changes in scores over time. The difficulties are easiest to explain using an example to illustrate the unusual scoring on the PISA.

Suppose the US scored exactly 500 one year on both exams. On the TIMSS, this score would mean that American students are performing at the same level as the average student in 1995, but the score would not tell us anything about how the US compares to other countries. If students in the US got a 500 again the following year, it would mean that their performance did not change. While the score of the US does not depend on the score of other countries, the ranking could easily change.

In contrast, a score of 500 on the PISA has a different meaning. It tells us that US students performed at the average of all students who took the exam *this year*, but it would not tell us anything about how student performance compares to prior years. By itself the difference does not seem problematic, but one can see the potential for confusion if we look at scores over time or what happens when additional countries take the exam. If the following year the US again scored 500, it would not necessarily mean that student performance hadn't changed. Even if the performance of US students does not change, the score could change based on the performance of other countries. If a high-performing country joined the PISA exam, or if the score in another country rose, the overall average performance would rise. This would mean that the score for US students would fall below 500, even if their performance stayed the same. In some sense, the PISA score is similar to the rank, where the number can change as other countries do better or worse.

Massachusetts Performance on TIMSS and PISA

Figure 4 - MA and US Scores on TIMSS, 8th Grade



As Figure 4 shows, Massachusetts outperformed the US on every 8th grade TIMSS exam. The 2007 results in 4th grade, which are not shown, were similar to the 8th grade results that year, with Massachusetts students scoring approximately 30–40 points higher than the US average.

As Figure 4 shows, Massachusetts outperformed the US on every 8th grade TIMSS exam. The 2007 results in 4th grade, which are not shown, were similar to the 8th grade results that year, with Massachusetts students scoring approximately 30–40 points higher than the US average.

Massachusetts not only scored higher than the US average, but performance also increased substantially and more rapidly than in the rest of the country. The rapid growth meant that the gap between Massachusetts and the rest of the US widened considerably. In 1999, Massachusetts only scored 11 points higher on math and 18 points higher on science. By 2007 the gap was greater than 35 points on both exams, and by 2011 it was 52 points in math and 42 points in science, despite gains in math performance during this time by the US as a whole. Massachusetts students' scores were also significantly better than the overall TIMSS average, particularly as the Commonwealth's performance improved in the early 2000s.

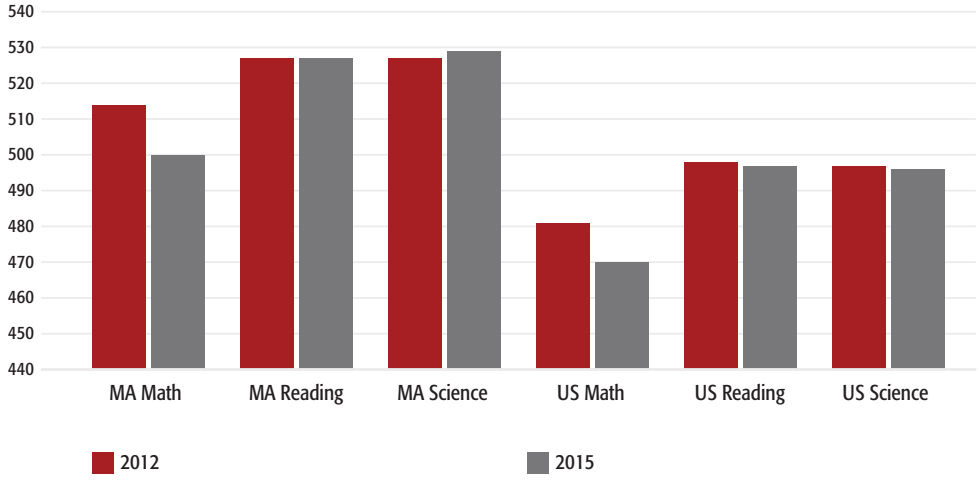
The high scores translate into extremely strong rankings, as shown in Table 9. In 1999, 8th grade students from 15 countries scored higher than students in Massachusetts on math and from 11 countries on science. In 2007 and 2011, after the dramatic improvement documented above, students from only a handful of countries would outperform Commonwealth students. The strong performance of 8th graders also showed up in 4th grade students, who proved to be some of the best in the world.

