Trends in International Mathematics and Science Study (TIMSS), 2011: Summary of Massachusetts Results

November, 2013
This document was prepared by the
Massachusetts Department of Elementary and Secondary Education
Mitchell D. Chester, Ed.D.
Commissioner

The Massachusetts Department of Elementary and Secondary Education, an affirmative action employer, is committed to ensuring that all of its programs and facilities are accessible to all members of the public. We do not discriminate on the basis of age, color, disability, gender identity, national origin, race, religion, sex or sexual orientation.

Inquiries regarding the Department’s compliance with Title IX and other civil rights laws may be directed to the Human Resources Director, 75 Pleasant St., Malden, MA 02148 781-338-6105.

© 2013 Massachusetts Department of Elementary and Secondary Education
Permission is hereby granted to copy any or all parts of this document for non-commercial educational purposes. Please credit the “Massachusetts Department of Elementary and Secondary Education.”

Massachusetts Department of Elementary and Secondary Education
75 Pleasant Street, Malden, MA 02148-4906
Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370
www.doe.mass.edu/
Table of Contents

Executive Summary .................................................................................................................. v

Introduction ............................................................................................................................. 1

Part I: Massachusetts Grade 8 Mathematics Performance, TIMSS 2011 ........................... 3

Average Student Achievement on the Mathematics Assessment ......................................... 6

Percentage of Students at the “Advanced” Performance Level ............................................. 7

Student Achievement by Cognitive Domain ....................................................................... 8

Student Achievement by Content Domain ....................................................................... 12

Student Achievement by Gender ....................................................................................... 15

Part II: Trends in Massachusetts Student Mathematics Achievement (1999 – 2011) ........... 5

Part III: Massachusetts Grade 8 Science Performance, TIMSS 2011 ............................... 17

Average Student Achievement on the Science Assessment ............................................... 20

Percentage of Students at the “Advanced” Performance Level .......................................... 21

Student Achievement by Cognitive Domain ..................................................................... 22

Student Achievement by Content Domain ..................................................................... 26

Student Achievement by Gender ..................................................................................... 29

Part IV: Trends in Massachusetts Student Science Achievement (1999 – 2011) ............... 19

Part V: Contextual Factors Linked to Mathematics and Science Achievement .............. 31

Home Environment Support .............................................................................................. 32

Home Educational Resources ............................................................................................ 32

Students’ Educational Expectations .................................................................................... 34

School Resources ............................................................................................................. 34

School Location .................................................................................................................. 35

School Composition by Student Economic Background ................................................... 36

Impact on Instruction ......................................................................................................... 37

Access to Science Labs and Lab Support ........................................................................... 39

Teacher Working Conditions ............................................................................................. 40

School Climate .................................................................................................................... 42

School Emphasis on Academic Success ............................................................................. 42
Executive Summary

Introduction

Eighth-grade students in Massachusetts took part in the Trends in International Mathematics and Science Study (TIMSS) in the fall of 2011. TIMSS is an international assessment of mathematics and science achievement that has been administered to fourth- and eighth-grade students every four years since 1995. In 2011, representative samples of students from 63 countries and 14 benchmarking entities, including nine U.S. states, participated in TIMSS. Participating countries and entities could choose to participate in the fourth-grade assessment, the eighth-grade assessment, or both.

Massachusetts, a benchmarking entity, participated in the eighth-grade assessment in 2011. A representative sample of 2,075 students from the Commonwealth took part in the study. Massachusetts also participated in the eighth-grade TIMSS in 1999 and 2007.

In December 2012, TIMSS issued a two-volume report on the results of the 2011 administration. This summary of Massachusetts results draws from data included in that report. The summary is designed for people who want to take a deeper look at the performance of Massachusetts students on the TIMSS assessments.

One of the key reasons Massachusetts participated as a benchmarking entity was to enable the Department of Elementary and Secondary Education to compare students in the Commonwealth with those from the highest-achieving countries in the world. For that reason, most of the tables and graphs in this summary report show either comparisons that involve Massachusetts and Singapore or comparisons that involve Massachusetts, the United States as a whole, and the three top-performing countries (Singapore, the Republic of Korea, and Chinese Taipei).

Overall Achievement

How did Massachusetts students do in 2011?

Scores on the TIMSS mathematics and science tests range from 0 to 1000. Both tests have an average scaled score of 500, with a standard deviation of 100.

In mathematics, Massachusetts eighth graders tied for fifth place in achievement (average score of 561), trailing only the four highest-performing Asian countries: Republic of Korea (613), Singapore (611), Chinese Taipei (609), and Hong Kong SAR (586). Japan, with an average score of 570, was statistically equal to Massachusetts. Massachusetts students scored significantly higher in mathematics when compared to their peers in the United States as a whole (509) and when compared to the international average of 500. (page 3)
In science, Massachusetts students tied for second place in achievement (average score of 567), behind only students from Singapore (590). Massachusetts students scored significantly higher in science when compared to their U.S. peers (525) and when compared to the international average of 500. (page 17)

What percentage of students scored at the “Advanced” benchmark level in the 2011 TIMSS?

Students who score 625 or above on the mathematics or science test are categorized within the “Advanced” benchmark level.

- In mathematics, 19% of Massachusetts eighth graders scored within the “Advanced” category; this compares to 48% of Singaporean students. (page 7)

- In science, 24% of Massachusetts students scored within the “Advanced” benchmark level; this compares to 40% of Singaporean students. (page 21)

What were Massachusetts students’ relative strengths and weaknesses in the 2011 mathematics assessment?

TIMSS reports scores in three cognitive domains: Knowing, Applying, and Reasoning. Reasoning items are the most demanding of students cognitively. In eighth-grade mathematics, TIMSS assesses four content domains: Number, Algebra, Geometry, and Data and Chance. Students’ strengths and weaknesses are identified by comparing their scores in each domain to their average score.

Cognitive Demand: Based on the 2011 results, Massachusetts students have a relative, statistically significant strength in the Knowing cognitive domain, indicating that they are well grounded in basic mathematical knowledge; however, they are relatively weak in the Applying cognitive domain, indicating that they have difficulty solving and representing solutions to problems. Massachusetts students do reasonably well (state average) in the more difficult Reasoning cognitive domain, suggesting that they are capable of thinking logically and systematically and have the potential to analyze and justify their answers in mathematics. (page 8)

Content Knowledge: Students in Massachusetts did significantly better on Number items (e.g., whole numbers, fractions, and decimals) and Data and Chance items (e.g., data interpretation and chance). Students performed significantly less well on items measuring the Geometry content domain (e.g., shapes and measurement). Students performed at the state average in Algebra (e.g., patterns and equations). (page 12)

“These are remarkable results, and I am so impressed how Massachusetts eighth-grade students and teachers continue to raise their game and build on past successes. These results are testament to the dividends that the Commonwealth’s investment in K–12 education is yielding. TIMSS is a renowned international assessment. We will use these results along with other data to identify where gaps in achievement still exist as we strive to ensure that all students are ready for success after high school.”

Commissioner Mitchell Chester, Department of Elementary and Secondary Education (December 2012)
What were Massachusetts students’ relative strengths and weaknesses in the 2011 science assessment?

The cognitive domains for the TIMSS science assessment are identical to those listed above for mathematics. The content domains for science are Biology, Earth Science, Physics, and Chemistry.

**Cognitive Demand:** Based on the 2011 results, Massachusetts students have a relative, statistically significant strength in the Knowing cognitive domain, indicating that they are well grounded in factual scientific knowledge. However, students are relatively and significantly weaker in applying this knowledge; students were relatively less able to compare, contrast, and interpret information, and to model their knowledge. Students performed at the state average in the Reasoning cognitive domain, suggesting that they have the capacity to think analytically, evaluate evidence, develop explanations, and draw conclusions. (page 22)

**Content Knowledge:** Students in Massachusetts performed significantly better on Biology items (e.g., life cycles, cells and their functions) and Earth Science items (e.g., Earth’s structure and processes), but significantly less well on items measuring the Physics content domain (e.g., physical states and changes in matter). Students performed at the state average on Chemistry items (e.g., chemical change). (page 26)

**Trends in Achievement**

Has the achievement of Massachusetts eighth graders improved over time?

Analysis of results from the 1999, 2007, and 2011 TIMSS administrations shows that Massachusetts eighth graders have made significant gains in achievement over time, in both mathematics and science.

- **In mathematics,** Massachusetts students made a 48-point gain in achievement between 1999 and 2011 (from 513 average scaled score points in 1999 to 561 in 2011). The trend in mathematics scores is shown in Figure 1, and related information can be found on page 6. The gain over time was statistically significant and of a moderate effect size.

- **In science,** Massachusetts students made a 34-point gain in achievement (from 533 average scaled score points in 1999 to 567 in 2011). The trend in science scores is shown in Figure 2, and related information can be found on page 20. The gain over time was statistically significant and of a small to moderate effect size.

Although the gain score in mathematics was of a moderate effect, the gain score in mathematics was the highest of any participating country or benchmarking entity for the 1999–2011 period. In science, the gain score was second highest among all countries and benchmarking entities.
Note: Massachusetts did not participate in the 2003 TIMSS.
Do Massachusetts girls do as well as Massachusetts boys on TIMSS?

In mathematics, Massachusetts boys out-scored girls in 1999, 2007, and 2011. However, the difference in average scaled scores in each year was not statistically significant. The comparison of mathematics scores across gender is presented in Figure 3, and related information can be found on page 15.

In science, Massachusetts boys significantly out-performed girls in 1999 (by 13 scaled score points) and in 2007 (by 10 scaled scored points). However, the gender gap in 2011 (6 scaled score points) was not significant, indicating that girls have made important strides in closing this achievement gap. The comparison of science scores across gender is shown in Figure 4, and related information can be found on page 29.

![Figure 3](image_url)

Note: Massachusetts did not participate in the 2003 TIMSS. SG: Singapore
The two-volume TIMSS report provides descriptive findings on how certain contextual factors impact student achievement. These findings are based on surveys completed by students, teachers, and principals. There are too many contextual factors reported by TIMSS to include in this summary. This summary provides information on a selection of factors included in the TIMSS report.

What are some contextual factors that appear to impact student learning?

- **Home Educational Resources:** The TIMSS study found that very few students in Massachusetts (less than 5%) reside in poorly resourced households. Nevertheless, the impact on these students’ learning appears large. On average, students who reported living in under-resourced households in Massachusetts scored over 100 points lower than their peers who live in well-resourced households on both the mathematics and science
assessments. This difference is equivalent to over a full standard deviation, a large effect. (page 32)

- **Teacher Working Conditions**: Students of teachers who reported poor working conditions (lack of materials and supplies, poor building condition, classroom overcrowding) also performed significantly worse than their peers. The average difference in scaled score points was 87 points in mathematics (equivalent to almost one standard deviation, a large effect) and 61 points in science (a moderate to large effect). (page 40)

- **Safe and Orderly Schools**: Between 4% and 7% of all students in Massachusetts have teachers who reported that their school environment was not safe and was disorderly. These students scored almost one standard deviation lower in both subjects (83 scaled score points in mathematics; 85 points in science) than students whose teachers considered their school environments to be safe. This is a large effect. (page 46)

- **Students Bullied at School**: Six percent of Massachusetts students participating in the study reported that they had been bullied once a week. These students scored approximately a quarter of a standard deviation lower (30 scaled score points in mathematics; 26 points in science) than students who reported almost never being bullied. (page 48)

- **Years of Teaching Experience**: Students of teachers with limited teaching experience in science (less than 5 years on the job) scored higher on the science test than students of teachers who could be considered veterans (more than 20 years on the job). The difference in average scaled scores was equivalent to a half of a standard deviation (45 points, a moderate effect). In contrast, students of relatively inexperienced teachers in mathematics scored similarly to students of veteran teachers. (page 50)

- **Student Confidence in Learning Mathematics and Science**: Students’ confidence in subject content also appears to have a large impact on their learning. On average, students who expressed no confidence in mathematics and science scored between 75 and 100 scaled score points lower on the respective tests when compared to their peers. These are both large effects. (page 65)

These are some of the contextual findings from the TIMSS reports (TIMSS, 2011a; TIMSS, 2011b) that are specific to Massachusetts. More rigorous analyses of the TIMSS results are required. These descriptive statistics ignore the possibility of contextual factors influencing each other in their relationship with achievement and do not take into account that students are nested within classrooms, within schools. However, the statistics do provide interesting comparisons and highlight factors that appear to impact student learning.
Conclusion
In 2011, Massachusetts eighth-grade students performed comparably on the TIMSS assessments to students in the top countries in the world in both mathematics and science. The trended data indicates that the performance of Massachusetts students has improved over time, and Massachusetts students continue to out-perform their U.S. peers. Further information on the performance of Massachusetts students on TIMSS can be found in this report and online at http://timss.bc.edu/.
INTRODUCTION

The Trends in International Mathematics and Science Study (TIMSS) assesses and compares student achievement in mathematics and science in grades 4 and 8. The goal of TIMSS is to provide countries with information that can be used to improve teaching and learning in mathematics and science (TIMSS, 2011a; TIMSS, 2011b). TIMSS began in 1995 with 45 countries participating and is conducted every four years. In 2011, more than 600,000 students from 63 countries and 14 benchmarking entities took part in the study. Countries, such as the United States (U.S.), that have participated in all five TIMSS administrations can compare student achievement over time.

Massachusetts was one of nine U.S. states invited to participate in the 2011 TIMSS by the U.S. Department of Education, which funded the cost of administering the assessment. The nine states each participated as a “benchmarking entity,” defined as a regional jurisdiction of a country. TIMSS participation provided Massachusetts with the opportunity to benchmark its students’ performance against that of students from countries around the world. In Massachusetts, 2,075 eighth-grade students from 56 schools took part in the 2011 TIMSS.

The TIMSS results came out in December 2012 and were published in a two-volume report totaling 1,020 pages. In addition to student achievement results, the TIMSS report presents feedback from students, teachers, and principals who completed surveys on issues ranging from bullying to whether schools have science laboratories. These contextual factors can affect student learning and achievement, and countries use the survey data to inform decisions and policies on student learning, classroom practice, and school administration.

This summary report is designed for people who want to take a deeper, more focused look at Massachusetts’ TIMSS results, extracted from the lengthy international report of results. This audience may include students, parents, teachers, school officials, researchers, and government officials. The report includes explanations (in sidebars) of technical terms and statistical concepts. In some sections of the report, readers will benefit from having some preexisting knowledge of statistical techniques.

