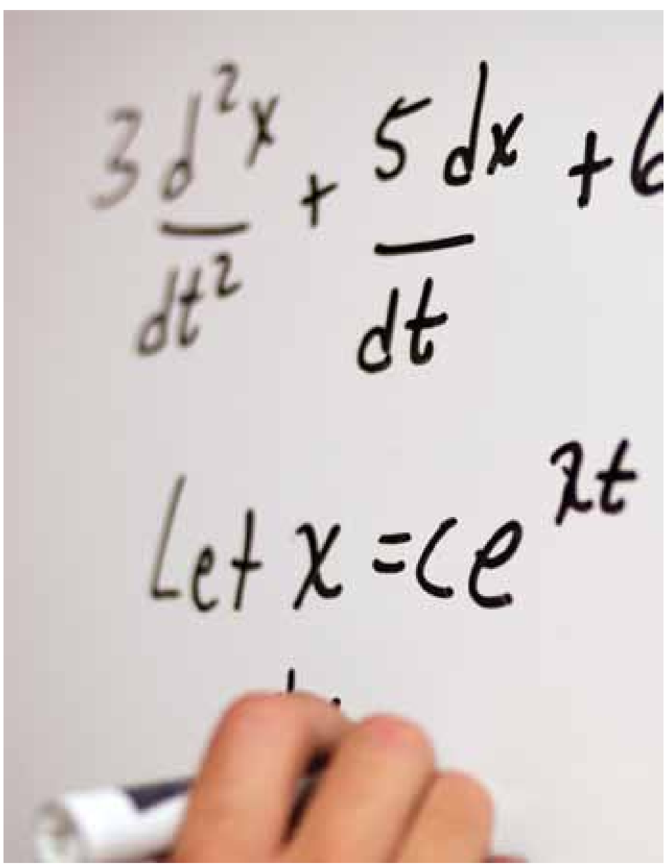
AMERICAN INSTITUTES FOR RESEARCH®

# The Second Derivative:
## *International Benchmarks in Mathematics For U.S. States and School Districts*

***Gary W. Phillips, Ph.D.***
*Vice President & Chief Scientist*
*AIR®*

*June, 2009*

*www.air.org*

## About This Report

The American Institutes for Research (AIR) has funded and conducted this report as part of our effort to make research relevant to policymakers and practitioners in the field of education. Our mission at AIR is to conduct and apply behavioral and social science research to improve people's lives and well-being, with a special emphasis on the disadvantaged. This report helps meet this goal by providing policymakers international benchmarks against which they can compare and monitor the educational performance of students.

In a highly interconnected world, U.S. students will require strong mathematic skills to compete against their peers around the globe. Reports such as The Second Derivative: International Benchmarks in Mathematics for U.S. States and School Districts help policymakers and educators to know how well they are doing in meeting this challenge and to track progress over time.

## About AIR®

Established in 1946, with headquarters in Washington, DC, and with nearly 30 offices in the United States and around the world, AIR is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of health, education, and workforce productivity.

# Contents

# List of Tables

# List of Figures



Copies of this paper can be downloaded by searching http://www.air.org, and questions can be addressed to the author at gwphillips@air.org. Proper citation is as follows: Phillips, G. W. (2009), The Second Derivative: International Benchmarks in Mathematics for U.S. States and School Districts. Washington, DC: American Institutes for Research.


# Executive Summary

*"In a global economy where the most valuable skill you can sell is your knowledge, a good education is no longer just a pathway to opportunity – it is a prerequisite. The countries that out-teach us today will out-compete us tomorrow." (Barack Obama, President of the United States, February 24, 2009)*

*"Fifty different goalposts is absolutely ridiculous. If we accomplish one thing in the coming years, it should be to eliminate the extreme variation in standards across America." (Arne Duncan, 9th United States Secretary of Education)*

The goal to have the best education in the world is captured in the U.S. Department of Education's initiative "*Race to the Top*." With a new administration in Washington there is a renewed enthusiasm among educators to *change* our educational ranking in the world. In calculus change is the *first derivative*. It is the ratio of how much we accomplish over how long it takes to accomplish it, or the *rise over the run*. This report helps with the numerator of this ratio by indicating what we are up against if we wish to be internationally competitive. The rise is formidable. The highest achieving countries are so far ahead of us we will never catch up if we run at the current pace. First derivative thinking will not solve this problem.

If we are to catch up with the best in the world and possibly get ahead of them we need to be thinking about the *second derivative*. We need to *accelerate* the pace of change. The best in the world do not stand still, do not wait, and do not give in. The best in the world are themselves struggling to stay ahead and are constantly trying to improve. If we continue to make imperceptible incremental improvements we will fall further behind or at best remain in the middle of the pack. In order to be the best in the world ourselves we need to understand what those ahead of us know, what they do, and why they do it. Then we need to implement radically new, groundbreaking policies that accelerate the pace at which our students learn in school.

This report[1] was developed out of an attempt to find a scientifically rigorous way to compare the mathematics performance of U.S. states and school districts against challenging international benchmarks. In order to compare ourselves with the best in the world in mathematics, this report provides a crosswalk between the data provided by the 2007 Trends in International Mathematics and Science Study (TIMSS) and the 2007 National Assessment of Educational Progress (NAEP). The report simplifies these comparisons by grading the countries, states, and school districts with a *comparable* grading system that is more familiar to policymakers, a grade of A, B, C, D, or BD (below a D). The report assumes that the international benchmark, against which we should calibrate our expectations and monitor our success, is a grade of B. The grade of B was chosen because this report shows it is statistically equivalent to the Proficient level on NAEP that has been recommended by the National Assessment Governing Board (NAGB) and No Child Left Behind (NCLB) as the level of performance we should expect from our students.

In ***benchmarking the United States*** against international achievement standards there were four overall findings.

> *First, the international average of mathematics learning in a broad cross-section of countries around the world is at a C level at Grade 4 and D+ at Grade 8. One disturbing finding was that there are a relatively large number of countries in which the students are performing below a D level of proficiency.*

> *Second, the U.S. average is at the C+ level of mathematics proficiency at Grade 4 and C at Grade 8. This suggests that the typical student in the United States is learning basic mathematics rather than the more complex mathematics required to meet global expectations.*

> *Third, the average of the participating Organization for Economic Cooperation and Development (OECD) countries is also at the C+ level for Grade 4 and a C level at Grade 8. The U.S. average is not significantly different from the OECD average.*

> *Fourth, a group of Asian countries consistently perform at the B+, B, and B- level and are learning mathematics, not just at a higher level than the United States, but at a quantum leap higher level than the United States. These are Chinese Taipei, South Korea, Singapore, Hong Kong SAR, and Japan. Clearly, these are some of the countries within the TIMSS that represent best practice and against which the United States should be making comparisons.*

In ***benchmarking the states*** against the same international achievement standards, there were several findings.

> *First, within the United States almost all states are performing at the C+ and C level, which is below the international benchmark recommended in this report. This is true for both Grade 4 and Grade 8.*

> *Second, there is a small set of states in which students are learning at the B or B- level. At Grade 4 these are Massachusetts, Minnesota, New Jersey, New Hampshire, Kansas, and Vermont. At Grade 8 this includes only Massachusetts.*

> *Third, there is a general tendency among the states to drop in performance from Grade 4 to Grade 8. Why states are less able to keep up in Grade 8 is an important mathematics education question that should be addressed by policymakers. This drop in performance is not seen among the high-achieving Asian countries.*

In ***benchmarking the school districts*** also against the same international achievement standards, there were two major findings.

[1] This report makes frequent references to terms such as international benchmarks, performance standards, achievement standards, and achievement levels. These terms reflect differences in terminology used by various groups. However, in each case these terms refer to cut-scores on an assessment. These standards reflect how much we expect the student to learn. They are related to, but not the same thing as, curriculum standards (what the student should learn) or content standards (what should be tested). Many U.S. educators are also interested in international benchmarks for curriculum and content standards in an effort to import the "best practices" from around the world for the benefit of American students. Benchmarking content and curriculum are important activities, but these activities are not the subject of this report, which strictly deals with benchmarking performance standards.

*First, most of the school districts that participated in the 2007 National Assessment of Educational Progress (NAEP) are performing at the C+ and C level, which is below the international benchmark recommended in this report. This is true for both Grade 4 and Grade 8. One difference between the states and the school districts is that none of the school districts are performing at the B level.*

*Second, there is also a general tendency among the districts to drop in performance from Grade 4 to Grade 8. For example, at Grade 4 only one district was performing at the D+ level – the District of Columbia. By Grade 8, five districts had fallen to the D+ level. These were Los Angeles, Chicago, Atlanta, the District of Columbia, and Cleveland.*

# Why International Benchmarking?

The race to the top starts with knowing where we stand and how high the bar is over which we need to jump. We also need to be able to monitor how fast we are running. Over the past several decades there has been a gradual evolution in our thinking about how to monitor student learning. This evolution is reflected in major changes in the National Assessment of Educational Progress (NAEP), which serves as an overall indicator of our country's educational success. At first we were happy just seeing if we were making progress. This was essentially comparing our performance today with our performance in the past. Then we decided we need to also compare ourselves to each other. This led to the State-NAEP, in which all 50 states are compared to each other. Along with this we established national standards for mathematical proficiency and used that as the national bar over which we would like to see our students jump. Under No Child Left Behind (NCLB) legislation each state has also established its own internal state bar which is usually not comparable to the national bar. In recent years globalization has reminded us that being better today than we were yesterday, or being better than our neighbor in the next state, is not good enough. Furthermore, establishing state or national bars uninformed by what is happening around the world is flying without radar.

*Merriam-Webster* defines a "benchmark" as "something that serves as a standard by which others may be measured or judged." International benchmarking for American education centers on identifying and collecting data about best practice from the most successful education systems around the world and then using that information to improve our own practices.

Currently there is considerable interest in international benchmarking. This interest has been generated by a steady stream of international data that show the United States performing below where we should be to remain internationally competitive. Also, the new administration in Washington, D.C., has made frequent references to the need for international benchmarking of state achievement standards. No one believes international benchmarking is a silver bullet that will solve all the problems with American education. But certainly it should be at the front of the list of strategies for making improvements. A recent report by the National Governors Association, the Council of Chief State School Officers, and Achieve (Benchmarking for Success, 2008) concludes that

> *"Governors recognize that new economic realities mean it no longer matters how one U.S. state compares to another on a national test; what matters is how a state's students compare to those in countries around the globe. America must seize this moment to ensure that we have workers whose knowledge, skills, and talents are competitive with the best in the world." (Page 1)*

The report concludes by stating that states must lead and look beyond their borders to

> *"fully understand how to benchmark expectations for student learning. They must significantly broaden the policy lens by drawing lessons from the highest performing, most equitable, and fastest advancing nations and states around the globe and adapting the very best educational practices to incorporate here at home...And state leaders have both the authority and an obligation to ensure that students attend globally competitive schools and school districts. America cannot maintain its place in the world—economically, socially, or culturally—unless all of its students gain the skills that allow them to compete on a global scale. The United States will only achieve true international competitiveness when state education policies and institutions are restructured to meet 21st century realities." (Page 39)*

Another recent report (From Competing to Leading, 2008) states

> *"The United States once enjoyed the position of global leader in education and now is struggling to compete. In measuring progress, most states compare themselves to other states rather than to international benchmarks. Because of the nation's diminished international standing, continuing to engage in interstate comparisons risks perpetuating regionally low standards and achievement, and ignores the necessity to adequately prepare a workforce that is mobile across both state and national boundaries. To move from competing to leading, states should spend less time comparing to one another and spend more time comparing to high-performing countries." (Page 5)*

This report developed out of an attempt to find a scientifically rigorous way to compare the mathematics performance of U.S. states and school districts against challenging international benchmarks. In order to compare ourselves with the best in the world in mathematics, this report provides a crosswalk between the data provided by the 2007 Trends in International Mathematics and Science Study (TIMSS) and the 2007 National Assessment of Educational Progress (NAEP). The report simplifies these comparisons by grading the countries, states, and school districts with a *comparable* grading system that policymakers are more likely to understand, a grade of A, B, C, D, or BD (below D). The report assumes that the international benchmark, against which we should calibrate our expectations and monitor our success, is a grade of B (a fuller discussion can be found in the section on Creating an International Grading Index).