Table 8 - Ranking of Massachusetts Students on TIMSS

	Foreign Countries	Ranking of Massachusetts			
		Math 08	Sci 08	Math 04	Sci 04
1999	38	16	12		
2007	48	6	4	4	2
2011	42	6	2		

On the PISA, Massachusetts again outperformed the US average on every exam (Figure 5). Massachusetts participated only twice, in 2012 and 2015.

Figure 5 - MA and US Scores on PISA Exams, 2012 and 2015



On the PISA, Massachusetts again outperformed the US average on every exam (Figure 7). Massachusetts participated only twice, in 2012 and 2015, and the only noticeable change over time is the decline in the math scores for both Massachusetts and the US. On the MCAS during these years, the scores of 10th grade students on the math exam were flat. However, as explained earlier, because of the way that PISA scales the test scores, the decline in scores may not represent a decline in performance and could be caused by a change in the performance of students in other countries.

Table 9 – Rank of Massachusetts and US on PISA, 2012 and 2015[18]

	2012 (65 countries)						2015 (72 countries)					
	Science		Reading		Math		Science		Reading		Math	
	US	MA	US	MA	US	MA	US	MA	US	MA	US	MA
Higher	23	7	20	4	30	10	41	2	45	1	31	12

Of the 65 countries that took the exam in 2012, Massachusetts was in the top 10 on all exams, with the strongest performance was in reading.

The rankings in Table 9 illustrate how Massachusetts compared to the rest of the world on the PISA exams, and how relative performance changed over time even if the scores were relatively stable. Of the 65 countries that took the exam in 2012, Massachusetts was in the top 10 on all exams, with the strongest performance was in reading. The high scores of the Commonwealth's students in reading is a unique piece of additional information that PISA provides compared to TIMSS. Massachusetts rankings in math were lower than on the TIMSS, but part of this difference is likely because PISA had almost twice as many participants as TIMSS.

The changes in ranking by 2015 are interesting. Despite little change in science and reading scores, the rankings improved—Massachusetts students rose to first in reading and second in science among 72 countries. The science ranking closely tracks the TIMSS results. At the same time, while math scores fell, the ranking only declined slightly between 2012 and 2015.

The results on the international exams, particularly on TIMSS, generally confirm the results from Figure 2 and Figure 3, which showed MCAS scores rising during the early 2000s. The

TIMSS and PISA results demonstrate that not only was student achievement in Massachusetts improving, but it also improved relative to students in the rest of the US and in other countries. By 2015, Massachusetts clearly had some of the best schools in the world. This is truly a success story for the Commonwealth. Unfortunately, because the Commonwealth has not participated in either exam since 2015, we cannot evaluate how changes in the curriculum and MCAS exam affected the performance of Massachusetts students relative to an international benchmark. We know that scores declined on the NextGen MCAS, but that is likely caused at least in part by higher standards; we do not know whether Massachusetts students have lost their lead or fallen behind students in other countries.

By 2015, Massachusetts clearly had some of the best schools in the world. This is truly a success story for the Commonwealth.

Subgroup Scores

The discussion of TIMSS and PISA above refers to the average score. However, both exams also provide information about the distribution of scores and the scores of various subgroups, and the TIMSS includes data that compares the experience and attitudes of students in different countries.

While Massachusetts students perform well overall, we can also examine the distribution of scores across students. Two countries/participants on an exam could have the same average, but one might have most students scoring near the average while the other might have a large disparity between high- and low-performing students. We can look at the distribution of scores among participants with similar overall averages to determine whether Massachusetts' strong performance is driven by consistently good performance or especially high scores at the top.

There is mixed evidence about the distribution of scores. On the TIMSS, the gap between high- and low-performing Massachusetts students was smaller than the gap in other participants with similar overall average scores. In other words, the range of scores in Massachusetts was narrower than in other high-performing participants, so that Massachusetts students do well, but fewer have exceptionally high or low scores. This pattern holds in both 2007 and 2011 across math and science in both 4th and 8th grade, although it is stronger on some exams than others. However, the relationship does not appear on the PISA exams in 2012 and 2015, suggesting that it is either a weak relationship, it only holds on the TIMSS exams, or it fades out by 10th grade.