How This Report Is Organized

This report is composed of seven main parts. Part I provides a summary of Massachusetts student achievement on the TIMSS 2011 mathematics assessment. Part II follows with a summary of the trends in Massachusetts student mathematics performance since 1999. Massachusetts student
achievement on the TIMSS 2011 science assessment is reported in Part III, with trends in science performance (1999–2011) summarized in Part IV. Part V provides a review of several contextual factors (such as teacher working conditions and school climate) that are often associated with student achievement and presents preliminary descriptive data on how these factors impact student learning in Massachusetts. A conclusion and references for the report are provided in Part VI and Part VII, respectively. Appendix I provides readers who are not familiar with TIMSS with important information on how TIMSS collects and reports data (e.g., explanations of TIMSS scaled scores and achievement benchmarks).

Throughout the report, to provide context for readers, Massachusetts student achievement is compared to the achievement of students in the United States and Singapore (the highest-achieving country). When contextual factors are being analyzed, Massachusetts results are compared to the international average and to results for the United States, Singapore, and two other high-achieving Asian countries (the Republic of Korea and Chinese Taipei).

Comparisons of the performance of Massachusetts students to that of students in other countries are available in the full TIMSS reports, posted at http://timss.bc.edu/. Questions concerning this summary report should be directed to the Student Assessment Services office at 781-338-3625.
This section provides information about the performance of Massachusetts eighth graders on the TIMSS 2011 mathematics assessment. Table 1 compares the performance of Massachusetts students with the performance of students from the top-performing countries and from the United States as a whole. In the table, the last column reports the statistical significance of the differences in average scaled scores between Massachusetts and the listed countries (see sidebar for an explanation of statistical significance).

Table 1

<table>
<thead>
<tr>
<th>Country/Benchmarking Entity</th>
<th>Average Scaled Score (Std. Error)</th>
<th>Standard Deviation</th>
<th>Rank*</th>
<th>Significance of Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea, Republic of</td>
<td>613 (2.9)</td>
<td>90</td>
<td>1</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Singapore</td>
<td>611 (3.8)</td>
<td>84</td>
<td>2</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>609 (3.2)</td>
<td>106</td>
<td>3</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Hong Kong SAR</td>
<td>586 (3.8)</td>
<td>84</td>
<td>4</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>Japan</td>
<td>570 (2.6)</td>
<td>85</td>
<td>5</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>561 (5.3)</td>
<td>73</td>
<td>5</td>
<td>------</td>
</tr>
<tr>
<td>United States</td>
<td>509 (2.6)</td>
<td>77</td>
<td>9</td>
<td>p &lt; 0.05</td>
</tr>
<tr>
<td>TIMSS Centerpoint</td>
<td>500</td>
<td>100</td>
<td>-----</td>
<td>p &lt; 0.05</td>
</tr>
</tbody>
</table>

*Rankings are based on 63 countries and 14 benchmarking entities who participated in TIMSS 2011.

**Statistical Significance**

TIMSS reports whether the difference in average student achievement between two countries (the test statistic) is statistically significant using a p-value of 0.05. The test statistic is compared to the hypothesis that there is zero difference between the two means (called the null hypothesis). The observed significance level (p-value) is the chance of getting a test statistic as large as or larger than the observed one. The chance is derived on the basis of the null hypothesis being right (i.e., there is no difference in the means between two countries). The smaller the chance is, the stronger the evidence against the null hypothesis. The p-value used by TIMSS (0.05) indicates that there is less than a 5% chance of getting a large test statistic when the null hypothesis is right.

**Note:** TIMSS does not correct for multiple comparisons. Therefore, it is important to use other evidence, such as effect sizes, to determine if differences in average scaled scores are meaningful.
When data are being compared (e.g., to determine if the difference between two means is statistically significant) and comparisons involve large sample sizes, even small differences will be reported as significant. In this report, effect sizes are used to help determine if differences are meaningful. Any effect sizes reported are Cohen’s D (see sidebar for an explanation) and use the international reference of one standard deviation being equivalent to 100 scaled score points. In using the international standard deviation in computing effect sizes, it is important to note that the effect sizes reported are likely underestimated.

Summary of Student Performance on the Mathematics Assessment:

- Massachusetts students (average scaled score of 561) performed comparably to students in Japan (570), the fifth-ranked country on the mathematics assessment.
- On average, Massachusetts students scored within the High (greater than or equal to 550) performance category.
- Massachusetts students scored significantly lower, on average, than students in the four top-performing countries: the Republic of Korea (613), Singapore (611), Chinese Taipei (609), and Hong Kong SAR (586). When Massachusetts is compared to the Republic of Korea and Singapore, the differences in average scaled scores (-52 points and -50 points, respectively) represent approximately half of a standard deviation (a moderate effect size).
- In comparison to students from the United States as a whole (509), students in Massachusetts scored significantly higher on the mathematics assessment; this difference (+52 points) is equivalent to half of a standard deviation, a moderate effect size.
- Massachusetts students scored 61 points higher than the TIMSS centerpoint (500), a difference that amounts to almost two thirds of a standard deviation (a moderate-to-large effect size).

---

**Effect Size**

As described on the previous page, TIMSS reports whether the difference in average student achievement between two countries is statistically significant. Reporting effect sizes is a common practice, used to assess how meaningful these statistically significant differences are.

An effect size is reported in standard units. It expresses the number of standard deviations that correspond to a value (in this case, the difference between two means). The international standard deviation of 100 (for achievement data) or 2 (for contextual data) is used in all calculations. The generic effect size formula for achievement data is as follows:

\[
\frac{\text{Mean 1} - \text{Mean 2}}{100}
\]

The following guidelines for Cohen’s D effect sizes are commonly used to assess the meaningfulness of an effect:

- **Large Effect:** 0.8
- **Moderate Effect:** 0.5
- **Small Effect:** 0.2
PART II
TRENDS IN MASSACHUSETTS STUDENT MATHEMATICS ACHIEVEMENT (1999–2011)

This section examines trends in Massachusetts student achievement in mathematics from 1999 to 2011. Massachusetts eighth graders participated in the TIMSS assessments in 1999 and 2007; Massachusetts did not participate in the 2003 TIMSS. Table 2 provides the sample size (total number of students assessed) for each of the TIMSS test administrations that Massachusetts participated in. The table also provides the participation rate (in parentheses), which corresponds to the percentage of students participating out of the total number invited to participate.

<table>
<thead>
<tr>
<th></th>
<th>1999</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>9,072 (85%)</td>
<td>7,377 (77%)</td>
<td>10,477 (81%)</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>Not Available</td>
<td>1,897 (92%)</td>
<td>2,075 (96%)</td>
</tr>
<tr>
<td>Singapore</td>
<td>4,966 (98%)</td>
<td>4,599 (95%)</td>
<td>5,927 (95%)</td>
</tr>
</tbody>
</table>

In the figures that follow, the performance of Massachusetts eighth graders in 2011 is compared to that of their peers in 2007 and 1999. For context, data from Singapore and the United States are also provided for the same time period. Students from Singapore have consistently performed well on TIMSS, and the use of Singapore as a comparison helps us gauge whether Massachusetts students are making progress in closing the achievement gap between us and the top-achieving nations. Included in this section is an assessment of the trends in (1) average student achievement, (2) percentage of students scoring at the “Advanced” performance level, (3) student achievement by cognitive domain, (4) student achievement by content domain, and (5) student achievement by gender.
1. Average Student Achievement on the Mathematics Assessment

Figure 1 shows changes over time in the average achievement of Massachusetts eighth graders in mathematics (blue trend line).

![Figure 1](image)

Note: Massachusetts did not participate in the 2003 TIMSS.

**Summary of Achievement Data:**

- The average performance of Massachusetts students increased from 513 in 1999 to 561 in 2011; this difference of 48 scaled score points (almost half a standard deviation) was statistically significant and of a moderate effect size.
- The average performance of U.S. and Singaporean students did not significantly increase between 1999 and 2011. The 18-point gain in Singapore’s average achievement between 2007 (593) and 2011 (611) was, however, statistically significant.
- The mathematics achievement gap between Massachusetts students and Singaporean students has decreased from 91 scaled score points in 1999 to 50 scaled score points in 2011. In 1999, this achievement gap represented approximately one full standard deviation (a large effect); the gap has been reduced to half of a standard deviation in 2011 (a moderate effect).
2. Percentage of Students at the “Advanced” Performance Level

Student performance on the TIMSS assessments is categorized into four performance levels: Low, Intermediate, High, and Advanced. To provide qualitative information on the competencies and skills needed to answer items at each performance level, TIMSS classifies and anchors (benchmarks) items that are representative of each performance level. For example, a multiple-choice item is associated with the highest level of competency (Advanced benchmark) if the item was answered correctly by 65% of “Advanced” students and less than 50% of “High”-performing students.

Students must score 625 or above to be categorized within the “Advanced” performance level in mathematics. Students in this category “can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning” (TIMSS, 2011a, p. 87). This ability differentiates “Advanced” students from those in the “High” performance category (550 or greater), who are less proficient in solving complex and non-routine problems. Figure 2 portrays the trend in the percentage of Massachusetts eighth graders who performed at the “Advanced” level.

**Figure 2**

**Mathematics (1999 - 2011)**

**Percentage of G8 Students Achieving Advanced International Benchmark**

- **1999**
- **2007**
- **2011**

<table>
<thead>
<tr>
<th>Year</th>
<th>U.S.</th>
<th>MA</th>
<th>Singapore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2007</td>
<td>8</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>2011</td>
<td>42</td>
<td>40</td>
<td>48</td>
</tr>
</tbody>
</table>

**Performance-Level Algebra Items**

**Low Benchmark**

\[ y = \frac{(a + b)}{c} \]

a = 8, b = 6, and c = 2

What is the value of y?

**Advanced Benchmark**

Solve this inequality:

\[ 9x - 6 < 4x + 4 \]

**Source:** TIMSS (2011a, p. 123, p. 137)
Summary of Benchmark Performance Level Data:

- The percentage of Massachusetts students scoring at the “Advanced” level more than doubled between 1999 and 2011, from 8% to 19%, a statistically significant increase. In 2011, almost one-fifth of Massachusetts students scored “Advanced.”
- In 1999, the percentage of Singaporean students scoring within the top performance category was more than five times the percentage of Massachusetts students achieving at the same level; in 2011, this multiple was reduced by half to 2.5 times.
- The percentage of U.S. students performing at the “Advanced” level stayed the same between 1999 and 2011, with less than 10% of U.S. students scoring within that top category.
- The percentage of students from Singapore who scored within the top performance category increased from 42% in 1999 to 48% in 2011. This increase was not statistically significant. The difference between 2007 and 2011 (8%) was statistically significant.

3. Student Achievement by Cognitive Domain

TIMSS reports scores in three cognitive domains: Knowing, Applying, and Reasoning. Each domain presents different cognitive challenges to students, with items in the Reasoning domain being, on average, the most difficult on the test.

- **Reasoning** items in mathematics require students to think logically and systematically and to be able to analyze and justify their answers to non-routine problems.
- Items from the Applying domain focus on students’ ability to solve relatively routine problems; these problems may require students to model their mathematical solutions.
- The Knowing domain requires students to recall, recognize, and compute basic mathematical facts; on average, items in this domain are the easiest items on the assessment.

Table 3 presents the percentage of items within each of the cognitive domains for the 2007 and 2011 mathematics assessments. Data were not available for 1999.
Table 3
Eighth-Grade Cognitive Domains in Mathematics TIMSS

<table>
<thead>
<tr>
<th>Cognitive Domain</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing</td>
<td>35%</td>
<td>35%</td>
</tr>
<tr>
<td>Applying</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>Reasoning</td>
<td>25%</td>
<td>25%</td>
</tr>
</tbody>
</table>

Figure 3 presents student achievement in mathematics broken down by cognitive domain for 2007 and for 2011. This section will first discuss the relative strengths and weaknesses of Massachusetts students in 2011 when compared to their peers and to their average performance. The summary next discusses trends in student achievement across time (2007–2011) in the cognitive domains.
Figure 3
Comparison of Mathematics Student Performance by Cognitive Domain

Mathematics (2007)
G8 Student Performance by Cognitive Domain

<table>
<thead>
<tr>
<th>Cognitive Domain</th>
<th>Knowing</th>
<th>Applying</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. (508)</td>
<td>517</td>
<td>502</td>
<td>506</td>
</tr>
<tr>
<td>MA (547)</td>
<td>554</td>
<td>543</td>
<td>546</td>
</tr>
<tr>
<td>Singapore (593)</td>
<td>592</td>
<td>597</td>
<td>589</td>
</tr>
</tbody>
</table>

Mathematics (2011)
G8 Student Performance by Cognitive Domain

<table>
<thead>
<tr>
<th>Cognitive Domain</th>
<th>Knowing</th>
<th>Applying</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. (509)</td>
<td>519</td>
<td>503</td>
<td>503</td>
</tr>
<tr>
<td>MA (561)</td>
<td>569</td>
<td>555</td>
<td>562</td>
</tr>
<tr>
<td>Singapore (611)</td>
<td>617</td>
<td>613</td>
<td>604</td>
</tr>
</tbody>
</table>
Summary of 2011 Mathematics Performance Data by Cognitive Domain:

- In each of the three cognitive domains, Massachusetts students in 2011 scored higher, on average, than U.S. students but lower, on average, than Singaporean students.

- In 2011, when compared to their overall average scaled score (561), Massachusetts students had a relative and statistically significant strength in the Knowing cognitive domain (569; +8 points) but a relative weakness in the Applying domain (555; -6 points). The average scaled score for Massachusetts students in the Reasoning domain (562) was similar to the overall state average.

- In 2011, U.S. students as a whole performed relatively weaker in Reasoning (503) and Applying (503) when compared to their overall average scaled score of 509. U.S. students performed above the national average for Knowing items (519), a relative strength.

- In 2011, Singapore students had a relative strength in the Knowing (617) and Applying (613) cognitive domains but a relative weakness in mathematical Reasoning skills (604) when compared to their national average score of 611.

Summary of Mathematics Performance Trends by Cognitive Domain:

- From 2007 to 2011, the average scaled scores of Massachusetts students increased across all three cognitive domains. Average scores increased by just over a tenth of a standard deviation in each domain (15 points in Knowing, 12 points in Applying, and 16 points in Reasoning). However, the gains in average scores were not statistically significant across administrations.