# Measuring State and School District Achievement

The measurement of state and school district achievement used in this document is based on a re-analysis of the data provided in the public report of the 2007 National Assessment of Educational Progress (NAEP). The NAEP is a congressionally authorized assessment of all 50 states and several territories. The assessment is carried out by the National Center for Education Statistics (NCES) with policy oversight by the independent National Assessment Governing Board (NAGB). Because of the persistent requests of urban school districts, the U.S. Congress authorized NAEP to assess, on a trial basis, six large urban school districts beginning in 2002. Since then, more districts have been added, resulting in 11 school districts in 2007 (and plans are under way to include even more districts in the future).

The challenging standards associated with NAEP provide a national context and an external compass with which states and school districts can steer educational policy to benefit their local systems. However, the superintendents in the 50 states and the urban school chiefs in these 11 large school districts also recognize the global nature of economic competition and acknowledge the importance of international educational expectations. They need reliable external data against which to benchmark the performance of their students.

# Options for International Benchmarking

Currently two international surveys collect international data in mathematics that could provide the data needed for international benchmarking. These are

*Program for International Student Assessment (PISA)*

*Trends in International Mathematics and Science Study (TIMSS)*

PISA is an assessment of 15-year-old students sponsored by the Organization for Economic Cooperation and Development (OECD) located in Paris. TIMSS is an assessment of Grade 4 and Grade 8 students sponsored by the International Association for the Evaluation of Educational Achievement (IEA), which is currently located in the Netherlands.

## Using PISA for International Benchmarking

One of the main advantages of using PISA for international benchmarking is that the assessment includes the 30 countries that make up the OECD (the survey also includes about an equal number of non-OECD countries). These OECD countries are some of the most advanced economies in the world and therefore America's most important trading partners and competitors. The TIMSS assessment includes a broader range of countries (including many developing countries) and about half of the OECD countries.

PISA is a literacy assessment and not a curriculum-based assessment. PISA measures how well students apply mathematics to real-world situations. It measures the cumulative "yield" of the student's total lifelong educational experience. In addition to covering what is learned in school, it also reflects what students learn from families, society, and popular culture.

In PISA, students are sampled by age (i.e., age 15) as opposed to being grade-based. In the United States most of the students in the sample are in Grade 10 (about 70%), but 2% are in Grades 7 and 8, 11% are in Grade 9, and 17% are in Grade 11.

## Using TIMSS for International Benchmarking

TIMSS is a curriculum-based assessment. It is intended to measure, internationally, the mathematics that should be learned in school. Although TIMSS collects data about societal goals, student attitudes, and values, the assessment does not commingle these literacy constructs with its definition of mathematical proficiency. In addition, TIMSS uses grade-based sampling (Grade 4 and Grade 8) so the sample is more relevant to how the American educational system is structured (i.e., by grade).

However, another factor related to TIMSS represents the reason it was selected for international benchmarking in this report. TIMSS can be statistically linked to NAEP, which measures essentially the same mathematical content as TIMSS (see Phillips, 2007, and Phillips and Dossey, 2008). Since all U.S. states take NAEP (as well as a few school districts), the statistical linkage allows each state to see how it would perform if the state-NAEP results were reported on the TIMSS scale. This statistical linkage was recently possible because NAEP and TIMSS were administered to equivalent representative samples of students, in the same subject (mathematics), in the same year (spring of 2007), and in the same grades (Grades 4 and 8). No similar basis exists to create a linkage between PISA and NAEP. That is because PISA is administered to age-15 students (which are primarily in the 10th grade in the United States), whereas NAEP is administered in the 8th grade.

The two primary ways the TIMSS reports its results are in terms of averages (or means) and achievement standards (referred to as *international benchmarks*). The labels and cut-points on the TIMSS scale for the international benchmarks are

*Advanced (625)*
*High (550)*
*Intermediate (475)*
*Low (400)*

These achievement standards apply to both the Grade 4 as well as the Grade 8 mathematics assessment. The achievement standards were initially established in the first TIMSS population in 1995, where they represented the 90th percentile (Advanced), the 75th percentile (High), the 50th percentile (Intermediate), and the 25th percentile (Low). The substantive content definitions of each of these international benchmarks are provided in the 2007 TIMSS report (see Mullis, Martin, & Foy, 2008, pages 68–69).

### TIMSS International Benchmarks for Grade 4 Mathematics

***Grade 4 Advanced:*** *Students can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning. They can apply proportional reasoning in a variety of contexts. They demonstrate a developing understanding of fractions and decimals. They can select appropriate information to solve multistep word problems. They can formulate or select a rule for a relationship. Students can apply geometric knowledge of a range of two- and three-dimensional shapes in a variety of situations. They can organize, interpret, and represent data to solve problems.*

***Grade 4 High:*** *Students can apply their knowledge and understanding to solve problems. Students can solve multistep word problems involving operations with whole numbers. They can use division in a variety of problem situations. They demonstrate understanding of place value and simple fractions. Students can extend patterns to find a later specified term and identify the relationship between ordered pairs. Students show some basic geometric knowledge. They can interpret and use data in tables and graphs to solve problems.*

***Grade 4 Intermediate:** Students can apply basic mathematical knowledge in straightforward situations. Students at this level demonstrate an understanding of whole numbers. They can extend simple numeric and geometric patterns. They are familiar with a range of two-dimensional shapes. They can read and interpret different representations of the same data.*

***Grade 4 Low:** Students have some basic mathematical knowledge. Students demonstrate an understanding of adding and subtracting with whole numbers. They demonstrate familiarity with triangles and informal coordinate systems. They can read information from simple bar graphs and tables.*

### TIMSS International Benchmarks for Grade 8 Mathematics

***Grade 8 Advanced:** Students can organize and draw conclusions from information, make generalizations, and solve nonroutine problems. They can solve a variety of ratio, proportion, and percent problems. They can apply their knowledge of numeric and algebraic concepts and relationships. Students can express generalizations algebraically and model situations. They can apply their knowledge of geometry in complex problem situations. Students can derive and use data from several sources to solve multistep problems.*

***Grade 8 High:** Students can apply their understanding and knowledge in a variety of relatively complex situations. They can relate and compute with fractions, decimals, and percents, operate with negative integers, and solve word problems involving proportions. Students can work with algebraic expressions and linear equations. Students use knowledge of geometric properties to solve problems, including area, volume, and angles. They can interpret data in a variety of graphs and tables and solve simple problems involving probability.*

***Grade 8 Intermediate:** Students can apply basic mathematical knowledge in straightforward situations. They can add and multiply to solve one-step word problems involving whole numbers and decimals. They can work with familiar fractions. They understand simple algebraic relationships. They demonstrate understanding of properties of triangles and basic geometric concepts. They can read and interpret graphs and tables. They recognize basic notions of likelihood.*

***Grade 8 Low:** Students have some knowledge of whole numbers and decimals, operations, and basic graphs.*

In 2007 the National Center for Education (NCES) released a report in which the state achievement standards on each state achievement test were mapped onto the NAEP scale (Mapping 2005 State Proficiency Standards Onto the NAEP Scales, 2007). This resulted in what was referred to as NAEP-equivalent achievement standards (the score on the NAEP scale that was equivalent to the state achievement standard). This result was obtained through statistically linking each state achievement test to the NAEP scale. The current report follows a similar strategy for statistical linking. However, in this report we are finding the TIMSS-equivalent score that is associated with the state-NAEP mean. This is the crosswalk that allows us to determine how the mean on the state-NAEP compares with the international benchmarks provided by TIMSS.[2]

[2] See the Technical Appendix for a description of the statistical linking study.

# Creating an International Grading Index[3]

This report uses these international benchmarks to conduct a secondary analysis of TIMSS. However, in order to facilitate communication with an American audience, the report employs the cut-scores on the TIMSS scale to create a new index using a metric more familiar to American educators and policymakers. For each country in TIMSS the country mean will be compared to the international benchmark. If the country average is at the Advanced level, the country is assigned an international grade of A. Similarly, if the country average is at the High level, it is assigned a B, for Intermediate it will be given a C, for Low it will be given a D, and if the country average is below Low it will given a BD (below a D).

Since the country averages are based on samples of students, the country average will have an associated margin of error. The margin of error will be incorporated into the international grade index by assigning a minus grade if the next highest international benchmark is within the 95% confidence interval of the average. For example, a country whose average is lower (but not significantly below) the Intermediate benchmark would be assigned a C- instead of a D. This gives the country the benefit of the doubt associated with the margin of error of the mean.

A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. For example, a C+ would occur when the country mean is more than halfway between the Intermediate and High international benchmarks on the TIMSS scale.

This report assumes that an international grade of B is the U.S. target and therefore the international benchmark against which we should make comparisons. An international grade of B was selected because the linking study underlying this report showed that a Proficient level on NAEP was statistically comparable to the High international benchmark (or an international grade of B) on TIMSS (see Table 7 and Table 8). The Proficient level has been recommended by NAGB and endorsed by NCLB as the target for mathematical performance.

[3] More details on the basis of the international grades can be found in the Technical Appendix.

There are several reasons why this international grading system is a good choice for comparing educational outcomes within the context of a global educational environment.

- The grading system is a familiar metric and is intuitively understandable to the public and policymakers.
- The grades are connected to rigorous international benchmarks (i.e., cut-scores on the TIMSS scale). This is indicated by the fact that only a few high-achieving countries and states received a B grade and no country or state received an A.
- The international benchmarks that underlie the grades were established through an international consensus process and have a scientifically based criterion-referenced interpretation (Olson, Martin, & Mullis, chapter 13).
- The grading system is comparable across Grades 4 and 8, across years of administration, across countries, and (because of the linking study) across states and school districts.

# International Grades in Mathematics for Countries

The international grades for each country are presented in Table 1 and Table 2 and Figure 1 and Figure 2. Figure 1, for example, shows the TIMSS performance with the country international grade. Figure 2 indicates how far above or below the international benchmark each country is performing using a B as the expected performance. The international grades in each country reveal several important findings that are important to American education policymakers.

The *first* is that the international average of mathematics learning in a broad cross-section of countries around the world is at a C level in Grade 4 and a D+ at Grade 8.[4] These averages are based on 36 countries at Grade 4 and 48 countries at Grade 8. These graphs show that there is a wide variation in mathematical learning around the world. A few countries do a good job of teaching mathematics to the overall population of students, but in many countries the average student is not learning much mathematics. If the United States is excluded and we calculate the international average for countries outside the United States, we find the international average is a D+ for Grade 4 and 8. This means the average student outside the United States has only a rudimentary understanding of basic mathematics.

The *second* important finding is that the U.S. average is at the C+ level of mathematics proficiency at Grade 4 and C at Grade 8. This suggests that the typical student performance in the United States is learning mathematics below the international benchmark of B and at a partially proficient level when compared against global expectations. The typical student within the United States has an intermediate level of mathematical understanding and applies *basic mathematics* rather than more complex mathematics to solve problems.

[4] The international averages in this report are weighted by the student population size and are indicated by the grey bars in the graph. The international averages include the United States, which makes up about a quarter of the student population among the countries in the study. Finally, it should be noted that the international average in this report is different from the international average in the 2007 TIMSS report (Mullis, Martin, & Foy, 2008), which is equal to 500. That is because the average reported in 2007 TIMSS is a "scale" average, not an international average. In order to report trends, the TIMSS survey established the international average of 500 in the 1995 assessment as the reference point. They then report whether the countries are improving compared to this start point. The current report uses the 2007 international average, not the scaled reference point.

The *third* finding is that the average of the participating OECD countries is also at the C+ level of performance at Grade 4 and C at Grade 8. The international OECD averages are based on 16 OECD countries at Grade 4 and 12 OECD countries at Grade 8. When the U.S. average is compared to the OECD average countries, we find that the U.S. average is not significantly different.[5] In Figure 1 and Figure 2 the dark bars of the chart represent countries that are significantly above or below the OECD international average (which is indicated by the grey bar). The white bars in the chart represent countries that are not significantly different from the OECD international average.