Some of the subgroup data is difficult to interpret, such as the scores segregated by parental education, socioeconomic status (SES), or language ability. Not surprisingly, TIMSS results show that students who had a parent complete university score higher than students who did not. The size of the gap across 8th grade math and science in 2007 appears to be slightly larger in Massachusetts than in the rest of the country or in the overall international average. While students in Massachusetts outscored the international average on the two exams by about 50 points each, the gap was roughly 17 points larger for students with a parent holding a college degree than for those with only a high school diploma. In other words, Massachusetts students with highly educated parents outperform their international peers by more than students with less educated parents. It is not clear why students with more highly educated parents appeared to have a larger advantage in Massachusetts than in other countries.

The data on performance by family SES or home language shows no obvious patterns. It is difficult to accurately measure differences in family income or SES across countries, but for what its worth, on PISA in 2015 the gaps between high and low SES students seem to be roughly similar in Massachusetts and other locations. On TIMSS, the performance gap for students who do not speak the language of the exam at home is slightly larger in science than in math in both 2007 and 2011 (see Table 11). The gap also appears to be larger in Massachusetts than in other countries.

The performance by gender on the exams is clearer: Massachusetts has had relatively stronger performance among boys than girls across exams and grades, and over most years of testing. Recall that on the MCAS, girls outperformed boys on ELA exams in both 2001 and 2019, while boys and girls had relatively similar scores on math exams. Also, girls improved relative to boys,

The performance by gender on the exams is clearer: Massachusetts has had relatively stronger performance among boys than girls across exams and grades, and over most years of testing.

with the gender gap getting larger (in girls favor) in ELA and girls starting to outperform boys on many math exams as well.

Table 11 - Gap in Scores by Language Spoken at Home[19]

	Massachusetts	International Average
2007 Math 4th	43	33
2007 Science 4th	56	45
2011 Math 8th	39	26
2011 Science 8th	75	33

On TIMSS the pattern starts out in a similar fashion, but by 2011 things have changed. In 1999, girls generally performed worse than boys on both math and science in 8th grade, and the gap in Massachusetts was roughly the same size as the international gap. Over the next 12 years, girls improved relative to boys, especially overseas. By 2011, the international average score for girls was above that of boys, while in Massachusetts boys still outperformed girls (this differs from the results on MCAS).

Table 12 - TIMSS Gender Gaps, 8th Grade Exams, Female - Male Scores

	1999			2011		
	Mass	**US**	**Int**	**Mass**	**US**	**Int**
Math	–7	–7	–4	–5	–3	4
Science	–13	–19	–15	–6	–11	6

This does not mean that the performance of Massachusetts girls fell on TIMSS — the opposite is true and girls' scores in Massachusetts grew *faster* than girls' scores in other countries, as shown in Table 13. However, the performance of Massachusetts boys was especially strong compared to their peers, with scores on both math and science surging while they were flat or fell in the rest of the United States and in other countries. Scores for girls in Massachusetts grew slightly faster than scores for boys, which confirms the pattern from the MCAS. The difference is that the improvement was not enough for girls to overtake the boys on TIMSS in Massachusetts.

Table 13 - Change in 8th Grade TIMSS Scores by Gender

	Mass		U.S.		Intl.	
	Boys	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**
1999 Math	517	510	505	498	489	485
2011 Math	563	558	511	508	465	469
Change	**46**	**48**	6	10	–24	–16
1999 Science	540	527	524	505	498	480
2011 Science	570	564	530	519	474	480
Change	**30**	**37**	6	14	–21	0

On the PISA exams, performance by gender on the reading exams certainly matches what we would expect based on MCAS, with girls scoring higher than boys. Girls were generally behind in math but almost even in science, which is not far from the TIMSS results. Between 2012 and 2015, the math gender gap in Massachusetts did not change much, while girls fell further behind in science and their lead narrowed substantially in reading. In other words, boys outperformed girls in Massachusetts on the PISA exams from 2012 to 2015, with the largest change in reading. The same pattern showed up in the international averages.