- In all three cognitive domains, the positive gap between Massachusetts performance and overall U.S. performance widened. For example, the gap in average Reasoning scores between Massachusetts students and U.S. students increased from 40 scaled score points in 2007 to 59 scaled score points in 2011.

- In two of the three cognitive domains (Knowing and Applying), the negative gap between student scores in Massachusetts and Singapore widened. For example, in the Knowing domain, the gap between Massachusetts students and Singaporean students increased from 38 to 48 scaled score points between 2007 and 2011. Singaporean student scores increased significantly across all three
cognitive domains between 2007 and 2011, and this contributed to the widening gap between Massachusetts and Singapore.

4. Student Achievement by Content Domain

TIMSS assesses four content domains in eighth-grade mathematics: *Number*, *Algebra*, *Geometry*, and *Data and Chance*. These four content areas are further divided into more specific content knowledge, as shown below:

- **Number**—whole numbers; fractions and decimals; integers and ratios; proportion and percent
- **Algebra**—patterns; algebraic expressions and equations/formulas; functions
- **Geometry**—geometric shapes; geometric measurement; location and movement
- **Data and Chance**—data organization and representation; data interpretation; chance

Table 4 presents the percentage of items within each content domain for 2007 and 2011. Data were not available for 1999.

<table>
<thead>
<tr>
<th>Content Domain</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>Algebra</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>Geometry</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Data and Chance</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Figure 4 presents student achievement in mathematics broken down by content domain for 2007 and for 2011. The summary will first discuss the relative strengths and weaknesses of Massachusetts students in 2011 when compared to their peers and to their average performance. The summary next discusses student achievement across time (2007–2011) in the content domains.
Figure 4
Comparison of Mathematics Student Performance by Content Domain

Mathematics (2007)
G8 Student Performance by Content Domain

Mathematics (2011)
G8 Student Performance by Content Domain
Summary of 2011 Mathematics Performance Data by Content Domain:

- In each of the four content domains, Massachusetts students in 2011 scored higher, on average, than U.S. students but lower, on average, than Singaporean students.

- In 2011, when compared to their overall average scaled score (561), Massachusetts students had a relative and statistically significant strength in the Number (567; +6 points) and Data and Chance (584; +23 points) content domains but a relative weakness in the Geometry domain (548; -13 points). The average scaled score for Massachusetts students in the Algebra domain (559) was similar to the overall state average.

- When compared to their overall average scaled score (509), U.S. students had a relative and statistically significant strength in the Algebra (512; +3), Number (514; +5), and Data and Chance (527; +18) content domains in 2011. U.S. students, as a whole, had a substantial and statistically significant relative weakness in Geometry (485; -24 points).

- In 2011, Singaporean students scored relatively consistently and at a high level across all four content domains. When compared to their overall average scaled score of 611, Singaporean students had a statistically significant strength in the Algebra content domain (614) but a statistically significant weakness in Data and Chance (607).

Summary of Mathematics Performance Trends by Content Domain:

- From 2007 to 2011, the average scaled scores of Massachusetts students increased across all four content domains. In Geometry, Massachusetts student performance increased by 25 scaled score points over that period; this difference was statistically significant. Student scores also increased by 13 points in the Number domain, 12 points in Algebra, and 5 points in Data and Chance, but these gains in average scaled scores were not statistically significant.

- In all four content domains, the positive gap between the performances of Massachusetts students and U.S. students widened. Most notably, the difference in average scaled scores in Geometry grew from 43 points in 2007 to 63 points in 2011.

- The negative gap between the performances of Massachusetts students and Singaporean students grew smaller in two content domains: the Number domain (reduced from 51 points in 2007 to 44 points in 2011) and the Geometry domain (reduced from 67 points in 2007 to 61 points in 2011). In contrast, the negative gap between Massachusetts and Singapore widened in Algebra (from 44 points in 2007 to 57 points in 2011).
2007 to 55 points in 2011) and in Data and Chance (from 10 points in 2007 to 23 points in 2011).

5. Student Achievement by Gender
This section presents trends in Massachusetts student achievement in mathematics, broken down by gender. In all administrations and in all countries, the percentage of girls in the samples was within one percentage point of 50%. To provide context, the trends are also provided for U.S. students and Singaporean students. The data are shown in Figure 5, with the darker trend lines representing boys within each country.

Figure 5

Note: Massachusetts did not participate in the 2003 TIMSS. SG: Singapore

Summary of Mathematics Performance Trends by Gender:
- Massachusetts boys scored, on average, higher than girls across all administrations of the eighth-grade mathematics assessment: 1999 (7-point gap), 2007 (6-point gap), and 2011 (5-point gap). However, the difference in average scores was not significant in any of those administrations, indicating that the average scaled scores over time are comparable across gender.
• In Massachusetts, girls and boys both appear to have contributed to the improved performance on the assessment between 1999 and 2011. The average scaled score for girls increased by 48 points between 1999 and 2011; the average scaled score for boys increased by 46 points over the same period. These increases are both equivalent to almost half a standard deviation (moderate effect size).

• U.S. boys scored significantly higher than girls on the 1999 (7-point difference) and 2003 (5-point difference) administrations. However, the gender gaps in 2007 and 2011 (3 points each year) were not significant, indicating that U.S. girls and U.S. boys performed comparably on the 2007 and 2011 TIMSS assessments. Between 1999 and 2011, there was little improvement in average scaled scores for either U.S. girls or U.S. boys, with differences equivalent to a tenth of a standard deviation (a small effect).

• Singaporean girls have scored, on average, significantly higher than boys on the TIMSS mathematics assessment since 2003, with the scaled score difference in the 2011 assessment being equivalent to 8 points. When compared to Massachusetts students, girls and boys in Singapore exhibit a more unpredictable achievement pattern in mathematics, with average scaled scores fluctuating in the last three TIMSS administrations.
PART III
MASSACHUSETTS GRADE 8 SCIENCE PERFORMANCE, TIMSS 2011

This section provides information about the performance of Massachusetts eighth graders on the TIMSS 2011 science assessment. Table 5 compares the performance of Massachusetts students with the performance of students from the top-performing countries and from the United States as a whole. In the table, the last column reports the statistical significance of the differences in average scaled scores between Massachusetts and the listed countries (see sidebar on page 3 for an explanation of statistical significance).

Table 5
Grade 8 Performance on TIMSS 2011 Science Assessment

<table>
<thead>
<tr>
<th>Country/Benchmarking Entity</th>
<th>Average Scaled Score (Standard Error)</th>
<th>Standard Deviation</th>
<th>Rank*</th>
<th>Significance of Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>590 (4.3)</td>
<td>97</td>
<td>1</td>
<td>( p &lt; 0.05 )</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>567 (5.1)</td>
<td>81</td>
<td>2</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>564 (2.3)</td>
<td>84</td>
<td>2</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Korea, Republic of</td>
<td>560 (2.0)</td>
<td>77</td>
<td>3</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Japan</td>
<td>558 (2.4)</td>
<td>76</td>
<td>4</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Finland</td>
<td>552 (2.5)</td>
<td>65</td>
<td>5</td>
<td>Not Significant</td>
</tr>
<tr>
<td>United States</td>
<td>525 (2.6)</td>
<td>81</td>
<td>10</td>
<td>( p &lt; 0.05 )</td>
</tr>
<tr>
<td>TIMSS Centerpoint</td>
<td>500</td>
<td>100</td>
<td>-----</td>
<td>( p &lt; 0.05 )</td>
</tr>
</tbody>
</table>

*Rankings are based on 63 countries and 14 benchmarking entities who participated in TIMSS 2011.

Summary of Student Performance on the Science Assessment:

- Massachusetts students (average scaled score of 567) performed comparably to students from Chinese Taipei (564), the second-ranked country on the science assessment.
- On average, Massachusetts students performed within the High (greater than or equal to 550) performance category.
- Massachusetts students scored significantly lower, on average, than students in the top-performing country of Singapore (590); the difference in average scaled scores (-23 points) is equivalent to a quarter of a standard deviation, a small effect size.
- In comparison to students from the United States as a whole (525), students in Massachusetts scored significantly higher on the science assessment; the difference in average scaled scores (+42 points) represents approximately half of a standard deviation (a moderate effect size).
Massachusetts students scored 67 points higher than the TIMSS centerpoint (500), a difference that amounts to two-thirds of a standard deviation (a moderate-to-large effect size).
PART IV
TRENDS IN MASSACHUSETTS STUDENT SCIENCE ACHIEVEMENT (1999–2011)

This section examines trends in Massachusetts student achievement in science from 1999 to 2011. Massachusetts eighth graders participated in the TIMSS assessments in 1999 and in 2007; Massachusetts did not participate in the 2003 TIMSS. Table 6 provides the sample size (total number of students assessed) for each of the TIMSS test administrations that Massachusetts participated in. The table also provides the participation rate (in parentheses), which corresponds to the percentage of students participating out of the total number invited to participate.

<table>
<thead>
<tr>
<th></th>
<th>1999</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United States</strong></td>
<td>9,072</td>
<td>7,377</td>
<td>10,477</td>
</tr>
<tr>
<td><strong>Massachusetts</strong></td>
<td>Not Available</td>
<td>1,897 (92%)</td>
<td>2,075 (96%)</td>
</tr>
<tr>
<td><strong>Singapore</strong></td>
<td>4,966 (98%)</td>
<td>4,599 (95%)</td>
<td>5,927 (95%)</td>
</tr>
</tbody>
</table>

In the figures that follow, the performance of Massachusetts eighth graders in 2011 is compared to that of their peers in 2007 and 1999. For context, data from Singapore (ranked first in science in 2011) and the United States are also provided for the same period. Included in this section is an assessment of the trends in (1) average student achievement, (2) percentage of students scoring at the “Advanced” performance level, (3) student achievement by cognitive domain, (4) student achievement by content domain, and (5) student achievement by gender.
1. Average Student Achievement on the Science Assessment

Figure 6 shows changes over time in the average achievement of Massachusetts eighth graders in science.

**Figure 6**

<table>
<thead>
<tr>
<th>Year of Administration</th>
<th>U.S.</th>
<th>MA</th>
<th>Singapore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>515</td>
<td>533</td>
<td>568</td>
</tr>
<tr>
<td>2003</td>
<td>527</td>
<td>556</td>
<td>578</td>
</tr>
<tr>
<td>2007</td>
<td>520</td>
<td>567</td>
<td>567</td>
</tr>
<tr>
<td>2011</td>
<td>525</td>
<td>567</td>
<td>590</td>
</tr>
</tbody>
</table>

Note: Massachusetts did not participate in the 2003 TIMSS.

**Summary of Science Achievement Data:**

- The average performance of Massachusetts students increased from 533 in 1999 to 567 in 2011; this difference of 34 scaled score points (a third of a standard deviation) was statistically significant.
- The average performance of U.S. students did not significantly increase between 1999 (515) and 2011 (525). In contrast, Singapore’s 22-point gain in average achievement between 1999 (568) and 2011 (590) was statistically significant.
- The science achievement gap between Massachusetts and Singapore decreased from 35 scaled score points in 1999 to 23 scaled score points in 2011. In 1999, this achievement gap represented approximately a third of a standard deviation; the gap was reduced to approximately a quarter of a standard deviation in 2011.
2. Percentage of Students at the “Advanced” Performance Level

Student performance on the TIMSS assessments is categorized into four performance levels: Low, Intermediate, High, and Advanced. To provide qualitative information on the competencies and skills needed to answer items at each performance level, TIMSS classifies and anchors (benchmarks) items that are representative of each performance level. For example, a multiple-choice item is associated with the highest level of competency (Advanced benchmark) if the item was answered correctly by 65% of “Advanced” students and less than 50% of “High”-performing students.

Students have to score 625 or above to be categorized within the “Advanced” performance level in science. Students in this category can “communicate an understanding of complex and abstract concepts in biology, chemistry, physics, and earth science” (TIMSS, 2011b, p. 111). The ability to combine information from several sources, draw conclusions, and provide coherent written explanations to convey their scientific knowledge differentiates these students from students within the “High” performance category (greater than or equal to 550), who are less proficient in these respects.

Figure 7 portrays the trend in the percentage of Massachusetts eighth graders who performed at the “Advanced” level in science.

Figure 7

<table>
<thead>
<tr>
<th>Participant</th>
<th>1999</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>12</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>MA</td>
<td>15</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Singapore</td>
<td>29</td>
<td>32</td>
<td>40</td>
</tr>
</tbody>
</table>

Performance-Level Chemistry Items

Low Benchmark

What is the chemical formula for carbon dioxide?

☐ CO
☐ CO₂
☐ C
☐ O₂

TIMSS (2011a, p. 121)

Advanced Benchmark

(Short Answer)

Ahmet put some powder into a test tube. He then added liquid to the powder and shook the test tube. A chemical reaction took place.

Describe two things he might observe as the chemical reaction took place.

TIMSS (2011a, p. 136)
Summary of Benchmark Performance Level Data:

- The percentage of Massachusetts students scoring at the “Advanced” level increased by a factor of 1.6 from 1999 to 2011 (from 15% in 1999 to 24% in 2011). This difference (9%) is statistically significant.
- In 1999, the percentage of Singaporean students scoring within the top performance category was almost twice the percentage of Massachusetts students achieving at the same level; in 2011, this multiple was reduced to 1.7 times.
- The percentage of U.S. students performing at the “Advanced” level did not improve between 1999 and 2011, with approximately 10% of U.S. students scoring within the top performance level in each administration.
- The percentage of students from Singapore who scored within the top performance category increased from 29% in 1999 to 40% in 2011. This difference (11%) was statistically significant.

3. Student Achievement by Cognitive Domain
TIMSS reports scores in three cognitive domains: Knowing, Applying, and Reasoning. Each domain presents different cognitive challenges to students, with items in the Reasoning domain being, on average, the most difficult on the test.

- Reasoning items in science require students to have the ability to think analytically, evaluate evidence, develop explanations, and draw conclusions.
- Items from the Applying domain focus on students’ ability to compare, contrast, and interpret information, and to model their scientific knowledge.
- The Knowing domain, in contrast, relies on students recalling, recognizing, and describing basic scientific facts. These items are, on average, the easiest items on the assessment.