The *fourth* important finding is that a group of Asian countries consistently perform at the B+, B, *and* B- levels and are learning mathematics, not just at a higher level than the United States, but at a level that is a quantum leap higher than the United States. These are Chinese Taipei, South Korea, Singapore, Hong Kong SAR, and Japan. The typical student within these countries has a higher level of mathematical understanding and uses *more complex mathematics* to solve problems. To get a feel for how far ahead these countries are, we convert the difference between the United States and the highest achieving country to effect-sizes (which are standard deviation units) and compare them to the difference between the lowest performing state and the highest performing state. For Grade 4 the highest achieving country (Hong Kong) is about one standard deviation ahead of the United States, which is comparable to the difference between the highest achieving state (Massachusetts) and the lowest achieving state (Mississippi). The same difference is observed at Grade 8 as well. At Grade 8 Chinese Taipei is about one standard deviation above the United States, and Massachusetts is about one standard deviation above Mississippi. One standard deviation represents a *huge* achievement gap. Clearly, these are some of the countries in TIMSS that represent best practice and against which the United States should be making comparisons.

[5] As mentioned above the TIMSS assessment does not include all 30 OECD countries. However, in the PISA assessment (which does include all OECD countries) the U.S. scores below the OECD average. This would probably be the case in TIMSS as well if all OECD countries were included in the assessment.

**Table 1: International Grades for Countries in 2007 Mathematics, Grade 4**[6]

| Country | TIMSS Mean | Grade |
|---|---|---|
| Hong Kong SAR | 607 | B+ |
| Singapore | 599 | B+ |
| Chinese Taipei | 576 | B |
| Japan | 568 | B |
| Kazakhstan | 549 | B- |
| Russian Federation | 544 | B- |
| England | 541 | C+ |
| Latvia | 537 | C+ |
| Netherlands | 535 | C+ |
| **OECD Mean** | **531** | **C+** |
| Lithuania | 530 | C+ |
| **United States** | **529** | **C+** |
| Germany | 525 | C+ |
| Denmark | 523 | C+ |
| Australia | 516 | C+ |
| Hungary | 510 | C |
| Italy | 507 | C |
| Austria | 505 | C |
| Sweden | 503 | C |
| Slovenia | 502 | C |
| Armenia | 500 | C |
| Slovak Republic | 496 | C |
| Scotland | 494 | C |
| New Zealand | 492 | C |
| Czech Republic | 486 | C |
| **International Mean** | **482** | **C** |
| Norway | 473 | C- |
| Ukraine | 469 | D+ |
| Georgia | 438 | D+ |
| Iran, Islamic Rep. of | 402 | D |
| Algeria | 378 | BD |
| Colombia | 355 | BD |
| Morocco | 341 | BD |
| El Salvador | 330 | BD |
| Tunisia | 327 | BD |
| Kuwait | 316 | BD |
| Qatar | 296 | BD |
| Yemen | 224 | BD |

[6] Note: The above table reports on the TIMSS international benchmark level of the typical student in the country (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next highest achievement level is within the 95% confidence interval of the country average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). Source of data: Mullis, Martin, & Foy, 2008.

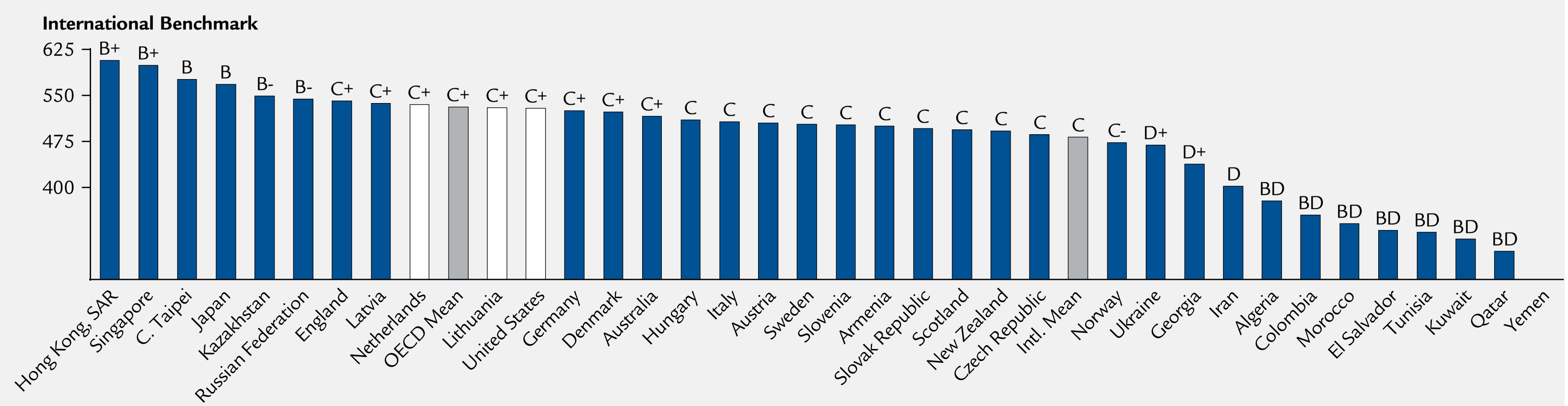
**Table 2: International Grades for Countries in 2007 Mathematics, Grade 8[7]**

| Country | TIMSS Mean | Grade |
|---|---|---|
| Chinese Taipei | 598 | B+ |
| Korea, Rep. of | 597 | B+ |
| Singapore | 593 | B+ |
| Hong Kong SAR | 572 | B |
| Japan | 570 | B |
| Hungary | 517 | C+ |
| England | 513 | C+ |
| Russian Federation | 512 | C |
| **OECD Mean** | **511** | **C** |
| **United States** | **508** | **C** |
| Lithuania | 506 | C |
| Czech Republic | 504 | C |
| Slovenia | 501 | C |
| Armenia | 499 | C |
| Australia | 496 | C |
| Sweden | 491 | C |
| Malta | 488 | C |
| Scotland | 487 | C |
| Serbia | 486 | C |
| Italy | 480 | C |
| Malaysia | 474 | C- |
| Norway | 469 | D+ |
| Cyprus | 465 | D+ |
| Bulgaria | 464 | D+ |
| Israel | 463 | D+ |
| Ukraine | 462 | D+ |
| Romania | 461 | D+ |
| **International Mean** | **461** | **D+** |
| Bosnia-Herzegovina | 456 | D+ |
| Lebanon | 449 | D+ |
| Thailand | 441 | D+ |
| Turkey | 432 | D |
| Jordan | 427 | D |
| Tunisia | 420 | D |
| Georgia | 410 | D |
| Iran, Islamic Rep. of | 403 | D |
| Bahrain | 398 | D- |
| Indonesia | 397 | D- |
| Syrian Arab Republic | 395 | D- |
| Egypt | 391 | BD |
| Algeria | 387 | BD |
| Colombia | 380 | BD |
| Oman | 372 | BD |
| Palestinian Nat'l Auth. | 367 | BD |
| Botswana | 364 | BD |
| Kuwait | 354 | BD |
| El Salvador | 340 | BD |
| Saudi Arabia | 329 | BD |
| Ghana | 309 | BD |
| Qatar | 307 | BD |

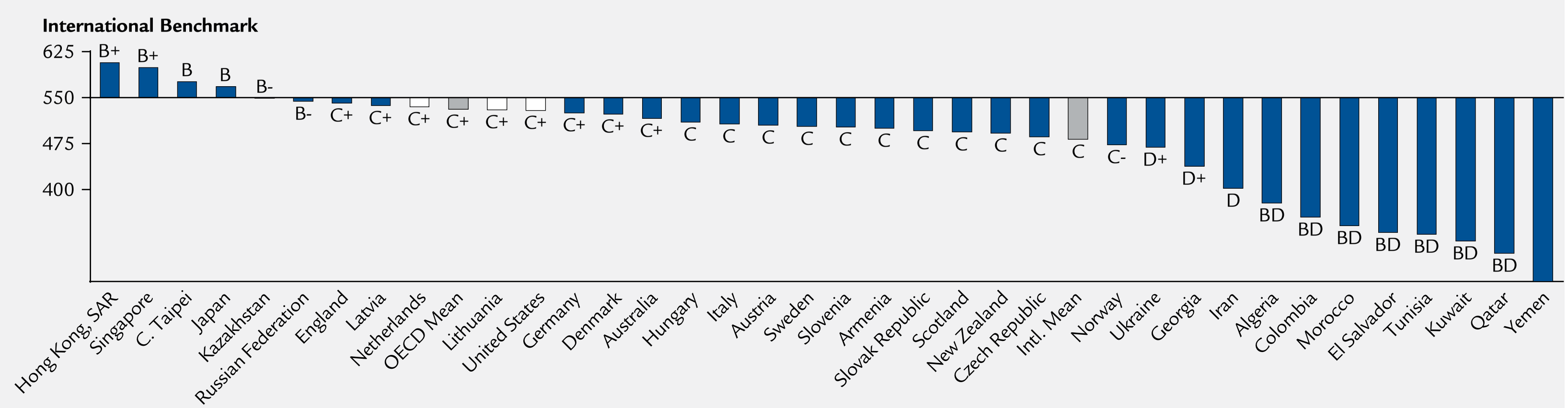
[7] Note: The above table reports on the TIMSS international benchmark level of the typical student in the country (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next highest achievement level is within the 95% confidence interval of the country average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). Source of data: Mullis, Martin, & Foy, 2008.

## Figure 1: International Grades for Countries, Mathematics, Grade 4

### International Grade Associated With the TIMSS Mean Achievement in Each Country, 2007 Mathematics, Grade 4



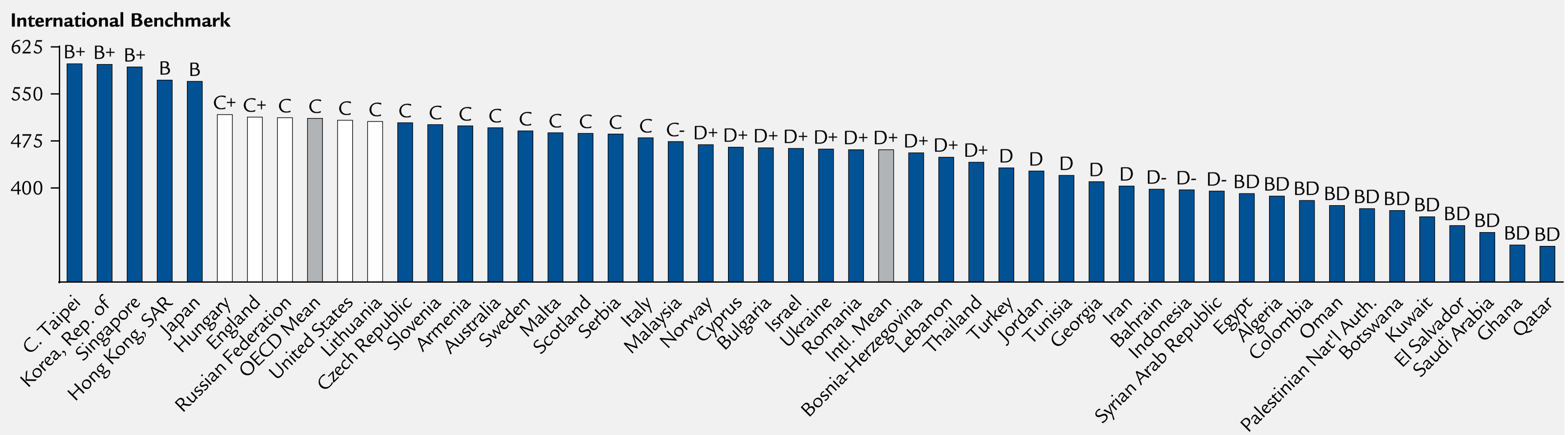

### TIMSS Mean Achievement in Each Country Benchmarked Against an International Grade of "B," 2007 Mathematics, Grade 4



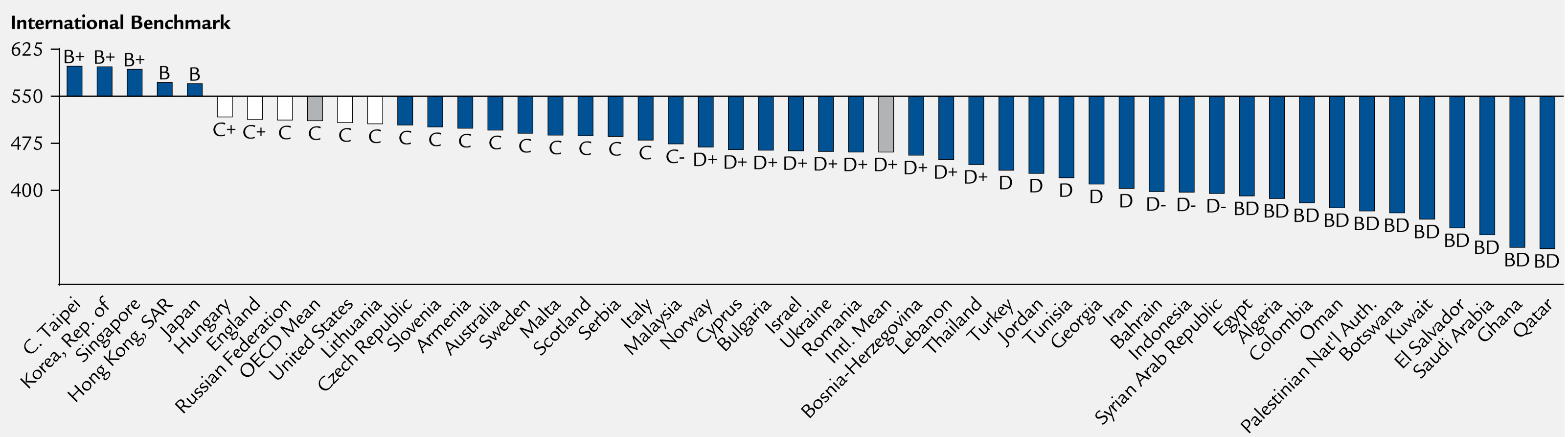

Source of data: Mullis, Martin, & Foy (2008), TIMSS 2007 International Mathematics Report. Grades determined by Phillips (2009).