Table 14 - Gender Gaps on PISA Exams, Female – Male Scores, 2012 and 2015

	Mass			Intl		
	Math	SCI	Reading	Math	SCI	Reading
2012	–10	3	32	–11	–1	38
2015	–9	–10	18	–8	–4	57
Change	1	–7	–14	3	–3	–11

In addition to providing information on the distribution of scores and results by gender, parental education, and income, TIMSS and PISA provide other interesting data or survey results. Among the notable pieces of information for Massachusetts:

- In fourth grade in Massachusetts and the rest of the United States, fewer students than in other countries reported that they spent a large amount of time on math homework. At the same time, fewer also reported that they do little homework. If these statements are accurate, American students appeared to be more likely to have a moderate amount of HW in 4th grade. However, by 8th grade students in both Massachusetts and the United States reported increased amounts of homework, enough so both groups reported more homework than did students in the rest of the world. It is not clear if the amount of HW changed or if student perceptions changed.
- Roughly two-thirds of 4th graders in Massachusetts reported highly positive attitudes towards math, slightly below the international average of 72 percent. By 8th grade student attitudes had deteriorated substantially, especially in Massachusetts: only 41 percent of Massachusetts students had a positive attitude, compared to 54 percent internationally.
- Despite their dislike of the subject, approximately 80 percent of 8th grade students in Massachusetts and elsewhere placed a high value on the importance of math.
- One finding confirms stereotypes about American children—both 4th and 8th grade students in Massachusetts have substantially more confidence in their math abilities than students overseas. Almost three-quarters of local 4th graders were highly confident in their ability in 2011 compared to 57 percent among all participants. By 8th grade the percentages have fallen, but students in Massachusetts are still much more confident than the international average, 60 percent to 43 percent.
- Confidence is especially prevalent among boys, both in Massachusetts and overseas. Internationally, about 5 percent more boys than girls are confident across 4th and 8th grade, while in Massachusetts the gap in favor of boys is roughly 8 percent.

Roughly two-thirds of 4th graders in Massachusetts reported highly positive attitudes towards math, slightly below the international average of 72 percent.

Conclusion

Massachusetts participated in TIMSS in 1999, 2007, and 2011, and in PISA in 2012 and 2015. The results from those exams provided a useful check of how students in Massachusetts stacked up against students around the world, as well as whether performance gains on MCAS had translated to gains against peers overseas.

The strong performance on the international exams across several years and subjects, especially on TIMSS, confirmed the quality of Massachusetts K–12 schools. The results attract workers and business to the Commonwealth; employees know that their children will have the opportunity to attend high-quality schools, and employers know that they will be able to find well-educated employees.

The strong performance on the international exams across several years and subjects, especially on TIMSS, confirmed the quality of Massachusetts K–12 schools.

Since 2015, the last year that the Commonwealth participated in either exam, the state changed curriculum and shifted to the Next Generation MCAS. The changes make it more difficult to evaluate student performance over time. Outside benchmarks, such as the results of NAEP and the international exams, provide useful context to scores on the MCAS and Next Generation MCAS, and also allow the public to evaluate how the changes to the curriculum affect Commonwealth students.

While the NAEP is useful, it is not clear whether comparing Massachusetts to other states sets a high enough goal. While of course we would like our schools to be among the best in the country, we also want them to be among the best in the world. We recommend that the Commonwealth rejoin one or both international exams. The relatively modest expense would be a small part of the billions of dollars spent on K–12 educate each year. Since the 1993 education reform law, regular assessments for our students have played a part in improving schools and districts by providing feedback and allowing for accountability. TIMSS and PISA exams would serve the same purpose by assessing the quality and rigor of the MCAS and curriculum.

References

- Candal, 2024, "MCAS, NAEP, and Educational Accountability." Pioneer Institute.
- Haney, W., 2000, "The Myth of the Texas Miracle in Education." Education Policy Analysis Archives, 8, 41. https://doi.org/10.14507/epaa.v8n41.2000
- PISA results and reports for multiple years, published by the National Center for Educational Statistics. Retrieved from https://nces.ed.gov/surveys/pisa/
- TIMSS Assessment Reports for multiple years, published by the TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from https://timssandpirls.bc.edu/timss-landing.html

Endnotes

1 Candal, 2024.

2 For example: https://www.edweek.org/policy-politics/map-a-f-grades-rankingsyfor-states-on-school-quality/2021/09 and https://www.usnews.com/news/best-states/rankings/education.