Table 7 presents the percentage of items within each of the cognitive domains for the 2007 and 2011 TIMSS science assessments. (Data were not available for 1999.) Notably, the percentage of Reasoning items on the 2011 assessment was lower than the percentage on the 2007 assessment.
Table 7
Cognitive Domains in Science TIMSS

<table>
<thead>
<tr>
<th>Cognitive Domain</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowing</td>
<td>30%</td>
<td>35%</td>
</tr>
<tr>
<td>Applying</td>
<td>35%</td>
<td>35%</td>
</tr>
<tr>
<td>Reasoning</td>
<td>35%</td>
<td>30%</td>
</tr>
</tbody>
</table>

Figure 8 presents student achievement in science broken down by cognitive domain for 2007 and for 2011. This section will first discuss the relative strengths and weaknesses of Massachusetts students in 2011 when compared to their peers and to their average performance. The summary next discusses trends in student achievement across time (2007–2011) in the cognitive domains.

Summary of 2011 Science Performance Data by Cognitive Domain:

- In each of the three cognitive domains, Massachusetts students in 2011 scored higher, on average, than U.S. students but lower, on average, than Singaporean students.
- In 2011, when compared to their overall average scaled score (567), Massachusetts students had a relative and statistically significant strength in the Knowing cognitive domain (576; +9 points) but a relative and statistically significant weakness in the Applying domain (561; -6 points). Massachusetts students scored at the state average (567) on science Reasoning items.
- The pattern of results for U.S. students was similar to the pattern for Massachusetts. Compared to their overall score (525), U.S. students had a relative and statistically significant strength in Knowing items (527) but a relative and statistically significant weakness in Applying items (522). U.S. students performed at approximately the national average for science Reasoning items (524).
- In 2011, Singapore students scored consistently and at high levels across all three cognitive domains, with average scaled scores ranging from 588 (Knowing) to 592 (Reasoning).

Summary of Science Performance Trends by Cognitive Domain:

- From 2007 to 2011, the average scaled scores of Massachusetts students increased in two of three cognitive domains. In the Knowing domain, the average student score increased 25 points (from 551 to 576), a statistically significant difference. The average score for the Applying domain increased 8 points (from 553 to 561),
but this difference was not significant. The average Massachusetts score in the *Reasoning* domain remained constant over the four years.

- In all three cognitive domains, the positive performance gap between Massachusetts students and U.S. students widened. For example, the gap in *Knowing* scores between Massachusetts students and U.S. students increased from 35 scaled score points in 2007 to 49 scaled score points in 2011. Smaller increases in the performance gap were evident for the *Applying* domain (+3 points) and the *Reasoning* domain (+5 points).

- In all three cognitive domains, the negative performance gap between Massachusetts students and Singaporean students widened. Most notable is the widening achievement gap in science *Reasoning* skills, which increased from 1 scaled score point in 2007 to 25 scaled score points in 2011. Singaporean students’ scores increased significantly across all three cognitive domains between 2007 and 2011, and this contributed to the widening gap between Massachusetts and Singapore.
Figure 8
Comparison of Science Student Performance by Cognitive Domain

Science (2007)
G8 Student Performance by Cognitive Domain

Knowing 516 517 529
Applying 551 553 567
Reasoning 570 567 568

U.S. (520)  MA (556)  Singapore (567)

Science (2011)
G8 Student Performance by Cognitive Domain

Knowing 527 522 524
Applying 576 561 567
Reasoning 588 589 592

U.S. (525)  MA (567)  Singapore (590)
4. Student Achievement by Content Domain
TIMSS assesses four content domains in eighth-grade science: Biology, Chemistry, Physics, and Earth Science. These four content areas are further divided into more specific content knowledge, as shown below:

- **Biology**—cells and their functions; life cycles; reproduction and heredity; ecosystems
- **Chemistry**—classification and composition of matter; properties of matter; chemical change
- **Physics**—physical states and changes in matter; energy transformations, heat, and temperature; light and sound; electricity and magnetism; forces and motion
- **Earth Science**—Earth’s structure and physical features; Earth’s processes, cycles, and history; Earth’s resources, their use, and conservation; Earth in the solar system and universe

Table 8 presents the percentage of items within each content domain for 2007 and 2011. Data were not available for 1999.

<table>
<thead>
<tr>
<th>Content Domain</th>
<th>2007</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>35%</td>
<td>35%</td>
</tr>
<tr>
<td>Chemistry</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Physics</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>Earth Science</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Figure 9 presents student achievement data in science broken down by content domain for 2007 and for 2011. The summary will first discuss the relative strengths and weaknesses of Massachusetts students in 2011 when compared to their peers and to their average performance. The summary next discusses student achievement across time (2007–2011) in the content domains.
Figure 9
Comparison of Science Student Performance by Content Domain
Summary of 2011 Science Performance Data by Content Domain:

- In each of the four content domains, Massachusetts students in 2011 scored higher, on average, than U.S. students. Massachusetts students scored lower, on average, than Singaporean students in three of the four content domains, but had a higher average scaled score in Earth Science.

- In 2011, when compared to their overall average scaled score (567), Massachusetts students had a relative and statistically significant strength in the Biology (575; +8 points) and Earth Science (577; +10 points) content domains but a relative weakness in the Physics domain (555; -12 points). Massachusetts students scored at approximately the state average in the Chemistry domain (568).

- In 2011, U.S. students performed relatively well in the Biology (530; +5 points) and Earth Science (533; +8 points) content domains when compared to their overall average scaled score of 525. U.S. students had a relative weakness in the Physics (513; -12 points) and Chemistry (520; -5 points) content domains. These differences were all statistically significant.

- In 2011, when compared to their overall average scaled score (590), Singaporean students had a relative strength in Physics (602; +12 points) and Biology (594; +4 points) but a substantial relative weakness in Earth Science (566; -24 points). These differences were all statistically significant. Singaporean students performed at the national average of 590 in Chemistry.

Summary of Science Performance Trends by Content Domain:

- Between 2007 and 2011, Massachusetts student performance increased significantly in both the Chemistry and Physics content domains. The average scaled score in Chemistry increased from 546 in 2007 to 568 in 2011 (+22 points); in Physics, the average scaled score increased from 539 in 2007 to 555 in 2011 (+16 points).

- Massachusetts student performance increased by 10 points in both the Biology and Earth Science domains between 2007 and 2011; these gains were not statistically significant.

- In all four content domains, the positive gap between Massachusetts performance and U.S. performance widened. Most
notably, the achievement gap in Chemistry grew from 36 points in 2007 to 48 points in 2011. A similar increase was evident in Biology, where the achievement gap between Massachusetts and U.S. students increased from 34 points in 2007 to 45 points in 2011.

- The negative achievement gap between Massachusetts and Singaporean students widened in three of the four content domains between 2007 and 2011. For example, in Biology, Singaporean students outperformed Massachusetts students by only 2 points in 2007; this gap widened to 19 points in 2011. In Earth Science, Massachusetts students have a relative strength when compared to Singaporean students. However, Singaporean students reduced this achievement gap to 11 points in 2011, down from 20 points in 2007.

5. Student Achievement by Gender

This section presents trends in Massachusetts student achievement in science, broken down by gender. In all administrations and in all countries, the percentage of girls in the samples was within one percentage point of 50%. To provide context, the trends are also provided for U.S. students and Singaporean students. The data are shown in Figure 10, with the darker trend lines representing boys within each country.

Summary of Science Performance Trends by Gender:

- Massachusetts boys have scored, on average, higher than girls across all administrations of the eighth-grade science assessment. The gap in achievement was significant in 1999 (13 points) and in 2007 (10 points). However, girls in Massachusetts have closed the gap, and the difference in average scaled scores in 2011 (6 points) was not significant.

- In Massachusetts, girls and boys both appear to have contributed to the improved performance on the science assessment between 1999 and 2011. The average scaled score for girls increased by 37 points between 1999 and 2011; the average score for boys increased by 30 points over the same time period. These increases are both equivalent to a third of a standard deviation (a small-to-moderate effect size).

- When compared to U.S. girls, U.S. boys have scored consistently and significantly higher on the TIMSS science assessment over the years. The 19-point gender gap that existed in 1999 closed to 11 points in 2011, but this gap is still significant. Between 1999 and 2011, the improvements in average scaled scores for U.S. girls (14 points) and U.S. boys (6 points) were of a small effect size (approximately a tenth of a standard deviation).

- Since 2003, Singaporean girls have performed comparably to Singaporean boys on the TIMSS science assessment. The increase in the average scaled score for Singaporean boys between 1999 and 2011 was of a small effect size (13 points); in contrast, the increase in the average scaled score for girls between 1999 and 2011 was equivalent to a small-to-medium effect size (32 points).
Figure 10

Science by Gender
G8 Trends in Achievement (1999 - 2011)

Note: Massachusetts did not participate in the 2003 TIMSS. SG: Singapore
Along with data on student achievement, the two-volume TIMSS 2011 report also presents data from surveys administered to principals, teachers, and students on contextual factors that support student learning and achievement. The TIMSS surveys inquired about the following five key areas of support: (1) home environment support, (2) school resources, (3) school climate, (4) teacher preparation, and (5) classroom instruction. Survey respondents were asked about multiple contextual factors within each of these five areas.

The figures and tables on the following pages present the survey findings for Massachusetts in each of the five key areas of support. In the figures and tables, Massachusetts is compared to the United States as a whole and to the three top-achieving East Asian countries: Singapore, Chinese Taipei, and the Republic of Korea. Note that not all contextual factors within each area of support are reported. An analysis of how the contextual factors combine to impact Massachusetts student achievement is beyond the scope of this summary report; as a result, this section of the report is purely descriptive.

For readers who want to review data for other countries, each figure and table is footnoted with the associated exhibit number from the two-volume TIMSS report, which is available online at http://timss.bc.edu/. In the TIMSS report, survey results are linked to average student achievement for the country or benchmarking participant.

How TIMSS Contextual Data Are Reported

TIMSS reports contextual data in different ways. In some cases, survey responses are tabulated and reported as simple percentages. For example, 9 (page 34) shows the percentage of students who expect to go onto attain a university or post-graduate degree in the future. The data reported in the table come directly from a student report.

In other cases, the responses from survey participants are calibrated and put on a scale. Take, for example, the TIMSS Safe and Orderly School scale, shown in Figure 20 (page 47). Mathematics teachers were asked to indicate their level of agreement with five statements evaluating their perceptions of how safe and orderly their school environment is. Four response options were provided (agree a lot; agree a little; disagree a little; or disagree a lot), and these responses were scored 3, 2, 1, and 0, respectively. Using a mathematical transformation, the raw scores from the survey responses were calibrated onto a single, standardized scale. On the Safe and Orderly School scale, a higher average score indicates that teachers in a country reported fewer issues related to safety and disorderly behavior.
On all of the scales shown in this section, the centerpoint (denoted international average) is ten, and the standard deviation is two. For each contextual factor, readers are provided with information on what it means to score “higher” on the scale. To make the scale data for each contextual factor more interpretable, TIMSS categorizes responses into high, medium, and low endorsement levels. For example, on the TIMSS Safe and Orderly School scale highlighted above, a scaled score of 10.7 or higher indicates that, on average, teachers in a country perceive that their schools are “safe and orderly”; a scaled score between 6.9 and 10.6 indicates a perception that schools are “somewhat safe and orderly”; and a scaled score of 6.8 or lower indicates a perception that schools are “not safe and orderly.” In this section, the scaled scores associated with high, medium and low endorsement levels, when applicable, are denoted on each figure.

Where possible, student achievement associated with the high and low endorsement levels is reported. Any effect sizes related to achievement data are reported with respect to the international standard deviation of 100. All data reported in this section are taken from the two-volume TIMSS report.

1. Home Environment Support
Research has shown the importance of the home environment in supporting student learning, with the presence of more resources within the home associated with higher achievement (TIMSS, 2011a; TIMSS, 2011b). This section addresses contextual factors that make up a supportive home environment.

Home Educational Resources (Student Report)
Students were surveyed about three “resources”: their parents’ level of education, the number of books in the home, and the presence of an internet connection and/or the availability of one’s own room as home study supports. Survey results were reported on the Home Educational Resources scale, which provides an overall measure of the richness of the home environment to support student learning. Figure 11 below shows the average scaled scores for Massachusetts and the comparison countries; a higher scaled score indicates a greater availability of educational resources in the home.
Summary of Home Environment Support: Home Educational Resources:

- Students from Massachusetts, on average, have access to a moderate amount of home resources to support their education.
- When compared to students from the United States as a whole, Massachusetts students have access to relatively more home resources to support their education; however, the difference in average scaled scores is equivalent to only a small-to-moderate effect (approximately a third of a standard deviation).
- The average scaled score for Massachusetts (11.5) is between a half and two thirds of a standard deviation higher than the average scaled score for Chinese Taipei (10.4) and Singapore (10.3) and three quarters of a standard deviation higher than the international average (10). These differences represent moderate-to-large effects.
- In Massachusetts, students from relatively well-resourced homes scored, on average, a full standard deviation higher (103 points) in mathematics than students from homes with

*See Exhibit 4.3 for data associated with figure. The Home Educational Resources scale ranged from a low of 7.9 (Ghana) to a high of 11.6 (Norway).
few resources, a large effect. The effect is even larger in science, where the difference in achievement (137 points) equals one and a third standard deviations.

**Students’ Educational Expectations (Student Report)**

TIMSS 2011 asked students about their educational aspirations after high school. Table 9 shows the percentage of students expecting to complete a university degree or a postgraduate degree (such as a doctorate or master’s).

<table>
<thead>
<tr>
<th>Key Resource Sub-Factor</th>
<th>MA</th>
<th>U.S.</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
<th>International Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students’ educational expectations: university or postgraduate degree</td>
<td>85%</td>
<td>83%</td>
<td>69%</td>
<td>73%</td>
<td>73%</td>
<td>56%</td>
</tr>
</tbody>
</table>

*See Exhibit 4.8 for average student achievement associated with this sub-factor.*

**Summary of Home Environment Support: Students’ Educational Expectations:**

- In Massachusetts, 85% of eighth graders plan to pursue a university or postgraduate degree. This percentage is two points higher than the percentage in the United States (83%), and 12–16 points higher than the percentages in the top-achieving East Asian countries of Singapore (69%), Chinese Taipei (73%), and the Republic of Korea (73%).
- Students in Massachusetts who expect to go to a four-year college or university outperformed their peers who expect to end their education at the end of or during high school. The differences in mathematics scores (63 points) and science scores (67 points) are both equivalent to approximately two thirds of a standard deviation (a moderate-to-large effect).