## Figure 2: International Grades for Countries, Mathematics, Grade 8

**International Grade Associated With the TIMSS Mean Achievement in Each Country, 2007 Mathematics, Grade 8**




**TIMSS Mean Achievement in Each Country Benchmarked Against an International Grade of "B," 2007 Mathematics, Grade 8**




Source of data: Mullis, Martin, & Foy (2008), TIMSS 2007 International Mathematics Report. Grades determined by Phillips (2009).




# International Grades in Mathematics for States

The main purpose of this report is to provide states and school districts in the United States a way to benchmark their performance against international standards. There are several approaches that might be followed.

The *first* approach might be for states and school districts to administer PISA or TIMSS to a representative sample of students within the state or school district. This would require an expenditure of funds from the state or school district to pay for the data collection of PISA or TIMSS. This would also impose an additional testing burden on students and schools, many of which are already being tested on the state criterion-referenced test and possibly a norm-referenced test, as well as meeting NAEP and other testing requirements. This approach is exactly what was done in the states of Massachusetts and Minnesota, which participated in the 2007 TIMSS. Although state and school district participation in PISA or TIMSS might be the best approach, it is not currently affordable or practical for all 50 states and is even more unrealistic for school districts.

There is a *second* approach that is almost as good as the first one but has the benefit of requiring no additional funding and placing no extra testing burden on students and schools. The second approach is statistical linking. Because NAEP and TIMSS cover essentially the same content, are administered to equivalent national samples within the United States, are occasionally administered within the same year (2007), and are administered in the same grades (Grades 4 and 8), the two assessments can be statistically linked. After the linking is complete we can then estimate how students within the United States would perform on TIMSS based on how they performed on NAEP. Since each state participates in state-NAEP, we can estimate how each state would perform on TIMSS. This is the crosswalk that gives states their international benchmarks. We can also validate these estimates by using the states of Massachusetts and Minnesota (states in which TIMSS was *actually* administered), where we can compare our state estimates to actual performance. The entire process of statistical linking and validation is provided in the Technical Appendix.

Table 3 and Table 4 and Figure 3 and Figure 4 display the results of the NAEP-TIMSS linking for states. The TIMSS average for each state is simply the NAEP average re-expressed in the metric of TIMSS. It is important to understand that the statistical linking conducted in this study does not "predict" a state-TIMSS score from a state-NAEP score. Therefore, we are not predicting scores on a test that was not administered in the state. Instead, the study converts the NAEP scores that were obtained in the state-NAEP to the TIMSS scale, which then allows us to see if the average state-NAEP performance reached the international benchmarks in TIMSS. In the tables below the entries for Massachusetts, Minnesota, and the United States are actual rather than estimated TIMSS results. The international grades for each state indicate several important findings.

The *first* finding is that an overwhelming majority of states in the United States are performing at the C+ and C level, which represents a level of mathematics learning that is below the international benchmark of B. Like the nation as a whole, the mathematics achievement of most states is at a relatively basic level rather than the more difficult and complex mathematics level associated with the international grade of B. This is true for both Grade 4 and Grade 8. In Figure 3 and Figure 4 the dark bars of the chart

**Table 3: International Grades for States in 2007 Mathematics, Grade 4**[8]

| State | Estimated TIMSS Mean | International Grade |
|---|---|---|
| Massachusetts | 572 | B |
| Minnesota | 554 | B |
| New Jersey | 552 | B |
| New Hampshire | 552 | B |
| Kansas | 551 | B |
| Vermont | 546 | B- |
| North Dakota | 544 | C+ |
| Indiana | 543 | C+ |
| Ohio | 542 | C+ |
| Wisconsin | 541 | C+ |
| Pennsylvania | 540 | C+ |
| Wyoming | 540 | C+ |
| Montana | 539 | C+ |
| Virginia | 539 | C+ |
| Iowa | 537 | C+ |
| Connecticut | 537 | C+ |
| New York | 536 | C+ |
| Washington | 536 | C+ |
| Maine | 536 | C+ |
| Texas | 536 | C+ |
| Florida | 535 | C+ |
| Delaware | 534 | C+ |
| North Carolina | 534 | C+ |
| South Dakota | 533 | C+ |
| Idaho | 532 | C+ |
| **OECD Mean** | **531** | **C+** |
| Maryland | 531 | C+ |
| Colorado | 530 | C+ |
| DoDEA | 530 | C+ |
| **United States** | **529** | **C+** |

represent states that are significantly above or below the OECD international average (which is indicated by the grey bar). The white bars in the chart represent states that are not significantly different from the OECD international average.

The *second* major pattern in the data is that there is a set of states in which students are learning at the B or B- level. At Grade 4 these are Massachusetts, Minnesota, New Jersey, New Hampshire, Kansas, and Vermont. At Grade 8 this includes only Massachusetts. The fact that several states have reached this level of mathematical proficiency indicates that it is possible in the United States for students to learn mathematics at a level that is competitive with the best in the world.

The *third* pattern is that there is a general tendency among the states to drop in performance from Grade 4 to Grade 8. This across-the-board drop in performance is dramatically illustrated by comparing the graphs in Figure 3 and Figure 4. In Grade 8 we see that almost every state has dropped and is further below the international benchmark of B than they were in Grade 4. Why states fall behind in Grade 8 is an important mathematics education question that should be addressed by policymakers. This drop in performance is not seen among the high-achieving Asian countries.

**Table 3: International Grades for States in 2007 Mathematics, Grade 4[8]—Continued**

| State | Estimated TIMSS Mean | International Grade |
|---|---|---|
| Missouri | 528 | C+ |
| Utah | 528 | C+ |
| Nebraska | 525 | C+ |
| Arkansas | 524 | C+ |
| Michigan | 523 | C+ |
| Illinois | 523 | C+ |
| Alaska | 523 | C+ |
| South Carolina | 522 | C+ |
| Oklahoma | 521 | C+ |
| West Virginia | 520 | C+ |
| Oregon | 519 | C+ |
| Rhode Island | 519 | C+ |
| Georgia | 517 | C+ |
| Kentucky | 517 | C+ |
| Hawaii | 515 | C+ |
| Tennessee | 511 | C |
| Arizona | 509 | C |
| Nevada | 508 | C |
| Louisiana | 504 | C |
| California | 504 | C |
| Alabama | 500 | C |
| New Mexico | 498 | C |
| Mississippi | 497 | C |
| **International Mean** | **482** | **C** |
| Washington, DC | 461 | D+ |

[8] Note: The above table reports on the TIMSS international benchmark level of the typical student in the state (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next highest achievement level is within the 95% confidence interval of the state average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). DoDEA = Department of Defense Education Activity. Source of data: Lee, Grigg, & Dion, 2007.

**Table 4: International Grades for States in 2007 Mathematics, Grade 8**[9]

| State | Estimated TIMSS Mean | International Grade |
|---|---|---|
| Massachusetts | 547 | B- |
| Minnesota | 532 | C+ |
| North Dakota | 530 | C+ |
| Vermont | 529 | C+ |
| Kansas | 527 | C+ |
| New Jersey | 524 | C+ |
| South Dakota | 524 | C+ |
| Virginia | 522 | C+ |
| New Hampshire | 522 | C+ |
| Montana | 521 | C+ |
| Wyoming | 520 | C+ |
| Maine | 519 | C+ |
| Colorado | 519 | C+ |
| Pennsylvania | 519 | C+ |
| Texas | 518 | C+ |
| Maryland | 518 | C+ |
| Wisconsin | 518 | C+ |
| Iowa | 517 | C+ |
| DoDEA | 516 | C+ |
| Indiana | 516 | C+ |
| Washington | 516 | C+ |
| Ohio | 516 | C+ |
| North Carolina | 514 | C+ |
| Oregon | 514 | C+ |
| Nebraska | 513 | C+ |
| Idaho | 513 | C+ |
| Delaware | 512 | C |
| Alaska | 511 | C |
| **OECD Mean** | **511** | **C** |
| Connecticut | 511 | C |

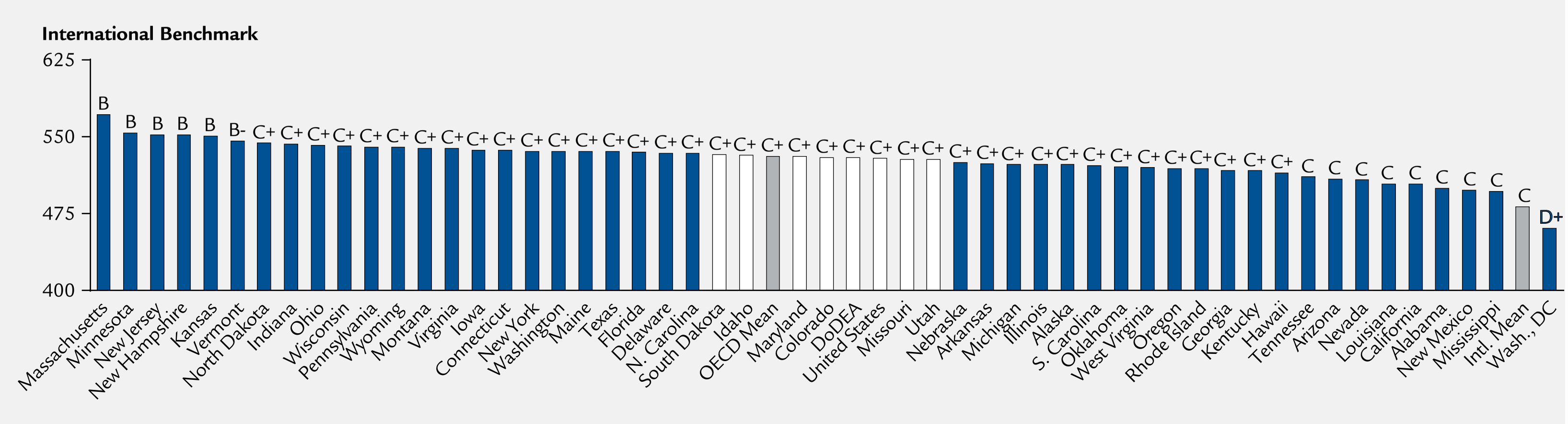
**Table 4: International Grades for States in 2007 Mathematics, Grade 8[9]—Continued**

| State | Estimated TIMSS Mean | International Grade |
|---|---|---|
| South Carolina | 509 | C |
| **United States** | **508** | **C** |
| Utah | 508 | C |
| Missouri | 507 | C |
| Illinois | 507 | C |
| New York | 506 | C |
| Kentucky | 503 | C |
| Florida | 500 | C |
| Michigan | 499 | C |
| Arizona | 496 | C |
| Rhode Island | 496 | C |
| Georgia | 494 | C |
| Oklahoma | 494 | C |
| Tennessee | 493 | C |
| Arkansas | 493 | C |
| Louisiana | 489 | C |
| Nevada | 486 | C |
| California | 485 | C |
| West Virginia | 484 | C |
| Hawaii | 482 | C |
| New Mexico | 479 | C |
| Alabama | 476 | C |
| Mississippi | 473 | C- |
| **International Mean** | **461** | **D+** |
| Washington, DC | 438 | D+ |