3 Students also took a short-lived history exam from 1999 to 2002.

4 State assessment scores can also be inaccurate or misleading for other reasons. For example, in the 1990s scores on the Texas standardized test improved, but the rise in scores may have reflected manipulation of which students took the exam rather than an improvement in student performance—see Haney, W. (2000).

5 The impact of COVID on learning is beyond the scope of this paper.

6 The count of yearly exams includes both MCAS and Next Generation MCAS exams.

7 The initial categories were advanced, proficient, needs improvement, and warning/failing. On the Next Generation MCAS, the categories were changed to exceeding, meeting, partially meeting, or not meeting expectations. For simplicity, the paper will refer to the percentage of students in the top two categories as a "score."

8 A high-stakes exam/graduation requirement can also lead to unintended consequences such as schools holding more students back to reduce the number of students who do not pass the test. This does not appear to be the case in Massachusetts. While there was an increase in the 9th grade retention in 2001, by the 2010s the retention rate was well below the rate in the late 1990s.

9 The change in ELA scores in 4th grade brought them more in line with scores in later grades. In the late 1990s and 2000, ELA scores averaged 20% in 4th grade and 58% in 8th grade. From 2001–2005, scores averaged 53% in 4th grade and 64% in 7th or 8th grade.

10 It is not clear why DESE changed the performance levels or believed that the new verbiage is an improvement. While the meaning of the labels is very similar, an advantage is that the changing terminology makes it clear that directly comparing scores across years can be problematic.

11 The period between the first Next Generation MCAS and the COVID pandemic was fairly short and showed no substantial changes.

12 2001 is used for most exams because of the large jump in scores between 2000 and 2001 that suggests the exam may have become easier that year. The Sci/Tech exams did not take place in 2001, and the table uses the earliest year after 2001 that the exam was given.

13 The percentage of students who graduate is much higher than the passing rates on the 10th grade exams because students who do not pass in 10th grade have multiple opportunities to retake the exam.

14 This table compares scores for lower-income students (later labeled as economically disadvantaged) compared to students who were not in the low-income group. This gap is larger than that reported in sources that compare the score for low-income students to the overall state average rather than to higher-income students.

15 Candal, 2024.

16 Technically PISA takes students from 15 years and 3 months of age to 16 years and 2 months. Most of the students are in 10th grade in the United States.

17 Stating that a score is significantly above another means that the gap is unlikely to be caused by the random variation in the samples. It means that we can be fairly confident that students in the country with the higher score actually outperform students in the country with the lower score.

18 This ranking is based on the number of countries that scored significantly higher than the US or Massachusetts. For example, if one country had a significantly higher score and two other countries had scores that were higher but not significantly higher, the ranking would be listed as second rather than fourth.

19 Difference between score for students speaking test language at home "always" and score for students speaking language at home "sometimes."

20 Candal, 2024.

About the Authors

Ken Ardon, Ph.D., is a professor of economics at Salem State University, where he has taught since 2004, as well as a member of Pioneer Institute's Center for School Reform Advisory Board. He received a Ph.D. in Economics from the University of California at Santa Barbara in 1999, where he co-authored a book on school spending and student achievement. He taught economics at Pomona College before moving to Massachusetts. From 2000 to 2004, Dr. Ardon worked for the Commonwealth of Massachusetts in the Executive Office of Administration and Finance.

Mission

Pioneer Institute develops and communicates dynamic ideas that advance prosperity and a vibrant civic life in Massachusetts and beyond.

Vision

Success for Pioneer is when the citizens of our state and nation prosper and our society thrives because we enjoy world-class options in education, healthcare, transportation, and economic opportunity, and when our government is limited, accountable, and transparent.

Values

Pioneer believes that America is at its best when our citizenry is well-educated, committed to liberty, personal responsibility, and free enterprise, and both willing and able to test their beliefs based on facts and the free exchange of ideas.

PIONEER INSTITUTE
PUBLIC POLICY RESEARCH

185 Devonshire Street, Suite 1101 Boston MA 02110 617.723.2277
www.pioneerinstitute.org Facebook.com/PioneerInstitute Twitter.com/PioneerBoston