**2. School Resources**

Students who attend well-resourced schools and whose peers are from more affluent backgrounds are associated with higher achievement (TIMSS, 2011a; TIMSS, 2011b). The location of the school and students’ socio-economic backgrounds are external factors that can impact student learning. In addition, in-school resources also impact student learning; these factors include the availability of school and classroom resources (reported by principals) and teacher working conditions (reported by teachers).
School Location (Principal Report)

A school’s location can influence student achievement. The school may reside in a disadvantaged neighborhood or have limited access to resources (e.g., libraries). Data related to school location, gathered from the survey of principals, are shown in Table 10.

Table 10
School Location (Percentage of Students)

<table>
<thead>
<tr>
<th>Key Resource Sub-Factors</th>
<th>MA</th>
<th>U.S.</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>School located in city/town with population greater than 100,000</td>
<td>9%</td>
<td>30%</td>
<td>100%</td>
<td>63%</td>
<td>87%</td>
</tr>
<tr>
<td>School located in city/town with population 15,001–100,000</td>
<td>67%</td>
<td>43%</td>
<td>0%</td>
<td>34%</td>
<td>10%</td>
</tr>
<tr>
<td>School located in city/town with population less than 15,000</td>
<td>24%</td>
<td>27%</td>
<td>0%</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>

*See Exhibit 5.2 for average student achievement associated with these sub-factors

Summary of School Resources: School Location:

- Two-thirds of Massachusetts eighth graders who participated in TIMSS 2011 attend schools in medium-sized cities or towns (those with a population of 15,001 to 100,000). This contrasts with Singapore, Chinese Taipei, and the Republic of Korea, where most students are located in schools within large cities (population of greater than 100,000). Just over 40% of participating U.S. students reside in medium-sized cities or towns; the international average for this category was 28%.

- Massachusetts students whose schools are within urban areas scored almost two thirds of a standard deviation lower in mathematics (61 points) than students whose schools are within small suburban/rural areas; this is a moderate-to-large effect. In science, the difference in achievement (86 points) is greater than eight tenths of a standard deviation (a large effect).
Research has shown that if students from disadvantaged economic backgrounds attend schools where the student body is, on average, composed of students from more affluent homes, these students will achieve at a higher level when compared to disadvantaged students who attend schools where the student body is, on average, from less affluent homes (TIMSS, 2011a; TIMSS, 2011b). TIMSS asked school principals to identify the percentage of students in their schools from affluent and disadvantaged homes. Based on principals’ responses, schools were categorized as “More Affluent,” “More Disadvantaged,” or “Neither More Affluent nor More Disadvantaged.” Table 11 shows the percentage of students within each category of schools.

Summary of School Resources: School Composition by Student Economic Background:

- In Massachusetts, almost a third of eighth graders (29%) attend schools categorized as “more affluent”; 26% attend schools categorized as “more disadvantaged”.
- In the United States as a whole, the percentage of students enrolled in schools categorized as “more disadvantaged” (55%) is more than twice the percentage in Massachusetts. In Singapore (11%) and Chinese Taipei (14%), this percentage is approximately half the percentage in Massachusetts.
- In the Republic of Korea, the percentage of students enrolled in schools categorized as “more disadvantaged” (32%) is comparable to Massachusetts.
- In Massachusetts, students enrolled in “more affluent” schools scored, on average, 68 points higher in mathematics than those enrolled in schools categorized as “more disadvantaged”; this is equivalent to a moderate-to-large effect size. In science, the score difference (89 points) corresponds to a large effect size.

<table>
<thead>
<tr>
<th>Economic Background (Principal Report)</th>
<th>Math Scaled Score Difference</th>
<th>Science Scaled Score Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Affluent:</td>
<td>589 (9.1; 29%)</td>
<td>599 (7.6; 29%)</td>
</tr>
<tr>
<td>More Disadvantaged:</td>
<td>521 (13.4; 26%)</td>
<td>510 (14.9; 26%)</td>
</tr>
</tbody>
</table>

Students who attend schools classified as “more affluent” scored, on average, 68 points higher in mathematics and 89 points higher in science than students who attend schools classified as “more disadvantaged.”
Table 11
School Composition by Student Economic Background (Percentage of Students)

<table>
<thead>
<tr>
<th>Key Resource Sub-Factors</th>
<th>MA</th>
<th>U.S.</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students from schools categorized as “More Affluent”</td>
<td>29%</td>
<td>22%</td>
<td>27%</td>
<td>17%</td>
<td>18%</td>
</tr>
<tr>
<td>Students from schools categorized as “Neither More Affluent nor More Disadvantaged”</td>
<td>45%</td>
<td>23%</td>
<td>61%</td>
<td>69%</td>
<td>51%</td>
</tr>
<tr>
<td>Students from schools categorized as “More Disadvantaged”</td>
<td>26%</td>
<td>55%</td>
<td>11%</td>
<td>14%</td>
<td>32%</td>
</tr>
</tbody>
</table>

*See Exhibit 5.4 for average student achievement associated with these sub-factors

Impact on Instruction (Principal Report)

Principals were asked to rate the extent to which mathematics and science instruction at their school was impacted by a lack of resources. Principals reported on approximately a dozen school and classroom resources, including instructional materials (e.g., textbooks), supplies (e.g., pencils), heating/cooling/lighting, buildings, instructional space, staff, and computers. TIMSS then calibrated the raw scores from the survey responses onto two standardized scales: the Mathematics Resource Shortages scale and the Science Resource Shortages scale. On both scales, a higher scaled score indicates less impact of school resource shortages on instruction. Figure 12 and Figure 13 below show the average scaled scores for mathematics and science respectively.

Summary of School Resources: Impact on Instruction:
- In Massachusetts and the comparison countries, principals report a similar degree of impact of school and classroom resource shortages on instruction for mathematics and science.
- Using TIMSS criteria, in Massachusetts, instruction in mathematics is, on average, not affected by resource shortages; in the United States as a whole, instruction in mathematics is “somewhat affected.” However, the difference in scaled scores (0.1) between Massachusetts and the United States is of a very small effect. Principals in Massachusetts and the United States as a whole report

**Massachusetts**

Impact of Shortages on Instruction (Principal Report)

Students whose principals indicated no impact of resource shortages on instruction scored, on average, 20 points higher in mathematics and 29 points higher in science than students whose principals reported some impact.

**Math Scaled Score Difference**
- Not Affected: 571 (8.8; 49%)
- Somewhat Affected: 551 (9.4; 50%)

**Science Scaled Score Difference**
- Not Affected: 584 (10.2; 42%)
- Somewhat Affected: 555 (9.1; 57%)

First number in parentheses is standard error; second number is % of students in each category.
that instruction in science is “somewhat affected” by a shortage of resources.

- Principals in Singapore and the Republic of Korea indicate that instruction in mathematics and science is not impacted by inadequate resources. In contrast, principals in Chinese Taipei feel that resource shortages somewhat affect instruction. Compared to Chinese Taipei, Massachusetts’s average scaled score was a quarter to a third of a standard deviation higher in mathematics (a small-to-moderate effect) and a fifth of a standard deviation higher in science.

- In Massachusetts, students of principals who reported no impact on instruction due to resource shortages scored, on average, 20 points higher in mathematics and 29 points higher in science than students of principals who reported some impact. These scaled score differences correspond to a small effect and a small-to-moderate effect for mathematics and science respectively (see sidebar).

**Figure 12**

*See Exhibit 5.9 for data associated with figure. The scale ranged from a low of 8.3 (Ukraine) to a high of 11.9 (Slovenia).*
Principals were asked to report on two resources specific to science: the availability of science laboratories within schools, and the availability of assistance for teachers when students are conducting science experiments. Table 12 reports on these factors for Massachusetts and the comparison countries.

Table 12
School Resources: Access to Science Labs and Lab Support
(Percentage of Students)

<table>
<thead>
<tr>
<th>Science Resource Sub-Factors</th>
<th>MA</th>
<th>U.S.</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools have a science laboratory</td>
<td>83%</td>
<td>81%</td>
<td>100%</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>Teachers do <strong>not</strong> have assistance during experiments</td>
<td>67%</td>
<td>68%</td>
<td>11%</td>
<td>12%</td>
<td>37%</td>
</tr>
</tbody>
</table>

*See Science Exhibit 5.16 for average student achievement associated with these sub-factors.

**Summary of School Resources: Access to Science Laboratories and Support:**
- Four-fifths of eighth graders in Massachusetts (83%) have access to a science laboratory; this is similar to the percentage in the United States (81%) but markedly lower than percentages in the top-achieving East Asian countries of Singapore (100%), Chinese Taipei (99%), and the Republic of Korea (100%). The international average is 80%.
Two out of every three eighth graders in Massachusetts (67%) do not have access to supplemental assistance (other than their teacher) when conducting science experiments. This percentage is comparable to the percentage for the United States (68%). Supplemental assistance during science experiments is much more common in Singapore, Chinese Taipei, and the Republic of Korea.

In Massachusetts, students who have access to a science laboratory scored, on average, 46 points higher on the TIMSS science assessment than students who don’t (see sidebar). This difference is equivalent to almost half of a standard deviation, a moderate effect. Students who received supplemental assistance during science experiments scored, on average, 14 points lower than students who didn’t (a small effect).

**Teacher Working Conditions (Teacher Report)**

TIMSS 2011 surveyed math and science teachers for their views on the working conditions in their schools. Five factors that make up teacher working conditions were assessed: building condition, classroom overcrowding, teaching hours, adequacy of workspace, and materials and supplies. TIMSS then calibrated the raw scores from the survey responses onto one standardized scale: the Teacher Working Conditions scale. Figure 14 and Figure 15 below show the average scaled scores for mathematics teachers and science teachers, respectively. In both figures, higher scores are indicative of fewer problems with working conditions.

**Summary of School Resources: Teacher Working Conditions:**

- Mathematics teachers in Massachusetts and in the United States as a whole report similar, relatively positive working conditions (on average, Massachusetts teachers report “hardly any” problems with their working conditions).

- In comparison to teachers from East Asian countries, mathematics teachers in Massachusetts rate their working conditions more highly, with the difference (2.8 scaled score points) between Massachusetts and the Republic of Korea being equivalent to almost one and a half standard deviations (a large effect).

- Massachusetts science teachers and their peers within

**Massachusetts Teacher Working Conditions (Teacher Report)**

Students whose teachers indicated hardly any problems with their working conditions scored, on average, 87 points higher in mathematics and 61 points higher in science than students whose teachers reported moderate problems.

**Math Scaled Score Difference**

- **Hardly Any Problems:** 563 (7.5; 53%)
- **Moderate Problems:** 476 (13.1; 4%)

**Science Scaled Score Difference**

- **Hardly Any Problems:** 575 (11.1; 41%)
- **Moderate Problems:** 514 (26.9; 6%)

First number in parentheses is standard error; second number is % of students in each category.
the United States as a whole report comparable working conditions. Both sets of science teachers indicate that, on average, they experience some “minor problems” with their working conditions.

- In comparison to teachers from East Asian countries, Massachusetts science teachers rate their working conditions more highly, with the difference (2.0 points) between Massachusetts and the Republic of Korea being equivalent to one full standard deviation (a large effect).
- Students of teachers who report hardly any problems with their working conditions scored, on average, 87 points higher in mathematics (a large effect) and 61 points higher in science (a moderate-to-large effect) than students of teachers who report moderate problems. Roughly 5% of Massachusetts teachers report having moderate problems with their working conditions.

**Figure 14**

Mathematics Teacher Working Conditions (2011)

- Hardly Any Problems: ≥ 11.7; Minor Problems: 9.0 - 11.6; Moderate Problems: ≤ 8.9

*See Exhibit 5.11 for data associated with figure. The Teacher Working Conditions scale in mathematics ranged from 7.7 (Botswana) to 12.5 (Florida).*
3. School Climate
TIMSS 2011 reported on several contextual factors related to school climate. Factors included in this summary are: (1) principals’ and teachers’ perceptions of the emphasis their schools place on academic success; (2) principals’ perceptions of school discipline and safety; (3) teachers’ views of the safety and orderliness of their school environment; and (4) students’ self-reported frequency of bullying.

**School Emphasis on Academic Success (Principal and Teacher Report)**
A positive school environment has been shown to lessen the impact on student learning of coming from a disadvantaged background (TIMSS, 2011a; TIMSS, 2011b). TIMSS asked both principals and teachers (mathematics and science) to report on the emphasis their schools place on academic success, as indicated by the following factors: (a) teachers’ understanding of curricular goals, (b) teachers’ degree of success in implementing the school’s curricula, (c) teachers’ expectations for student achievement, (d) parental support for student achievement, and (e) students’ desire to do well. TIMSS used the reports from principals and teachers to create a scale called the School Emphasis on Academic Success scale, which incorporates all of the indicators listed above. The figures below show how Massachusetts and the comparison countries scored on the scale based on principal reports (Figure 16), mathematics teacher reports (Figure 17), and science teacher reports (Figure 18). In all of the figures, a higher scaled score indicates more school emphasis on academic success.
Figure 16

School Emphasis on Academic Success (2011)
Principal Reports

Average scores on this scale ranged from 8.0 (Tunisia) to 11.8 (Dubai, UAE).

*See Exhibit 6.2 for data associated with figure.

Figure 17

School Emphasis on Academic Success (2011)
Mathematics Teacher Reports

Average scores ranged from 8.5 (Morocco) to 11.4 (Qatar and Alberta, Canada).

*See Exhibit 6.4 for data associated with figure.
Summary of School Climate: School Emphasis on Academic Success:

- Massachusetts principals and teachers rate their schools’ emphasis on academic success similarly (on average, a “high emphasis” on academic success).
- Compared to their peers in the United States, Singapore, and the Republic of Korea, Massachusetts principals report slightly more emphasis on academic success. The difference in average scaled scores between Massachusetts and the United States is equal to a quarter of a standard deviation (a small-to-moderate effect). The difference between Massachusetts and the Republic of Korea is equivalent to approximately a third of a standard deviation (a small-to-moderate effect).
- Massachusetts mathematics and science teachers report more emphasis on academic success, compared to their peers in the United States as a whole; these differences are equal to a quarter of a standard deviation for mathematics (a small effect) and four
tenths of a standard deviation for science (a small-to-moderate effect).
• Similarly, Massachusetts teachers (both mathematics and science) report more emphasis on academic success when compared to their East Asian counterparts, with the difference between Massachusetts and Singapore in each subject equal to just over a half of a standard deviation (a moderate effect).
• In Massachusetts, students of teachers who reported a “very high” school emphasis on academic success scored, on average, approximately a half of a standard deviation higher in mathematics (47 points) and science (56 points) when compared to students of teachers who reported a “medium” emphasis on academic success (see sidebar). Similarly, students of principals who reported a “very high” emphasis on academic success scored, on average, 30 points higher (small-to-moderate effect) in mathematics and 36 points higher (small-to-moderate effect) in science when compared to students of principals who reported “medium” emphasis (data not shown).