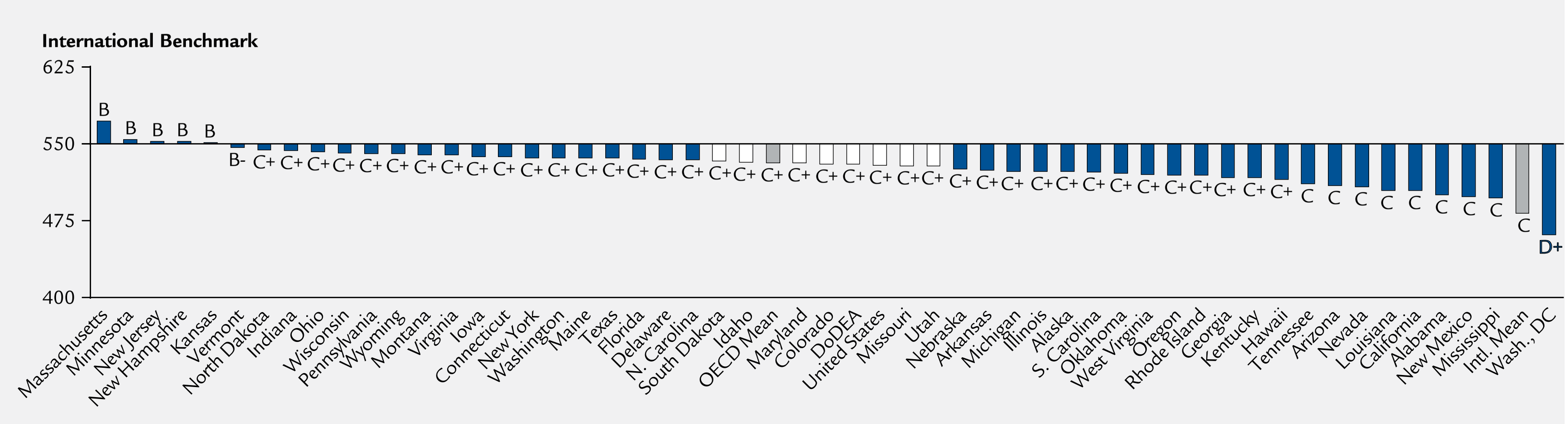
[9] Note: The above table reports on the TIMSS international benchmark level of the typical student in the state (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next highest achievement level is within the 95% confidence interval of the state average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). DoDEA = Department of Defense Education Activity. Source of data: Lee, Grigg, & Dion, 2007.

## Figure 3: International Grades for States, Mathematics, Grade 4

**International Grade Associated With the Estimated TIMSS Mean Achievement in Each State, 2007 Mathematics, Grade 4**



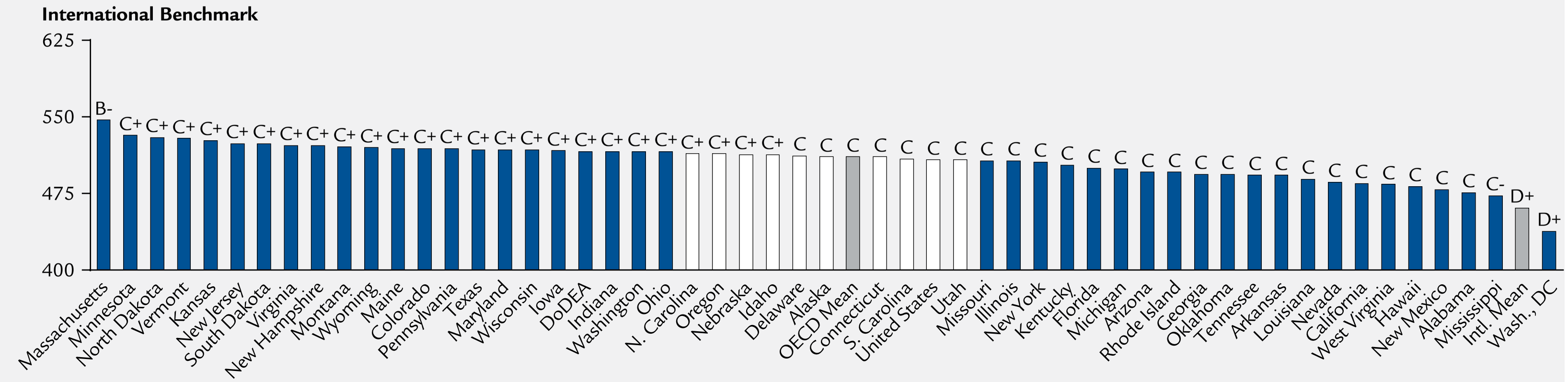

**Estimated TIMSS Mean Achievement in Each State Benchmarked Against an International Grade of "B," 2007 Mathematics, Grade 4**



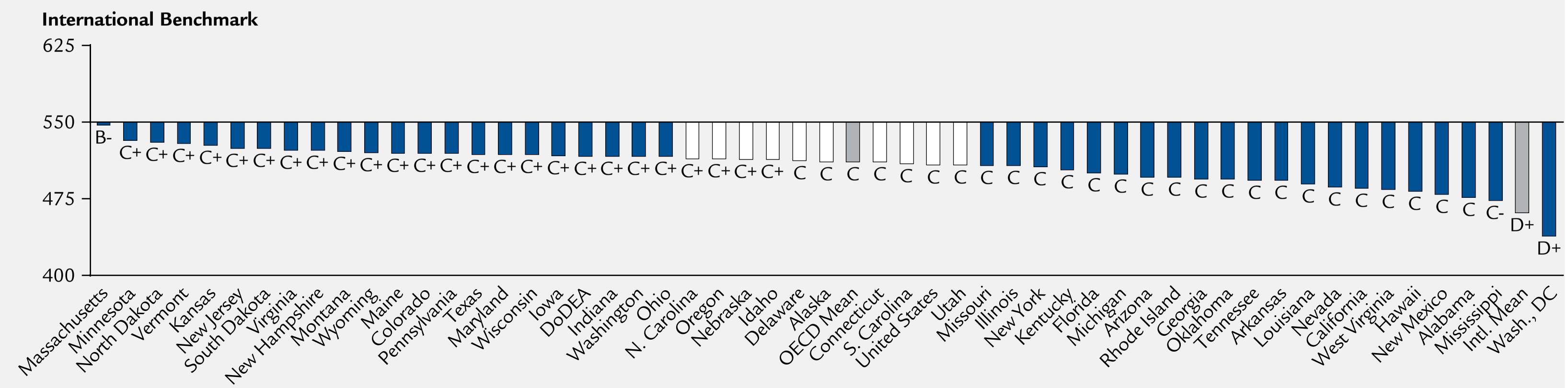

Source of data: Lee, Grigg, & Dion (2007), The Nation's Report Card: Mathematics 2007. Grades determined by Phillips (2009).

## Figure 4: International Grades for States, Mathematics, Grade 8

**International Grade Associated With the Estimated TIMSS Mean Achievement in Each State, 2007 Mathematics, Grade 8**




**Estimated TIMSS Mean Achievement in Each State Benchmarked Against an International Grade of "B," 2007 Mathematics, Grade 8**




Source of data: Lee, Grigg, & Dion (2007), The Nation's Report Card: Mathematics 2007. Grades determined by Phillips (2009).


# International Grades in Mathematics for School Districts

In 2007, 11 large city school districts within the United States (in addition to the District of Columbia) also participated in the NAEP assessment. We can use the statistical linking study (discussed in the Technical Appendix) to also estimate their TIMSS performance. Again, it is important to emphasize that the statistical linking conducted in this study does not "predict" a district's TIMSS performance from the district NAEP performance. In other words, we are not predicting scores on a test that was not administered in the school district. Instead, the study converts the NAEP scores that were obtained in the school district to the TIMSS scale, which then allows us to see if the average district NAEP performance reached the international benchmarks in TIMSS. In the following tables the entries for the United States are actual rather than estimated TIMSS results.

The results from the international benchmarking of school districts are contained in Table 5 and Table 6 and Figure 5 and Figure 6. The findings in the study parallel those for the states in most cases.

The *first* finding is that most of the school districts that participated in the NAEP 2007 are performing at the C level, which represents a relatively basic international level of mathematics performance. This is true for both Grade 4 and Grade 8. One difference between the states and the school districts is that none of the school districts are performing at the B level. In Figure 5 and Figure 6 the dark bars of the chart represent school districts that are significantly below the OECD international average (which is indicated by the grey bar). The white bars in the chart represent school districts that are not significantly different from the OECD international average. Across the United States, the average student in large central cities (cities with populations greater than 250,000) also performed at the C level.

The *second* pattern is that there is also a general tendency among the districts to drop in performance from Grade 4 to Grade 8. This broad drop in performance is illustrated by comparing Figure 5 and Figure 6. In Grade 8 we see that every district has dropped in performance and is further below the international benchmark of B than they were in Grade 4.

**Table 5: International Grades for School Districts in 2007 Mathematics, Grade 4[10]**

| District | Estimated TIMSS Mean | International Grade |
|---|---|---|
| Charlotte | 540 | C+ |
| Austin | 532 | C+ |
| **OECD Mean** | **531** | **C+** |
| **United States** | **529** | **C+** |
| New York City | 519 | C+ |
| Houston | 515 | C+ |
| San Diego | 514 | C+ |
| Boston | 511 | C |
| Large Central Cities | 503 | C |
| Atlanta | 487 | C |
| **International Mean** | **482** | **C** |
| Los Angeles | 480 | C |
| Chicago | 476 | C |
| Cleveland | 465 | C- |
| Washington, DC | 461 | D+ |

[10] Note: The above table reports on the TIMSS international benchmark level of the typical student in the school district (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next highest achievement level is within the 95% confidence interval of the district average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). Source of data: Lutkus, Grigg, & Dion, 2007.

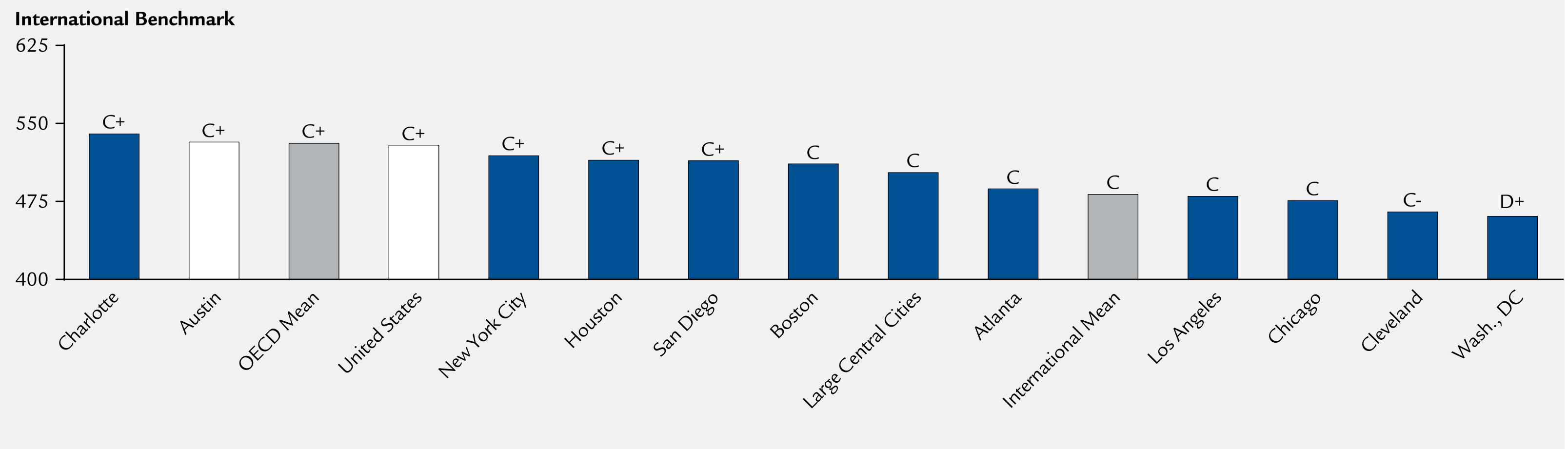
**Table 6: International Grades for School Districts in 2007 Mathematics, Grade 8[11]**

| District | Estimated TIMSS Mean | International Grade |
|---|---|---|
| Charlotte | 511 | C |
| Austin | 511 | C |
| **OECD Mean** | **511** | **C** |
| **United States** | **508** | **C** |
| Boston | 498 | C |
| Houston | 492 | C |
| San Diego | 489 | C |
| New York City | 483 | C |
| Large Central Cities | 482 | C |
| Chicago | 464 | D+ |
| **International Mean** | **461** | **D+** |
| Los Angeles | 457 | D+ |
| Cleveland | 456 | D+ |
| Atlanta | 455 | D+ |
| Washington, DC | 438 | D+ |