Discipline and School Safety (Principal Report)
TIMSS 2011 surveyed principals on eleven measures of school discipline and safety, asking to what degree each was a problem in the eighth grade of their schools. Factors such as tardiness, absenteeism, cheating, vandalism, and intimidation of teachers and fellow students were assessed. TIMSS then combined the survey results on a scale called the School Discipline and Safety scale. Figure 19 below shows how Massachusetts and the comparison countries scored on the scale. A higher scaled score indicates fewer discipline and safety problems were reported.

Summary of School Climate: Discipline and School Safety:
• Principals from both Massachusetts and the United States as a whole indicate that they experience, on average, “minor problems” with discipline and safety issues in their schools.
• On average, Massachusetts principals report fewer issues with school discipline and safety than their peers in the United States as a whole; the difference in scaled scores (0.5) is equal to a quarter of a standard deviation (a small-to-moderate effect).
• On average, principals in the Republic of Korea report more problems with school discipline and safety than Massachusetts principals, with the

Massachusetts School Discipline and Safety (Principal Report)
Students whose principals reported hardly any problems with school discipline and safety scored, on average, 55 points higher in mathematics and 68 points higher in science than students whose principals reported moderate problems.

Math Scaled Score Difference
Hardly Any Problems: 571 (10.9; 23%)
Moderate Problems: 516 (20.7; 11%)

Science Scaled Score Difference
Hardly Any Problems: 581 (9.7; 23%)
Moderate Problems: 513 (25.7; 11%)
difference (0.5 points) between Massachusetts and the Republic of Korea being equivalent to a quarter of a standard deviation (a small-to-moderate effect). In contrast, principals in Chinese Taipei and Singapore report fewer problems. The difference in average scaled scores (0.8) between Massachusetts and Chinese Taipei is equivalent to four tenths of a standard deviation (a small-to-moderate effect).

- In Massachusetts, students of principals who reported hardly any problems with discipline and school safety outperformed their peers whose principals reported moderate problems. The difference in achievement amounted to just over a half of a standard deviation in mathematics (55 points; a moderate-to-large effect) and two thirds of a standard deviation in science (68 points; a moderate-to-large effect).

**Figure 19**

<table>
<thead>
<tr>
<th>Average Scaled Score</th>
<th>U.S.</th>
<th>MA</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
<th>International Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1</td>
<td>10.6</td>
<td>10.9</td>
<td>11.4</td>
<td>10.1</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*See Exhibit 6.10 for data associated with figure. Average scores ranged from 7.5 (Syrian Arab Republic) to 11.8 (Kazakhstan).

**Safe and Orderly School (Teacher Report)**

TIMSS 2011 also asked teachers to rate the safety and orderliness of their school environment. Five measures of school safety were assessed: the safety of the neighborhood in which the school is located, the school’s security policies and practices, student behavior, students’ respect for teachers, and teachers’ personal feelings of safety. The figures below show how Massachusetts and the comparison countries scored on the Safe and Orderly School scale as reported by mathematics teachers (Figure 20) and science teachers (Figure 21). A higher scaled score indicates that fewer safety problems were reported.
**Figure 20**

Safe and Orderly School (2011)
Mathematics Teacher Reports

Average Scaled Score

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Scaled Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>10.4</td>
</tr>
<tr>
<td>MA</td>
<td>11</td>
</tr>
<tr>
<td>Singapore</td>
<td>10.7</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>9.1</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>8.5</td>
</tr>
<tr>
<td>International Average</td>
<td>10</td>
</tr>
</tbody>
</table>

Safe and Orderly: ≥ 10.7; Somewhat Safe and Orderly: 6.9 - 10.6; Not Safe and Orderly: ≤ 6.8

*See Exhibit 6.8 for data associated with figure. Average scores ranged from 8.0 (Botswana) to 11.4 (Minnesota).

**Figure 21**

Safe and Orderly School (2011)
Science Teacher Reports

Average Scaled Score

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Scaled Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>10.2</td>
</tr>
<tr>
<td>MA</td>
<td>10.9</td>
</tr>
<tr>
<td>Singapore</td>
<td>10.6</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>9.2</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>8.4</td>
</tr>
<tr>
<td>International Average</td>
<td>10</td>
</tr>
</tbody>
</table>

Safe and Orderly: ≥ 10.7; Somewhat Safe and Orderly: 6.9 - 10.6; Not Safe and Orderly: ≤ 6.8

*See Exhibit 6.8 for data associated with figure. Average scores ranged from 7.9 (Botswana) to 11.5 (Minnesota).
Summary of School Climate: Safe and Orderly School:

- Mathematics and science teachers in Massachusetts report, on average, being within “safe and orderly” schools.
- On average, teachers in Massachusetts report fewer problems with the safety and orderliness of their school environment when compared to teachers in the United States as a whole. The difference in average scaled scores between Massachusetts and the United States corresponds to approximately a third of a standard deviation for both mathematics and science teacher reports (a small-to-moderate effect).
- Teachers from the Republic of Korea report more problems with the safety and orderliness of their schools than teachers from Massachusetts or from the other comparison countries. In both mathematics and science, the scaled score difference (2.5 points) between Massachusetts and the Republic of Korea is equivalent to one and a quarter standard deviations (a large effect).
- In Massachusetts, students of teachers who reported a safe and orderly school environment outperformed their peers whose teachers reported an unsafe environment (see sidebar). The difference in achievement amounted to just over three quarters of a standard deviation in both mathematics (83 points) and science (85 points). These differences are of a large effect size.

Massachusetts
School Safety and Order
(Teacher Report)

Students whose teachers reported a safe and orderly school environment scored, on average, 83 points higher in mathematics and 85 points higher in science than students whose teachers reported an unsafe and disorderly environment.

Math Scaled Score Difference
Safe and Orderly:  573 (6.9; 65%)
Not Safe and Orderly:  490 (13.9; 7%)

Science Scaled Score Difference
Safe and Orderly:  575 (7.4; 61%)
Not Safe and Orderly:  490 (42.2; 4%)

First number in parentheses is standard error; second number is % of students in each category.

Students Bullied at School (Student Report)

In the student survey, students were asked to report how often they had experienced six bullying behaviors: being made fun of; being left out of games; being the subject of lies spread by another student; having something stolen; being hit or hurt by another student; and being made to do things they didn’t want to do by other students. Based on student responses, TIMSS created the Students Bullied at School scale. Figure 22 shows how Massachusetts and the comparison countries scored on the scale. A higher scaled score indicates that, in the country in question, the average student experiences fewer incidences of bullying.
Summary of School Climate: Students Bullied at School:

- On average, eighth graders in Massachusetts report being bullied “almost never.” The survey results did show that 6% of Massachusetts eighth graders report being bullied “about weekly.”
- Students in Massachusetts report fewer bullying incidents, on average, than their peers in the United States as a whole. The average scaled score difference between Massachusetts and the United States is of a small effect size (equivalent to a fifth of a standard deviation).
- Like Massachusetts students, students in the three East Asian countries report “almost never” being bullied in school. However, Singapore’s average scaled score is below the international average. The scaled score difference between Singapore and Massachusetts is equal to four tenths of a standard deviation, a small-to-moderate effect.
- In Massachusetts, the difference in achievement (30 points in mathematics, and 26 points in science) between students who almost never experience bullying and those who experience it weekly is equivalent to a small-to-moderate effect (see sidebar).
4. Teacher Preparation
TIMSS 2011 surveyed teachers on a variety of measures of preparedness, including years of experience in the field, teacher confidence in their subject, and career satisfaction. These data are reported in this section.

Years of Teaching Experience (Teacher Report)
Teachers were surveyed about their number of years of teaching experience. Table 13 and Table 14 report these data for mathematics and science teachers, respectively. The tables show the percentage of eighth graders in Massachusetts and the comparison countries taught by teachers at each level of experience.

Table 13
Teacher Preparation: Mathematics Teachers’ Years of Experience (Percentage of Students)

<table>
<thead>
<tr>
<th>Key Sub-Factors: Math Teachers’ Years of Experience</th>
<th>MA</th>
<th>U.S.</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 years or more</td>
<td>10%</td>
<td>26%</td>
<td>10%</td>
<td>24%</td>
<td>34%</td>
</tr>
<tr>
<td>At least 10 but less than 20 years</td>
<td>33%</td>
<td>28%</td>
<td>16%</td>
<td>41%</td>
<td>22%</td>
</tr>
<tr>
<td>At least 5 but less than 10 years</td>
<td>39%</td>
<td>28%</td>
<td>26%</td>
<td>26%</td>
<td>17%</td>
</tr>
<tr>
<td>Less than 5 years</td>
<td>18%</td>
<td>17%</td>
<td>47%</td>
<td>9%</td>
<td>27%</td>
</tr>
<tr>
<td>Average years of experience</td>
<td>11</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

*See Exhibit 7.6 for average student achievement associated with these sub-factors.

Table 14
Teacher Preparation: Science Teachers’ Years of Experience (Percentage of Students)

<table>
<thead>
<tr>
<th>Key Sub-Factors: Science Teachers’ Years of Experience</th>
<th>MA</th>
<th>U.S.</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 years or more</td>
<td>17%</td>
<td>24%</td>
<td>13%</td>
<td>28%</td>
<td>42%</td>
</tr>
<tr>
<td>At least 10 but less than 20 years</td>
<td>37%</td>
<td>38%</td>
<td>17%</td>
<td>28%</td>
<td>17%</td>
</tr>
<tr>
<td>At least 5 but less than 10 years</td>
<td>38%</td>
<td>21%</td>
<td>25%</td>
<td>26%</td>
<td>20%</td>
</tr>
<tr>
<td>Less than 5 years</td>
<td>9%</td>
<td>16%</td>
<td>46%</td>
<td>18%</td>
<td>21%</td>
</tr>
<tr>
<td>Average years of experience</td>
<td>13</td>
<td>14</td>
<td>8</td>
<td>13</td>
<td>15</td>
</tr>
</tbody>
</table>

*See Exhibit 7.6 for average student achievement associated with these sub-factors.

Summary of Teacher Preparation: Years of Teaching Experience:
- In Massachusetts, eighth-grade mathematics teachers have, on average, relatively less teaching experience than their science counterparts.
- In mathematics and science, the vast majority of eighth graders in Massachusetts are taught by teachers in the two middle reporting categories (i.e., from 5 to less than 20
years of experience), indicating a relatively mature teaching body. In both disciplines, the United States, Chinese Taipei, and the Republic of Korea have a lower overall percentage of students taught by teachers in these middle categories, but a higher percentage taught by teachers with 20 years or more of experience.

- Of note, Singapore has a relatively young teaching body in both disciplines, with almost half of Singaporean students being taught by teachers with less than 5 years of experience.
- In Massachusetts, students of relatively inexperienced mathematics teachers (< 5 years) performed comparably to students of teachers with the greatest experience (≥ 20 years). In science, students of relatively inexperienced teachers outperformed their peers by 45 points when compared to students taught by the most experienced teachers. This difference is equivalent to almost half a standard deviation (a moderate effect).

**Teacher Confidence (Teacher Report)**

Teachers were asked how confident they feel in performing routine teaching-related activities with students, such as answering student questions, explaining concepts, providing challenging tasks for capable students, adapting teaching to engage students, and increasing students’ appreciation of mathematics and science. Teacher responses were placed on one scale for mathematics (Confidence in Teaching Mathematics scale) and one for science (Confidence in Teaching Science scale). The data for Massachusetts and the comparison countries are shown in Figure 23 (mathematics) and Figure 24 (science). A higher scaled score indicates greater teacher confidence.

**Massachusetts**

**Years of Teaching Experience (Teacher Report)**

Students taught by the least experienced teachers scored, on average, 10 points lower in mathematics but 45 points higher in science than students taught by the most experienced teachers.

**Math Scaled Score Difference**

- 20 years or more: 566 (20.3; 10%)
- Less than 5 years: 556 (17.9; 18%)

**Science Scaled Score Difference**

- 20 years or more: 549 (24.3; 17%)
- Less than 5 years: 594 (24.3; 9%)

*First number in parentheses is standard error; second number is % of students in each category.*
**Figure 23**

Teacher Confidence in Teaching Mathematics (2011)

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Scaled Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>10.6</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>10.8</td>
</tr>
<tr>
<td>Singapore</td>
<td>9.1</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>9.4</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>8.6</td>
</tr>
<tr>
<td>International Average</td>
<td>10</td>
</tr>
</tbody>
</table>

*See Exhibit 7.13 for data associated with figure. The Confidence in Teaching Mathematics scale ranged from 8.0 (Japan) to 11.5 (Kazakhstan).

**Figure 24**

Teacher Confidence in Teaching Science (2011)

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Scaled Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>10.5</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>10.7</td>
</tr>
<tr>
<td>Singapore</td>
<td>9.4</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>9.5</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>8.4</td>
</tr>
<tr>
<td>International Average</td>
<td>10</td>
</tr>
</tbody>
</table>

*See Exhibit 7.13 for data associated with figure. The Confidence in Teaching Science scale ranged from 7.9 (Japan) to 11.7 (Kazakhstan).
Summary of Teacher Preparation: Teacher Confidence:

- Eighth-grade mathematics and science teachers in Massachusetts and in the United States as a whole are “very confident” in the teaching of their subjects.

- In general, Massachusetts’ mathematics and science teachers are relatively more confident in their ability to teach their subjects than teachers from the three East Asian countries. In both disciplines, the difference in average scaled scores between Massachusetts and the Republic of Korea is over one standard deviation, which corresponds to a large effect size.

- In Massachusetts, students of mathematics teachers who expressed being “somewhat confident” scored higher on average than students of teachers who expressed being “very confident”; the difference of 26 scaled score points is equivalent to a quarter of a standard deviation (a small-to-moderate effect).

- In Massachusetts, students of science teachers who expressed being “somewhat confident” scored, on average, comparably to students of science teachers who expressed being “very confident” in their teaching.