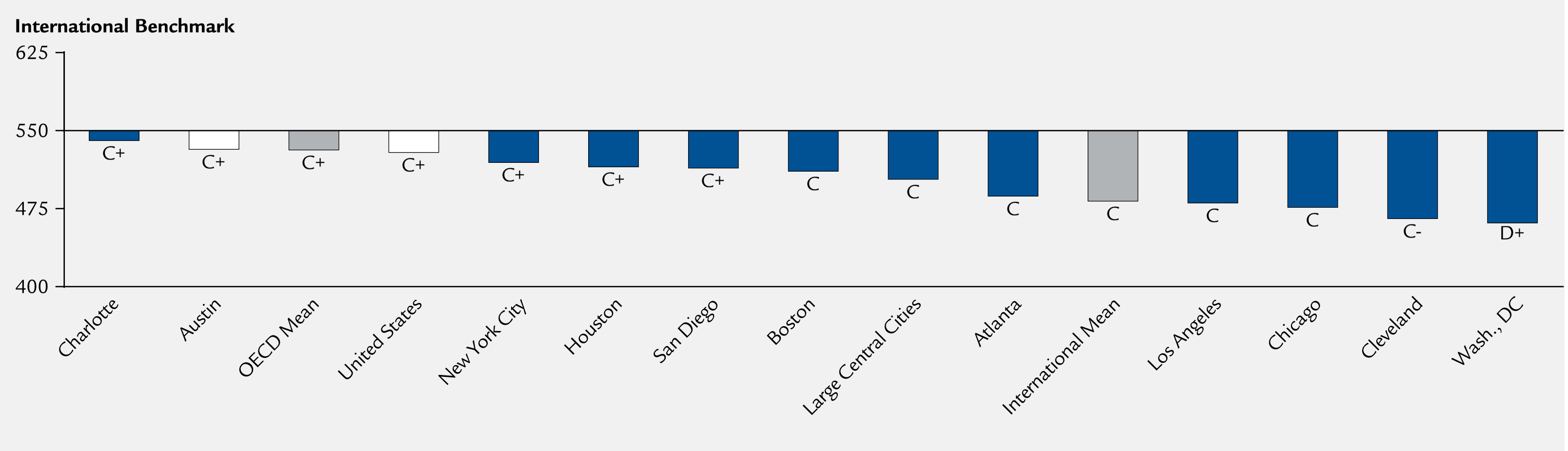
[11] Note: The above table reports on the TIMSS international benchmark level of the typical student in the school district (i.e., the mean student). The grade is based on A = Advanced, B = High, C = Intermediate, D = Low, and BD = below a D. A grade with a minus (e.g., B-) occurs when the next highest achievement level is within the 95% confidence interval of the district average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). Source of data: Lutkus, Grigg, & Dion, 2007.

## Figure 5: International Grades for School Districts, Mathematics, Grade 4

**International Grade Associated With the Estimated TIMSS Mean Achievement in Each School District, 2007 Mathematics, Grade 4**



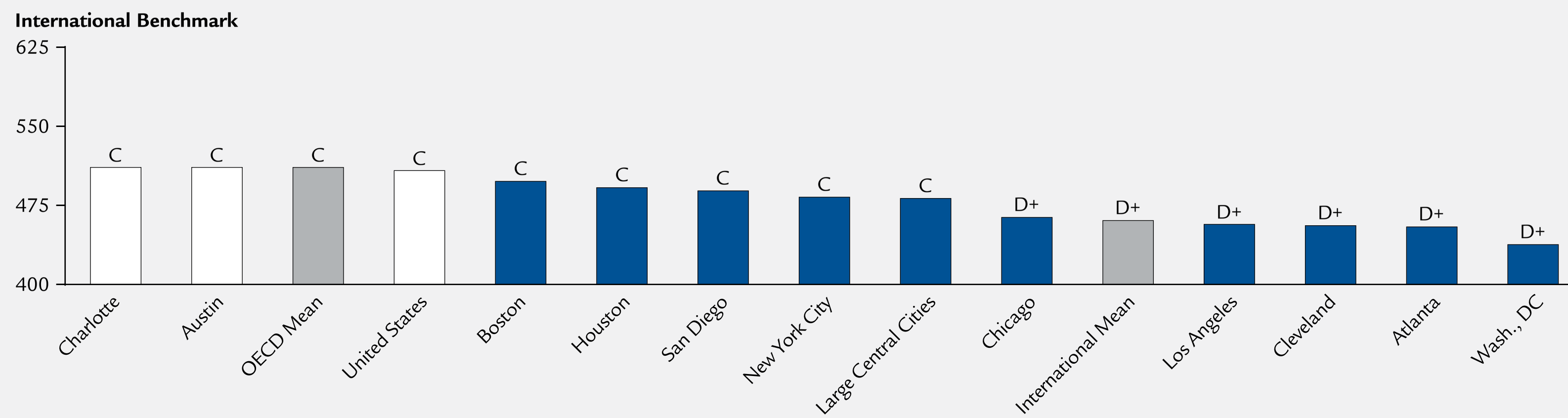

**International Grade Associated With the Estimated TIMSS Mean Achievement in Each School District, 2007 Mathematics, Grade 4**



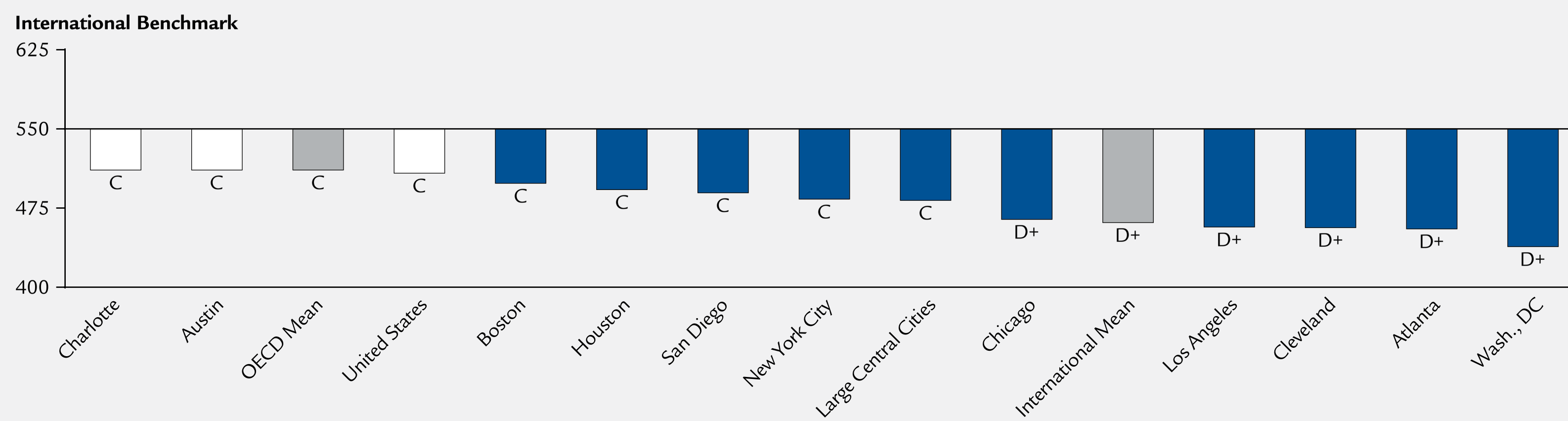

Source of data: Lutkus, Grigg, & Dion (2007), The Nation's Report Card: Trial Urban District Assessment in Mathematics 2007. Grades determined by Phillips (2009).

## Figure 6: International Grades for School Districts, Mathematics, Grade 8

**International Grade Associated With the Estimated TIMSS Mean Achievement in Each School District, 2007 Mathematics, Grade 8**




**Estimated TIMSS Mean Achievement in Each District Benchmarked Against an International Grade of "B," 2007 Mathematics, Grade 8**




Source of data: Lutkus, Grigg, & Dion (2007), The Nation's Report Card: Trial Urban District Assessment in Mathematics 2007. Grades determined by Phillips (2009).

# Relationship Between National Standards and International Standards

One question policymakers often ask is: How do our U.S. national standards compare to international standards? The statistical linking in this study makes it possible to answer this question, at least for NAEP national-achievement-level standards. We can compare the national achievement levels on NAEP (Basic, Proficient, and Advanced) to the international benchmarks on TIMSS (Intermediate, High, and Advanced). To do this we express the NAEP achievement standards in terms of the TIMSS scale and then compare the NAEP standards to the TIMSS standards. This comparison is provided in Table 7 and Table 8.

**Table 7: Comparing TIMSS International Benchmarks to NAEP Achievement Levels Estimated on the TIMSS Scale, Grade 4**

| TIMSS 2007 Math Grade 4 International Benchmarks | | NAEP 2007 Math Grade 4 Achievement Levels on the TIMSS Scale | | Linking Error in NAEP 2007 Achievement Levels on the TIMSS Scale | Compared to TIMSS, the NAEP 2007 Achievement Level Is Statistically |
|---|---|---|---|---|---|
| Intermediate | 475 | Basic | 461 | 3.03 | Lower |
| High | 550 | Proficient | 553 | 2.61 | Same |
| Advanced | 625 | Advanced | 640 | 3.73 | Higher |

Note: Comparisons in achievement standards are done with 95% confidence intervals.

**Table 8: Comparing TIMSS International Benchmarks to NAEP Achievement Levels Estimated on the TIMSS Scale, Grade 8**

| TIMSS 2007 Math Grade 8 International Benchmarks | | NAEP 2007 Math Grade 8 Achievement Levels on the TIMSS Scale | | Linking Error in NAEP 2007 Achievement Levels on the TIMSS Scale | Compared to TIMSS, the NAEP 2007 Achievement Level Is Statistically |
|---|---|---|---|---|---|
| Intermediate | 475 | Basic | 467 | 3.19 | Lower |
| High | 550 | Proficient | 546 | 3.16 | Same |
| Advanced | 625 | Advanced | 618 | 4.31 | Same |

Note: Comparisons in achievement standards are done with 95% confidence intervals.

These analyses lead to the following conclusions:

- The Basic achievement level on NAEP 2007 is significantly *lower* than the Intermediate International Benchmark on TIMSS in Grades 4 and 8.
- The Proficient achievement level on NAEP 2007 is statistically *comparable* to the High International Benchmark on TIMSS in Grades 4 and 8.
- The Advanced achievement level on NAEP 2007 is significantly *higher* than the Advanced International Benchmark on TIMSS at Grade 4, and statistically *comparable* to the Advanced International Benchmark at Grade 8.

These analyses indicate that the NAEP 2007 achievement levels are mostly comparable to the international standards. Relating them to the grading system in this report, we find the Basic level is a little lower than the international C, the Proficient level is a B, and the NAEP Advanced level is higher than an A at Grade 4 and comparable to an A at Grade 8.

# Relationship Between PISA and TIMSS

As mentioned above, two international mathematics assessments are currently available for international benchmarking. These are PISA and TIMSS. Although there is much in common between the two assessments, there is one main difference: PISA is primarily a math literacy assessment and TIMSS is primarily a math proficiency assessment.

Although PISA measures mathematical proficiency, PISA is primarily an assessment of mathematics literacy. The PISA framework (*Assessing, Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006,* page 72) defines mathematical literacy as

> *"an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen."*

From this definition we see that the orientation on PISA is not on how much mathematics you have learned; rather it is on how well you apply the mathematics you have learned to understanding and solving problems in the real world. The literacy orientation to mathematics in PISA is very consistent with the 21st Century Skills initiative, which is gaining momentum within the United States.