**Teacher Career Satisfaction (Teacher Report)**

Mathematics and science teachers were asked about their satisfaction with their career in teaching. Teachers are more motivated and ready to teach if they are satisfied with their jobs and working conditions (TIMSS, 2011a; TIMSS, 2011b). Six viewpoints were assessed: contentment with their profession; satisfaction with their current school; enthusiasm for teaching compared to when they started; desire to continue teaching for as long as possible; frustration level with being a teacher; and the importance they attribute to their work. These measures were combined on the Teacher Career Satisfaction scale. The data for Massachusetts and the comparison countries are shown below in Figure 25 (mathematics) and Figure 26 (science). A higher scaled score indicates greater teacher satisfaction with their careers.
Figure 25

Teacher Career Satisfaction (2011)
Mathematics Teacher Reports

*Satisfactory: ≥ 10.4; Somewhat Satisfied: 7.1 - 10.3; Less Than Satisfied: ≤ 7.0

Average Scaled Score

<table>
<thead>
<tr>
<th></th>
<th>U.S.</th>
<th>MA</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
<th>International Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>10.1</td>
<td>10.3</td>
<td>9.2</td>
<td>9.4</td>
<td>8.2</td>
<td>10</td>
</tr>
</tbody>
</table>

*See Exhibit 7.16 for data associated with figure. The Teacher Career Satisfaction scale in mathematics ranged from 8.2 (the Republic of Korea) to 12.3 (Honduras).

Figure 26

Teacher Career Satisfaction (2011)
Science Teacher Reports

*Satisfactory: ≥ 10.4; Somewhat Satisfied: 7.1 - 10.3; Less Than Satisfied: ≤ 7.0

Average Scaled Score

<table>
<thead>
<tr>
<th></th>
<th>U.S.</th>
<th>MA</th>
<th>Singapore</th>
<th>Chinese Taipei</th>
<th>Republic of Korea</th>
<th>International Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>9.7</td>
<td>10.2</td>
<td>9.2</td>
<td>9.6</td>
<td>8.3</td>
<td>10</td>
</tr>
</tbody>
</table>

*See Exhibit 7.16 for data associated with figure. The Teacher Career Satisfaction scale in science ranged from 8.3 (the Republic of Korea) to 12.2 (Honduras).
Summary of Teacher Preparation: Teacher Career Satisfaction:

- Eighth-grade mathematics and science teachers in Massachusetts are, on average “somewhat satisfied” with their careers, with both groups expressing comparable levels of satisfaction.
- When compared to their counterparts in the United States as a whole, mathematics teachers in Massachusetts report comparable satisfaction levels. Science teachers in Massachusetts express higher levels of career satisfaction than science teachers in the United States as a whole; the difference in average scaled scores (0.5) is equivalent to a quarter of a standard deviation (a small-to-moderate effect).
- In general, Massachusetts’ mathematics and science teachers are relatively more satisfied with their careers than teachers from the three East Asian countries. In both disciplines, the difference in average scaled scores between Massachusetts and the Republic of Korea is equal to approximately one standard deviation, which corresponds to a large effect size.
- In Massachusetts, student achievement in mathematics and science was comparable across satisfaction groups (data not shown).

5. Classroom Instruction

Teachers and students were surveyed on several indicators that relate to classroom instruction in mathematics and science. For example, teachers reported on the degree to which they collaborate to improve their teaching and how often they use practices that are thought to engage students in learning. Students were asked about the following indicators of classroom instruction: (1) the degree to which they value math and science; (2) the degree to which they feel engaged by their learning; and (3) their level of confidence in learning mathematics and science. In the past, research has shown that students who have a more positive attitude toward their classroom learning tend to perform better on assessments (TIMSS, 2011a; TIMSS, 2011b). The data for these classroom instruction indicators are reported in this section for Massachusetts and the comparison countries.

Teacher Collaboration (Teacher Report)

Teachers reported on the extent to which they interact with other teachers to plan instruction, improve teaching strategies, try out new ideas, and observe their peers to learn more about teaching. These measures were placed on the Collaborate to Improve Teaching scale. The data for Massachusetts and the comparison countries are shown in Figure 27 (mathematics) and Figure 28 (science). A higher scaled score indicates greater teacher collaboration.
The Collaborate to Improve Teaching scale in mathematics ranged from 8.2 (Morocco) to 11.4 (Israel).

The Collaborate to Improve Teaching scale in science ranged from 8.4 (Morocco) to 11.2 (Qatar).
Summary of Classroom Instruction: Teacher Collaboration:

- Eighth-grade mathematics and science teachers in Massachusetts are, on average, “collaborative”; however, the average scaled scores in mathematics and science are both below the international average.

- Mathematics and science teachers in Massachusetts report lower collaboration rates than their peers in the United States as a whole; in both subjects, the difference in average scaled scores between Massachusetts and the United States is of a small-to-moderate effect size.

- In general, teachers in Massachusetts and Singapore report comparable levels of collaboration. When compared to teachers in Chinese Taipei, Massachusetts teachers report relatively more collaboration. The differences in average scaled scores between Massachusetts and Chinese Taipei are of a small effect for mathematics and of a small-to-moderate effect for science.

- In Massachusetts, in both mathematics and science, students of “very collaborative” teachers scored comparably to students of “somewhat collaborative” teachers (see sidebar). The difference in average achievement for both mathematics (16 points) and science (11 points) corresponds to a small effect size.

### Instruction to Engage Students in Learning (Teacher Report)

Teachers were surveyed on how often they use four teaching practices to engage students in their learning. These practices are: summarizing the lesson’s learning goals, questioning to elicit reasons and explanations, encouraging students to show improvement, and praising students for good effort. TIMSS then combined the survey results on the Engaging Students in Learning scale. The data for Massachusetts and the comparison countries are shown below in Figure 29 (mathematics) and Figure 30 (science). A higher scaled score indicates more frequent use of the engagement practices. For example, a scaled score of 8.7 or higher indicates that teachers, on average, use engagement practices in “most lessons.”
The Engaging Students in Learning scale in mathematics ranged from 8.4 (Chinese Taipei) to 11.2 (North Carolina).

The Engaging Students in Learning scale in science ranged from 8.2 (Japan) to 11.1 (Connecticut).
Summary of Classroom Instruction: Instruction to Engage Students in Learning:

- Mathematics and science teachers in Massachusetts use engagement practices, on average, in “most lessons.”
- Mathematics and science teachers in Massachusetts and in the United States as a whole report using engagement practices with a similar frequency.
- Massachusetts mathematics and science teachers report more frequent use of engagement practices than their counterparts in the three East Asian countries; the difference in average scaled scores between Massachusetts and Chinese Taipei teachers is greater than 0.8 of a standard deviation (i.e., a large effect size) in both mathematics and science.
- In Massachusetts, students of teachers who reported using engagement practices in most lessons scored, on average, lower than students of teachers who reported using them in only half of their lessons; this finding was apparent in both mathematics and science (see sidebar). The differences in average achievement for mathematics (38 points) and science (16 points) correspond to a small-to-moderate effect size and to a small effect size, respectively.

Teacher Emphasis on Science Investigation (Teacher Report)

Specific to science, teachers were surveyed on how often they asked students to perform several practices related to science investigation. These practices include: observing and describing natural phenomena; observing experiments demonstrated by the teacher; planning and conducting their own investigations; using scientific formulas and laws to solve problems; explaining their study; and relating what they are studying to their daily lives (TIMSS, 2011b). The survey responses were combined on the Emphasize Science Investigation scale. The data for Massachusetts and the comparison countries are shown in Figure 31. A higher scaled score indicates more frequent use of the science investigation practices.

Summary of Classroom Instruction: Teacher Emphasis on Science Investigation:

- On average, science teachers in Massachusetts use instructional practices related to science investigation in “less than half the lessons.”
• Science teachers in the United States as a whole place more emphasis on science investigation than teachers from Massachusetts; the difference in average scaled scores (0.8) is equivalent to four tenths of a standard deviation (a small-to-moderate effect size).

• Massachusetts science teachers report comparable use of science investigation practices in their classrooms to teachers from Singapore and Chinese Taipei. The difference in average scaled scores (0.7) between Massachusetts and the Republic of Korea is equivalent to approximately a third of a standard deviation, a small-to-moderate effect size.

• In Massachusetts, students of teachers who reported using science investigation instructional practices in about half their lessons or more scored, on average, higher than students of teachers who reported using them in less than half of their lessons. The difference in average achievement (36 points) corresponds to a small-to-moderate effect size.

*See Exhibit 8.28 for data associated with figure. The Emphasize Science Investigation scale ranged from 7.5 (Norway) to 11.3 (Jordan and Palestinian National Authority).
Students’ Valuing of Mathematics and Science (Student Report)

Students were asked to rate their level of agreement with six statements designed to measure the value they place on studying mathematics or science. These statements were related to students’ perceptions of the relevance of each subject to their daily lives and to their college and career prospects, and of how important it was to do well in each subject. Student responses were placed on the Students Value Mathematics and Students Value Science scales. The data for Massachusetts and the comparison countries are shown below in Figure 32 (mathematics) and Figure 33 (science). A higher scaled score indicates greater student valuation of their subjects.

Summary of Classroom Instruction: Value of Subject:

• On average, students in Massachusetts report that they “somewhat value” mathematics and science. The average scaled score for both subjects is at (mathematics) or just below (science) the international average.

• Students’ valuation of mathematics and science is comparable in Massachusetts and in the United States as a whole.

• When compared to students in Chinese Taipei and the Republic of Korea, Massachusetts students rate the value of mathematics and science higher; the difference in average scaled scores between Massachusetts students and Chinese Taipei students corresponds to a large effect in mathematics and a moderate-to-large effect in science.

• Students who reported that they value math and science scored higher on the TIMSS assessments than students who reported that they do not value these subjects, as shown in the sidebar. The average scaled score differences for mathematics (32 points) and science (41 points) correspond to a third of a standard deviation and fourth tenths of a standard deviation, respectively (both small-to-moderate effects).
Figure 32
Value of Mathematics (2011)
Student Reports

*See Exhibit 8.3 for data associated with figure. The Students Value Mathematics scale ranged from 8.3 (Chinese Taipei) to 11.5 (Ghana and Morocco).

Figure 33
Value of Science (2011)
Student Reports

*See Exhibit 8.3 for data associated with figure. The Students Value Science scale ranged from 8.5 (Chinese Taipei and Japan) to 11.6 (Ghana). Please note: the data in this figure relate to countries that enroll students in science as a single subject. Other countries teach eighth-grade science as separate subjects.

62
Student Engagement in Mathematics and Science Lessons

Students were asked to rate their level of agreement with five statements designed to measure the extent to which their mathematics or science lessons engaged them. These statements were related to how relevant the lessons were to students, how easy the lessons were to understand, teacher expectations, and student interest in the content and activities embedded in the lessons. Student responses were placed on the Engaged in Mathematics Lessons scale and the Engaged in Science Lessons scale. The data for Massachusetts and the comparison countries are shown below in Figure 34 (mathematics) and Figure 35 (science). A higher scaled score indicates greater student engagement.

Summary of Classroom Instruction: Student Engagement:

• On average, eighth-grade students in Massachusetts feel “somewhat engaged” in their mathematics and science lessons. Of note, the average scaled score for mathematics is below the international average of 10.
• Student engagement in mathematics and science lessons is comparable in Massachusetts and in the United States as a whole; any difference in scaled scores is of a small effect.
• When compared to students in Chinese Taipei and the Republic of Korea, Massachusetts students rate their engagement in mathematics and science lessons higher; the difference in average scaled scores between Massachusetts and the Republic of Korea corresponds to a moderate-to-large effect in mathematics and a large effect in science.
• Massachusetts students who reported being engaged in their lessons scored higher on the TIMSS assessments than students who reported not being engaged, as shown in the sidebar. The average scaled score differences in mathematics (22 points) and science (24 points) both correspond to less than a quarter of a standard deviation (both differences are of a small effect size).
The Engaged in Mathematics Lessons scale ranged from 8.0 (the Republic of Korea) to 11.4 (Armenia).

The Engaged in Science Lessons scale ranged from 8.2 (Japan) to 11.3 (Tunisia). Please note: the data in this figure relate to countries that enroll students in science as a single subject. Other countries teach eighth-grade science as separate subjects.
Student Confidence in Learning Mathematics and Science

Students were asked to rate their level of agreement with nine statements designed to assess their confidence in learning mathematics or science. These statements were related to how difficult the subject is to the student, relative to his or her peers and to other subjects, and the student’s perception of how well the teacher supports and boosts his or her confidence. Student responses were placed on the Students Confident in Mathematics scale and the Students Confident in Science scale. The data for Massachusetts and the comparison countries are shown below in Figure 36 (mathematics) and Figure 37 (science). A higher scaled score indicates greater student confidence.

Summary of Classroom Instruction: Student Confidence:

- On average, students in Massachusetts feel “somewhat confident” in learning mathematics and science.
- Student confidence in learning mathematics and science is comparable in Massachusetts and in the United States as a whole; any difference in average scaled scores is of a small effect size.
- When compared to students in Chinese Taipei and the Republic of Korea, Massachusetts students rate their confidence in learning mathematics and science higher; the difference in average scaled scores between Massachusetts and Chinese Taipei is greater than one standard deviation in both mathematics and science (both differences are of a large effect size).
- Massachusetts students who reported more confidence in their learning scored higher on the TIMSS assessments than students who were less confident, as shown in the sidebar. The average scaled score differences in mathematics and in science both correspond to approximately eight tenths of a standard deviation (a large effect size).
Figure 36

Student Confidence in Learning Mathematics (2011)

*See Exhibit 8.5 for data associated with figure. The Students Confident in Mathematics scale ranged from 8.6 (Chinese Taipei and Japan) to 11.1 (Israel).

Figure 37

Student Confidence in Learning Science (2011)

*See Exhibit 8.5 for data associated with figure. The Students Confident in Science scale ranged from 8.3 (Chinese Taipei) to 11.1 (Tunisia). Please note: the data in this figure relate to countries that enroll students in science as a single subject. Other countries teach eighth-grade science as separate subjects.
PART VI
REPORT CONCLUSION
The report conclusion is divided into two parts. The first part summarizes Massachusetts student achievement on the 2011 TIMSS assessments and trends in achievement between 1999 and 2011. The second part discusses the contextual factors that appear to impact Massachusetts student achievement in mathematics and science.

Student Achievement
In 2011, Massachusetts eighth-grade students performed well on both the mathematics and science TIMSS assessments. In both subjects, they were among the highest achievers on the assessments and performed on par with students from many of the top-achieving Asian countries. Of the 63 countries and 14 benchmarking entities that participated in the 2011 TIMSS, Massachusetts ranked fifth overall in eighth-grade mathematics and second overall in eighth-grade science. When compared to their peers from the United States as a whole, Massachusetts students did exceptionally well in both subjects, with students scoring statistically higher, on average, than their U.S. peers.