Although TIMSS includes measures of mathematical literacy, TIMSS is primarily an assessment of mathematical proficiency. A description of what TIMSS measures can be found in the *TIMSS 2007 Assessment Framework* (Mullis et al., 2007).

> *The content and cognitive domains are the foundation of the TIMSS 2007 fourth- and eighth-grade assessments.... At the eighth grade, two of the four content domains are geometry and algebra, but since geometry and algebra generally are not taught as formal subjects in primary school, the domain assessed at the fourth grade focuses on geometric shapes and measures and introductory algebra concepts are included as part of number. At the fourth grade, the domain pertaining to data focuses on reading and displaying data whereas at eighth grade it includes more emphasis on interpretation of data and the fundamentals of probability.* (Page 13)

This description of what TIMSS measures is in stark contrast to that of PISA. It focuses more on what mathematics teachers should teach in the curriculum and what mathematics students should learn in school. It focuses less on how well they should apply mathematics to societal problems and how well the student is a reflective citizen or intelligent consumer. Because of this focus TIMSS has provided a treasure trove of information about the quality of curriculum and school resources around the world in the *TIMSS 2007 Encyclopedia* (Mullis et al., 2008). TIMSS describes three levels of the curriculum in each country. These are the *intended curriculum* (what the educational system and society intends for students to learn), the *implemented curriculum* (how the educational system is organized and what is actually taught), and the *attained curriculum* (what is actually learned). The more traditional orientation in TIMSS is very consistent with the subject-matter content orientation in NAEP and the curriculum orientation recommended by the National Council of Teachers of Mathematics (NCTM).

Both of these surveys provide important, but different, insights into education around the world. Policymakers and researchers are fortunate to have access to both surveys for purposes of international benchmarking and understanding best practice.

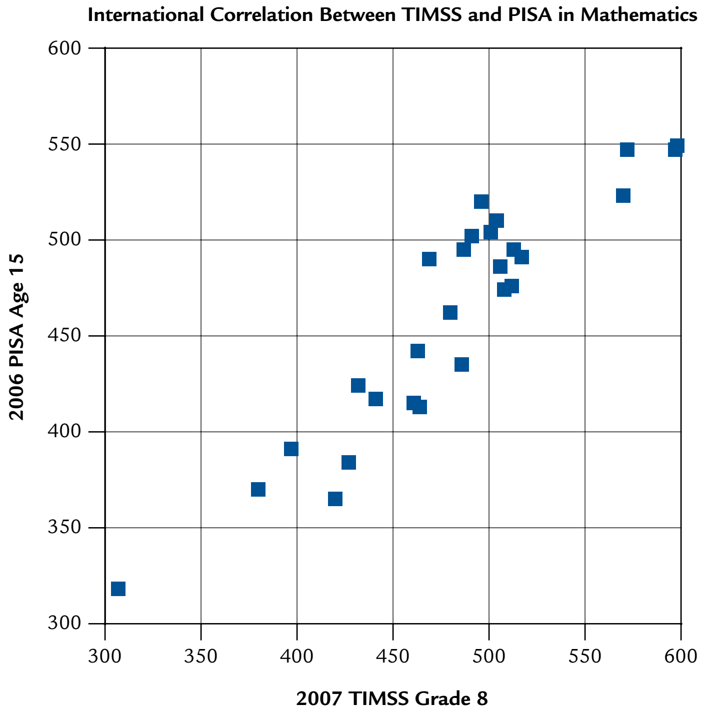
Even though PISA and TIMSS have different orientations it turns out that country performance is highly correlated across the two surveys. In other words, if a country performs well on PISA it is likely to do well on TIMSS. Table 9 shows the average performance for 2007 TIMSS Grade 8 and 2006 PISA age 15 displayed next to each other. These are the 27 countries that participated in both the 2007 TIMSS and the 2006 PISA. A graph of the relationship is shown in Figure 7. We see a high correlation between the TIMSS and PISA results. In fact, the correlation is equal to .93. This high correlation indicates that TIMSS and PISA should provide similar results for U.S. international benchmarking of performance standards. Countries that perform better than the United States on TIMSS are likely to overlap with the countries that perform better than the United States on PISA.

Although there is a high correlation between TIMSS and PISA, we still should study the educational practices and societal supports within these countries to learn about best practice. This is because the correlation in Table 9 is based on between-country variance, but almost all the variation in TIMSS and PISA is within-country variation. Therefore, the most useful information we can learn from these surveys will be found by examining variables that contribute to their success within each country. Between-country variation is more appropriate for international benchmarking *performance standards* (that is the subject of this report). But within-country variation is more appropriate for benchmarking *content standards* and curriculum and practices.

**Table 9: Average Performance on TIMSS and PISA**

| Countries in 2007 TIMSS Grade 8 and PISA 2006 Age 15 | 2007 TIMSS Grade 8 | 2006 PISA Age 15 |
|---|---|---|
| Chinese Taipei | 598 | 549 |
| Korea, Rep. of | 597 | 547 |
| Hong Kong SAR | 572 | 547 |
| Japan | 570 | 523 |
| Australia | 496 | 520 |
| Czech Republic | 504 | 510 |
| Slovenia | 501 | 504 |
| Sweden | 491 | 502 |
| England | 513 | 495 |
| Scotland | 487 | 495 |
| Hungary | 517 | 491 |
| Norway | 469 | 490 |
| Lithuania | 506 | 486 |
| Russian Federation | 512 | 476 |
| **United States** | **508** | **474** |
| Italy | 480 | 462 |
| Israel | 463 | 442 |
| Serbia | 486 | 435 |
| Turkey | 432 | 424 |
| Thailand | 441 | 417 |
| Romania | 461 | 415 |
| Bulgaria | 464 | 413 |
| Indonesia | 397 | 391 |
| Jordan | 427 | 384 |
| Colombia | 380 | 370 |
| Tunisia | 420 | 365 |
| Qatar | 307 | 318 |

### Figure 7: Correlation Between Average Performance on TIMSS and PISA



# Conclusions

This report develops a statistically rigorous and comparable metric that is intuitively easy to understand, that allows policymakers to benchmark the educational achievement for the United States as a whole, the states within our country, and several school districts against tough international achievement standards. The report shows that international benchmarking of educational performance standards is feasible and can be done in a cost-effective way.

There are at least three overall important substantive findings in this report.

The first is that America, and most of the states within the United States, and school districts are learning basic mathematics rather than more complex mathematics. This is indicated by the C+ in Grade 4 and C in Grade 8 for the United States, with similar performance for the overwhelming number of states and school districts.

The second theme is that there are a small set of Asian countries and a few states within the United States that are doing well in mathematics. That is indicated by the grades of B+, B, and B- for those countries and states. No country or state earned an A.

Third, our states and school districts fall comparatively further behind in Grade 8 than they do in Grade 4. Although the United States falls further behind in the higher grade, the highest achieving countries maintain their level of performance.

# References


*Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: Organization for Economic Co-operation and Development, 2006.

*Benchmarking for Success: Ensuring U.S. Students Receive a World-class Education*. National Governors Association, Washington, DC, 2008.

*From Competing to Leading: An International Benchmarking Blueprint*. Education Commission of the States, Denver, CO, 2008.

Johnson, E. G., Cohen, J., Chen, W., Jiang, T., & Zhang, Y. (2005). *2000 NAEP–1999 TIMSS Linking Report*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Holland, P. W. (2007). *A Framework and History for Score Linking*. In “Linking and Aligning Scores and Scales” (Dorans, Pommerich and Holland, Editors). Springer, New York.

Lee, J., Grigg, W. and Dion, G. *The Nation’s Report Card: Mathematics 2007* (NCES 2007–494). Washington, DC, 2007.

Lutkus, A., Grigg, W. and Dion, G. *The Nation’s Report Card: Trial Urban District Assessment in Mathematics 2007* (NCES 2008–452). Washington, DC, 2007.

*Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007–482). U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office, 2007.

Mullis, I., V.S., Martin, M., & Foy, P. *TIMSS 2007 International Mathematics Report*, Lynch School of Education: Boston College, Boston, MA, 2008.

Mullis, I.V.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., & Stanco, G.M. (Eds.). *TIMSS 2007 Encyclopedia: A Guide to Mathematics and Science Education Around the World*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, 2008.

Mullis, I.V.S., Martin, M.O., Ruddock, G. J., O’Sullivan, C. Y., Arora, A., & Erberber, E. *TIMSS 2007 Assessment Framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, 2007.

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

Phillips, G. W., (2007a). *Chance Favors the Prepared Mind: Mathematics and Science Indicators for Comparing States and Nations*. American Institutes for Research: Washington, DC, 2007.

Phillips, G. W. (2007b). *Expressing International Educational Achievement in Terms of U.S. Performance Standards: Linking NAEP Achievement Levels to TIMSS*. Washington, DC: American Institutes for Research, 2007.

Phillips, G. & Dossey, J. *Counting on the Future: International Benchmarks in Mathematics for American School Districts*. Washington, DC: American Institutes for Research, 2008.

*PISA 2006: Science Competencies for Tomorrow's World, Vol. 1*. Paris: Organization for Economic Co-operation and Development, 2007.

Wolter, K. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.


# Technical Appendix

*Linking the NAEP 2007 to the TIMSS 2007 in Mathematics*

## What Is Statistical Linking?

Statistical linking as used in this report is essentially expressing the results of one test (e.g., NAEP indicated by $x$) in terms of the metric of another test (e.g., TIMSS indicated by $y$).

An analogy might help illustrate what this means. Let's say we have a meteorologist $y$ hired by a company called the IEA in the Netherlands to conduct a study in 2007 to determine the average temperature of school buildings among 48 countries. One of the countries (the United States) is authorized by Congress to hire meteorologist $x$ under contract to a statistical agency NCES to conduct a similar study in 2007 among the 50 states within the United States. Each meteorologist draws separate (but randomly equivalent) nationally representative samples within the United States at the same time. Meteorologist $y$ measures temperature with the Celsius scale and meteorologist $x$ measures temperature with the Fahrenheit scale. Meteorologist $y$ finds the average temperature within the United States equal to 25 degrees and meteorologist $x$ finds the average temperature within the United States equal to 77 degrees. Policymakers want to compare the results of the two studies but they are confused and frustrated by the fact that the international study uses Celsius while the American study uses Fahrenheit.

Let's extend this analogy one step further to parallel the analyses in this report. During the previous administration in Washington, legislation called No Child Left Behind encouraged each state to establish their own school building temperature standard and report how many buildings were meeting that standard to the federal government. However, after a new administration was elected it was decided that every state having its own standard was ridiculous, so an effort was made to reduce the variability in state standards. In order to get a handle on how best to accomplish this, it was decided to start by benchmarking each state standard against the well respected international temperature standards established by meteorologist $y$ among the 48 countries.

A researcher at the American Institutes for Research uses a linking study which he conducts in the U.S. national sample to transform the Fahrenheit scale used in the state samples to the Celsius scale used in the international samples. He knows that if he converts

Celsius to Fahrenheit then $x = A + By = 32 + 1.8y$, where $A = 32$ and $B = 1.8$. On the other hand, if he converts Fahrenheit to Celsius then we find $y = \frac{1}{B}(x - A)$. However, both conversions assume that we know the parameters $A$ and $B$. If we do not know the parameters then they must be estimated from the sample data (that is the situation we are in when linking NAEP to TIMSS where both tests measure essentially the same construct but the results are expressed in different metrics and we do not know the parameters $A$ and $B$). The researcher estimates $A$ and $B$ empirically from the U.S. national sample data and acknowledges that these estimates are affected by measurement error as well as sampling error. Now we can see how the state standards stack up against the international standard. The researcher cautions that the "estimates" of Celsius measurements for each state have more error variance (the linking error variance) than they would if each state had "actual" Celsius measurements. Although "actual" Celsius measurements would be better than "estimates" of Celsius measures, the extra cost and burden of having both meteorologists conduct studies in each state is not cost-effective.

To complete the analogy, imagine that two states, Massachusetts and Minnesota, were able to hire meteorologist $y$ so these two states have measurements using both the Fahrenheit as well as the Celsius scale. These two states can be used to help validate the estimates obtained in all the states because in Massachusetts and Minnesota we can compare the *estimated* Celsius results against *actual* Celsius results. The validation comparison in the two states indicate that the linking estimates were good in most cases but not in all cases. The conclusion from these analyses was that the linking strategy was a cost-effective method that was good enough to provide reasonable estimates. However, if we needed more precise and more comprehensive information, then we should incur the extra expense of hiring meteorologist $y$ as well as $x$ in each state.