Massachusetts students’ achievement on the 2011 mathematics assessment resulted from a relatively good performance on items testing basic mathematical knowledge (above state average score) combined with a solid performance on items testing reasoning skills (at state average score). However, Massachusetts students performed relatively and significantly worse on items requiring them to apply their mathematical knowledge (e.g., by solving problems and modeling or representing solutions). Massachusetts students performed well in the Number and Data and Chance content domains and solidly on items in the Algebra domain. Massachusetts students, however, do have a relative and significant weakness in Geometry. One of the reasons Singapore, the second-ranked country in mathematics, performs so well is that Singaporean students score consistently high across all content domains when compared to Massachusetts students.

In science, Massachusetts students’ achievement resulted from their relative strength (above state average) in factual knowledge and basic science skills, combined with their solid reasoning skills. Similar to mathematics, students in Massachusetts had a relative and significant weakness in modeling their scientific knowledge (the Applying cognitive domain). Massachusetts students performed relatively well in the Biology and Earth Science content domains and at the state average for Chemistry. In contrast, students in Massachusetts were relatively and significantly weaker in the Physics content domain, where the score difference between Massachusetts and Singapore (the top-performing country) was substantial (almost 50 points).

The trends data provide a more complete picture of how well Massachusetts students are doing. Between 1999 and 2011, Massachusetts students made steady gains in both mathematics and
science; the increase in average scaled scores over that period was statistically significant in both subjects. When student performance is broken down by content domain, the data show that Massachusetts students made positive achievement gains in every content area in both subjects between 2007 and 2011 (data were only available for these two assessments). Of note, in mathematics, student performance in Geometry (a relative weakness in Massachusetts) exhibited the highest gain score (25 points) between 2007 and 2011. In science, the average scaled score also improved in Physics (a relative weakness) between 2007 and 2011, but this gain was of a smaller magnitude (16 points).

Girls and boys both contributed to Massachusetts’ improved performance on the TIMSS mathematics and science assessments between 1999 and 2011, with positive gains recorded by both genders over the time period. Although boys in Massachusetts significantly outperformed their female peers in science in 1999 and 2007, the gender gap in 2011 was not significant; this narrowing of the gender gap contributed to the overall improvement between 2007 and 2011 in this subject.

In contrast to Massachusetts students, students in the United States as a whole have not recorded improvements in their TIMSS performance over time, with their average performance essentially remaining flat in both mathematics and science. Singapore’s performance over time has been more erratic, with a decrease in achievement between the 2003 and 2007 assessments. However, the increase in achievement between 2007 and 2011 was statistically significant in both subjects, and Singaporean students have consistently outperformed Massachusetts students in all years.

The gap between Massachusetts students and Singaporean students appears to be closing over time; Massachusetts’ steadily improving performance is helping to close this gap. More students from Massachusetts are scoring in the advanced performance category; the percentage of students from Massachusetts scoring in the top performance level increased from 8% in 1999 to 19% in 2011 for mathematics and from 15% in 1999 to 24% in 2011 for science.

**Contextual Factors**

A large body of research has shown that student achievement may be impacted by contextual factors that are related to the individual student (e.g., resources in the home); to the teacher (e.g., number of years teaching); or to school resources and the school environment (e.g., school climate). The TIMSS 2011 report presents data from surveys administered to principals, teachers, and students on these kinds of contextual factors. Part V of this report summarizes results from the TIMSS surveys, focusing on the data for Massachusetts and a small group of comparison countries. Highlights from Part V are presented below, with contextual factors categorized by the size of their impact on student achievement. As described in Part I of this report, effect sizes were used to assess the significance of scaled score differences between student groups.
Readers should keep in mind that the discussion of contextual factors in this report is limited in that it is essentially descriptive in nature. The inter-relationship between contextual factors—and how this inter-relationship impacts achievement—is not reported. In order to investigate these types of inter-relationships and their influence on achievement, more sophisticated statistical analyses (e.g., multi-level linear modeling) are required. However, the descriptive statistics are illuminating and reveal that not all subgroups of Massachusetts students performed comparably on the two assessments.

Factors with Large or Moderate-to-Large Effects on Student Achievement

**Student Reports**

The contextual factors that appear to have a large impact on student achievement for students are a student’s home environment and his or her confidence in learning the subject matter. Based on the TIMSS surveys, a small percentage of Massachusetts eighth graders (4%) were categorized as having few educational resources at home. These students scored substantially lower on both assessments than well-resourced students, with the difference being equivalent to a large effect size. Student confidence in learning mathematics or science also appears to have a major impact on achievement, with the least-confident students scoring approximately eight tenths of a standard deviation lower in both subjects than their most-confident peers. Almost a third of Massachusetts eighth graders reported that they were not confident in mathematics, with a fifth of Massachusetts eighth graders expressing low confidence in science.

**Teacher Reports**

Only a small percentage of Massachusetts eighth-grade teachers (~5%) reported having moderate problems with their working conditions (e.g., over-crowded classrooms, inadequate supplies). A similar percentage felt that their schools are unsafe and disorderly. Students of these teachers scored substantially lower, on average, on both the mathematics and science TIMSS assessments when compared to students of teachers who reported a safe and orderly school environment and hardly any problems with their working conditions.

**Principal Reports**

Eleven percent of principals reported having moderate problems with discipline and safety issues in their schools. Students of these principals scored lower, on average, on the TIMSS assessments when compared to students of principals who reported hardly any problems with school discipline and safety. Students of urban principals (9% of eighth graders) also scored substantially lower, on average, than students of principals whose schools are located in small suburban/rural areas. Similarly, students whose schools have a disproportionally more disadvantaged student population scored, on average, much lower in mathematics and science than students from schools with a more affluent student population. Just over a quarter of
Massachusetts eighth graders attend schools that have a high percentage of students from disadvantaged homes.

**Factors with Moderate or Small-to-Moderate Effects on Student Achievement**

**Student Reports**
Student educational expectations in Massachusetts appear to influence student achievement in both subjects. Students who do not expect to continue their education beyond high school scored two thirds of a standard deviation lower in both subjects when compared to students who expressed a desire to go to a four-year college. Of a small-to-moderate effect on student achievement are factors that relate to the extent of bullying that occurs and how much value students place on studying mathematics and science. Students who are bullied “about weekly” (6% of the student population) and students who do not value mathematics (12% of students) or science (30% of students) scored somewhat lower on the TIMSS assessments.

**Teacher Reports**
Students of teachers who indicated that their school placed a very high emphasis on academic success (11–17% of students) outperformed their peers in both subjects when compared to students whose teachers expressed only a medium emphasis (18–25% of students). Specific to mathematics instruction, students of teachers who frequently employed certain engaging instructional practices (93% of students) performed worse on the TIMSS assessment when compared to students of teachers who employed the instructional practices less frequently (this negative effect was of a smaller magnitude in science). Also, specific to mathematics, students of teachers who reported being only somewhat confident in their teaching (8% of students) scored, on average, higher on the TIMSS assessment when compared to students of teachers who reported being very confident (92 percent of students). In contrast, teacher confidence in teaching science appeared to have no impact on student achievement.

Specific to science, students of teachers who emphasized student use of science investigation practices scored, on average, higher on the TIMSS science assessment when compared to students who used these practices less frequently. Also, specific to science, students of relatively inexperienced teachers (< 5 years teaching experience) outperformed their peers whose teachers had been teaching for 20 years or more. (In contrast, albeit of a small effect, students of relatively experienced mathematics teachers outperformed their peers whose teachers were inexperienced.)

**Principal Reports**
Students of principals who reported that their school placed a medium emphasis on academic success (22% of students) scored, on average, lower on the TIMSS assessments when compared to students of principals who reported a very high emphasis on academic success (27% of students). Specific to science, and of a moderate effect, students who had access to
laboratories (83% of students) performed higher on the TIMSS assessment than students who did not.

Factors with Small Effects on Student Achievement

**Student Reports**
Of a small effect, students who reported that they are not engaged by their learning environment (26% of students in mathematics and 18% of students in science) performed slightly less well than their peers who expressed being engaged (16% of students in mathematics and 33% of students in science).

**Teacher Reports**
Of small effect on student achievement is the teacher-related factor of teacher collaboration. Approximately a quarter of Massachusetts eighth-grade teachers reported that classroom instruction is a very collaborative process in mathematics and science. Students of these teachers performed slightly better than students of teachers who expressed that their school environment was somewhat collaborative.

**Principal Reports**
A lack of materials and other resources (e.g., instructional space, technologically competent staff) has a small effect on student achievement. Students of principals who reported that instruction at their school was somewhat affected by resource shortages scored, on average, lower in mathematics and science than students of principals who reported that instruction was not affected by resource shortages. Specific to science, principals also reported whether teachers have supplemental assistance when they conduct laboratory experiments and investigations. Students of teachers who had no supplemental assistance (67% of all students) scored, on average, higher than students whose teachers had assistance.

**Massachusetts in Comparison to Singapore**
Contextual factors can help to explain differences in student achievement (TIMSS, 2011a; TIMSS, 2011b). When comparing Massachusetts to the top-performing East Asian countries, some of the most interesting comparisons related to contextual factors come when Massachusetts is contrasted with Singapore. On the 2011 TIMSS assessments, Singaporean students had the highest average scaled score in science and the second highest in mathematics. Yet it is hard to pinpoint any contextual factors that might explain Singapore’s superior performance. For example, students and teachers in Massachusetts expressed more confidence in learning and teaching mathematics and science than students and teachers in Singapore. Similarly, students in Massachusetts reported more educational resources at home, on average, than their peers in Singapore. According to teacher surveys, Massachusetts schools place more emphasis on academic success when compared to schools in Singapore. The working conditions
of teachers in both countries are comparable as is the perception of how safe schools are. None of these factors would appear to help explain the difference in achievement between Massachusetts and Singapore.

When compared to Massachusetts, Singapore does have a much smaller percentage of students who attend schools whose student distribution has a high proportion of students from economically disadvantaged homes (26% in Massachusetts, 11% in Singapore). In both Singapore and Massachusetts, students in these types of schools perform substantially worse (large effect sizes) in mathematics and in science. Having a lower percentage of students in these types of schools may help explain Singapore’s superior performance on the TIMSS assessments. However, this evidence is at best anecdotal; without further, more sophisticated analyses of how all factors combine to impact achievement, it is hard to make any conclusive statements about the differences in performance between Massachusetts and Singapore and what factors account for these differences.
PART VII
REFERENCES


APPENDIX I

How TIMSS Selects Students and Reports Achievement Data

How TIMSS Selects Students for the Study
Education systems differ across the globe, with some students starting their schooling at age 4 and others starting at age 5 or 6. In order to participate in the grade 8 TIMSS assessments, students must be in a grade that represents 8 years of formal schooling, and the average age of students at the time of testing must be at least 13.5 years. In Massachusetts, this corresponds to grade 8. (Internationally, kindergarten is not considered a part of a student’s formal education.)

The cost and resources that would be required to assess all students in each participating country or benchmarking entity is too great. As a result, representative samples are assessed from each participant. A representative sample is a smaller number of students taken from the full target population who accurately reflect the characteristics (e.g., race, ethnicity, socio-economic status, geographic locality) of the student population within the participating entity. To ensure that a sample is representative of the target population, TIMSS first divides the target population into strata, and then randomly and proportionally chooses schools within the strata. Finally, TIMSS randomly chooses one or more classes of students to participate from within each selected school. A random sample ensures that any school or class of students has an equal probability of inclusion as any other school or class of students.

For example, for the U.S. sample, the eighth-grade population nationwide was first divided up by the type of school (public or private); region of the country (Northeast, Central, West, Southeast); community type (eight levels); and minority status (above or below 15 percent of the student population). The second stage involved randomly and proportionally selecting schools within the strata, and the final stage involved randomly picking one or more classrooms from within each school. All students from the selected classrooms were combined to form the sample.

For each country, TIMSS aims to have 150 schools with a minimum of 4,000 students participating in the sample. A smaller sample of 2,075 students represented Massachusetts.

How TIMSS Achievement Results Are Reported
TIMSS uses scaled scores to report student achievement. Why are scaled scores used? The mathematics and science content covered by TIMSS is vast. If a student was assessed on all possible content, the test would take several hours to complete. To reduce the testing burden on students, the subject content is divided into several test forms or booklets, with each booklet

---

1 See Part V for a discussion of how TIMSS contextual data are reported.
encompassing material that is representative of the subject content. Each test booklet contains
some items that are common to other booklets on the test, and these items are used to place
students’ scores onto one common achievement scale.

Every effort is made to ensure that the difficulty of the booklets is comparable; however, it is
extremely difficult to accomplish this task. As a result, in order to ensure that scores from each
test booklet can be interpreted consistently and with the same meaning, raw scores are
standardized onto one scale, and these scaled scores are used to compare results. Standardization
ensures that no matter which test booklet a student is given, the student’s score indicates the
same level of performance. This standardization process also allows student achievement to be
compared across content domains, across cognitive domains, and across time. TIMSS secures
items from public view and these items are used to place the five assessments (1995, 1999, etc.)
onto the same scale so trends in student achievement can also be assessed.

TIMSS Achievement Scaled Score Distribution
Student achievement in TIMSS is scored on a scale of 0 to 1,000; the centerpoint of the scale is
set to 500 with a standard deviation of 100. The average (mean) tells us where most of the scores
reside. A standard deviation is a measure of the spread of scores around the average. In general,
most scores (68%) are within one standard deviation of the average with 95% of scores within
two standard deviations (the other 5% of scores are farther away from the average score). In
TIMSS, most scores (68%) range from 400 to 600. When a country’s scores are tightly clustered
around the mean, the standard deviation is small and indicates little variability in the score
distribution. In contrast, a country with a large standard deviation has greater variability in its
students’ scores, with scores spread further out toward the extremes of the score distribution.

TIMSS Achievement Benchmarks
To help countries understand and interpret the meaning of the scaled scores, TIMSS categorizes
student performance into four performance levels or benchmarks: Advanced (above 625), High
(550), Intermediate (475), and Low (400). These performance benchmarks are qualitatively
described by the types of items and skills needed to achieve at each benchmark. Readers are
referred to http://timss.bc.edu/ for further information on how these benchmarks are set up. For
each country, TIMSS reports the percentage of students that fall into each performance
benchmark.