There is one important way in which the linking between NAEP and TIMSS is different from the temperature analogy. True measurements between Fahrenheit and Celsius are perfectly correlated, whereas the correlation between true scores on NAEP and TIMSS is less than 1.0. If it is possible to estimate this correlation, then that information could be used to improve the precision of the TIMSS estimates. The only way to estimate this correlation would be to administer NAEP and TIMSS in such a way that the correlation could be calculated (e.g., administer both NAEP and TIMSS to the same random sample of students).

## Linking NAEP 2007 to TIMSS 2007

In 2007 the National Center for Education (NCES) released a report in which the state achievement standards on each state achievement test were mapped onto the NAEP scale (*Mapping 2005 State Proficiency Standards Onto the NAEP Scales*, 2007). This resulted in what was referred to as *NAEP-equivalent* achievement standards (the score on the NAEP scale that was equivalent to the state achievement standard). This result was obtained through statistically linking each state achievement test to the NAEP scale. The current report follows a similar strategy for statistical linking. However, in this report we are finding the *TIMSS-equivalent* score that is associated with the state-NAEP or district-NAEP mean. This is the crosswalk that allows us to determine how the performance of the state and school district compares with the international benchmarks provided by TIMSS.

This report uses the statistical linking procedures outlined in Johnson et al. (2005) and Phillips (2007b), in which NAEP was linked to TIMSS by using statistical moderation. One major difference is that this report uses extant statistics from the NAEP 2007 and TIMSS 2007 published reports rather than recalculating them from the public-use data files and plausible values available from the NAEP and TIMSS assessments.

In the following discussion, $Y$ denotes TIMSS and $X$ denotes NAEP. In statistical moderation, the estimated $z$ score is a transformed $x$ score expressed in the $y$ metric

$$z = \hat{A} + \hat{B}(x)$$
$$= \left(\hat{\mu}_Y - \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\hat{\mu}_X\right) + \left(\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\right)x \quad (1.1)$$

In equation (1.1), $\hat{A}$ is an estimate of the intercept of a straight line, and $\hat{B}$ is an estimate of the slope defined by

$$\hat{A} = \hat{\mu}_Y - \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\hat{\mu}_X \quad (1.2)$$

$$\hat{B} = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \quad (1.3)$$

In equation (1.3), $\hat{\mu}_X$ and $\hat{\mu}_Y$ are the national means of the U.S. NAEP and U.S. TIMSS, respectively, while $\hat{\sigma}_X$ and $\hat{\sigma}_Y$ are the national standard deviations of the assessments.

## Linking Error Variance

The linking error variance in the *TIMSS-equivalent* means for each state can be determined through the following equation.

$$\hat{\sigma}^2_{\bar{z}} = \hat{B}^2\hat{\sigma}^2_{\bar{x}} + \hat{\sigma}^2_A + 2(\bar{x})\hat{\sigma}_{AB} + (\bar{x})^2\hat{\sigma}^2_B \quad (1.4)$$

According to Johnson et al. (2005), the error variances in this equation, $\hat{\sigma}^2_A$, $2\hat{\sigma}_{AB}$, and $\hat{\sigma}^2_B$ can be approximated by Taylor-series linearization (Wolter, 1985).

$$\hat{\sigma}^2_A = \hat{B}^2\hat{\sigma}^2_{\mu_X} + \hat{\sigma}^2_{\mu_Y} + \hat{\mu}^2_X\hat{B}^2\left[\frac{Var(\hat{\sigma}_Y)}{\hat{\sigma}^2_Y} + \frac{Var(\hat{\sigma}_X)}{\hat{\sigma}^2_X}\right]$$

$$2\hat{\sigma}_{AB} = -2\hat{\mu}_X\hat{B}^2\left[\frac{Var(\hat{\sigma}_Y)}{\hat{\sigma}^2_Y} + \frac{Var(\hat{\sigma}_X)}{\hat{\sigma}^2_X}\right]$$

$$\hat{\sigma}^2_B = \hat{B}^2\left[\frac{Var(\hat{\sigma}_Y)}{\hat{\sigma}^2_Y} + \frac{Var(\hat{\sigma}_X)}{\hat{\sigma}^2_X}\right] \quad (1.5)$$

Equations (1.4) and (1.5) were used with data in the U.S. linking sample to derive the estimates of linking error variance in this paper.

The statistics needed to use equations (1.1) through (1.5) are contained in Tables 10 and Table 11 on the following page.

The parameter $\hat{A}$ and $\hat{B}$ estimates are indicated in Table 12 and Table 13. These are the intercepts and slopes, respectively, needed to re-express NAEP results on the TIMSS scale.

## Relationship Between International Benchmarks and International Grades

One of the primary ways the TIMSS reports its results is in terms of achievement standards (fortuitously referred to in the report as international benchmarks). The labels and cut-points on the TIMSS scale for the international benchmarks are *Advanced* (625), *High* (550), *Intermediate* (475), and *Low* (400).

These achievement standards apply to both Grade 4 as well as Grade 8 mathematics assessment. The achievement standards were initially established in the first TIMSS population in 1995 where they represented the 90th percentile (Advanced), the 75th percentile (High), the 50th percentile (Intermediate), and the 25th percentile (Low). The substantive content definitions of each of these international benchmarks are provided in the 2007 TIMSS report (see Mullis, Martin, & Foy, 2008, pages 68–69). The relationship between the TIMSS international benchmarks and the international grade is presented in Table 14.

This report uses these international benchmarks to conduct a secondary analysis of TIMSS. However, in order to facilitate communication with an American audience the report will use the cut-scores on the TIMSS scale to create a new index using a metric all American educators and policymakers understand. For each country in TIMSS the country mean will be compared to the international benchmark. If the country average is at the Advanced level the country is assigned an international grade of A. Similarly, if the country average is at the High level they are assigned

**Table 10: Means and Standard Deviations for National Samples of Grade 4 TIMSS 2007 and NAEP 2007 in Mathematics**

| | Mean | Error of Mean | Standard Deviation | Error of Standard Deviation |
|---|---|---|---|---|
| TIMSS 2007 Math Grade 4 | 529.00 | 2.45 | 75.33 | 1.76 |
| NAEP 2007 Math Grade 4 | 239.72 | 0.17 | 28.63 | 0.10 |

Sources: Mullis, Martin, & Foy, 2008; Lee, Grigg, & Dion, 2007.

**Table 11: Means and Standard Deviations for National Samples of Grade 8 TIMSS 2007 and NAEP 2007 in Mathematics**

| | Mean | Error of Mean | Standard Deviation | Error of Standard Deviation |
|---|---|---|---|---|
| TIMSS 2007 Math Grade 8 | 508.45 | 2.83 | 76.74 | 2.04 |
| NAEP 2007 Math Grade 8 | 281.35 | 0.27 | 36.07 | 0.13 |

Sources: Mullis, Martin, & Foy, 2008; Lee, Grigg, & Dion, 2007.

**Table 12: Estimating TIMSS 2007 Mathematics From NAEP 2007, Mathematics, Grade 4**

| | Estimates of Linking Parameters A and B | |
|---|---|---|
| | *A* | *B* |
| Parameter | -101.79 | 2.63 |
| Standard error | 15.13 | 0.06 |
| Covariance | -0.93 | |

**Table 13: Estimating TIMSS 2007 Mathematics From NAEP 2007, Mathematics, Grade 8**

| | Estimates of Linking Parameters A and B | |
|---|---|---|
| | *A* | *B* |
| Parameter | -90.13 | 2.13 |
| Standard error | 16.29 | 0.06 |
| Covariance | -0.91 | |

**Table 14: Crosswalk Between the TIMSS International Benchmark, the TIMSS Scaled Score, and the International Grade**

| International Benchmark on TIMSS | Cut-score on TIMSS for International Grades | International Grade |
|---|---|---|
| Advanced | 625 | A |
| | 625 - 1.96*SE Mean | A- |
| | 587.5 | B+ |
| High | 550 | B |
| | 550 - 1.96*SE Mean | B- |
| | 512.5 | C+ |
| Intermediate | 475 | C |
| | 475 - 1.96*SE Mean | C- |
| | 437.5 | D+ |
| Low | 400 | D |
| | 400 - 1.96*SE Mean | D- |
| | Below (400 - 1.96*SE Mean) | BD |

a B, for "Intermediate" they will be given a C, for Low a D, and if the country average is below Low they are given a DB (below a D).

Since the country averages are based on samples of students, the country average will have an associated margin of error (column 3 in Table 15 through Table 20). The margin of error will be incorporated into the international grade index by assigning a minus (to the next highest) grade if the international benchmark is within the 95% confidence interval of the average. For example, a country whose average is lower (but not significantly below) the Intermediate benchmark would be assigned a C- instead of a D. This gives the country the benefit of the doubt associated with the margin of error of the mean. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. For example, a C+ would occur when the country mean is more than halfway between the Intermediate and High international benchmarks on the TIMSS scale.

**Table 15: Relationship Between International Grades and International Benchmarks in Mathematics for Countries, Grade 4**

| Country | TIMSS Mean | Standard Error of TIMSS Mean | International Benchmark Level of the Mean | Grade |
|---|---|---|---|---|
| Hong Kong SAR | 607 | 3.6 | High | B+ |
| Singapore | 599 | 3.7 | High | B+ |
| Chinese Taipei | 576 | 1.7 | High | B |
| Japan | 568 | 2.1 | High | B |
| Kazakhstan | 549 | 7.1 | Intermediate | B- |
| Russian Federation | 544 | 4.9 | Intermediate | B- |
| England | 541 | 2.9 | Intermediate | C+ |
| Latvia | 537 | 2.3 | Intermediate | C+ |
| Netherlands | 535 | 2.1 | Intermediate | C+ |
| **OECD Mean** | **531** | **1.3** | **Intermediate** | **C+** |
| Lithuania | 530 | 2.4 | Intermediate | C+ |
| **United States** | **529** | **2.4** | **Intermediate** | **C+** |
| Germany | 525 | 2.3 | Intermediate | C+ |
| Denmark | 523 | 2.4 | Intermediate | C+ |
| Australia | 516 | 3.5 | Intermediate | C+ |
| Hungary | 510 | 3.5 | Intermediate | C |
| Italy | 507 | 3.1 | Intermediate | C |
| Austria | 505 | 2.0 | Intermediate | C |
| Sweden | 503 | 2.5 | Intermediate | C |
| Slovenia | 502 | 1.8 | Intermediate | C |
| Armenia | 500 | 4.3 | Intermediate | C |
| Slovak Republic | 496 | 4.5 | Intermediate | C |
| Scotland | 494 | 2.2 | Intermediate | C |
| New Zealand | 492 | 2.3 | Intermediate | C |
| Czech Republic | 486 | 2.8 | Intermediate | C |
| **International Mean** | **482** | **1.0** | **Intermediate** | **C** |
| Norway | 473 | 2.5 | Low | C- |
| Ukraine | 469 | 2.9 | Low | D+ |
| Georgia | 438 | 4.2 | Low | D+ |
| Iran, Islamic Republic of | 402 | 4.1 | Low | D |
| Algeria | 378 | 5.2 | Below Low | BD |
| Colombia | 355 | 5.0 | Below Low | BD |
| Morocco | 341 | 4.7 | Below Low | BD |
| El Salvador | 330 | 4.1 | Below Low | BD |
| Tunisia | 327 | 4.5 | Below Low | BD |
| Kuwait | 316 | 3.6 | Below Low | BD |
| Qatar | 296 | 1.0 | Below Low | BD |
| Yemen | 224 | 6.0 | Below Low | BD |

Note: The above table reports on the TIMSS international benchmark level of the typical student in the country (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next-highest achievement level is within the 95% confidence interval of the country average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). *Source of data: Mullis, Martin, & Foy, 2008.*

**Table 16: Relationship Between International Grades and International Benchmarks in Mathematics for Countries, Grade 8**

| Country | TIMSS Mean | Standard Error of TIMSS Mean | International Benchmark Level of the Mean | Grade |
|---|---|---|---|---|
| Chinese Taipei | 598 | 4.5 | High | B+ |
| Korea, Republic of | 597 | 2.7 | High | B+ |
| Singapore | 593 | 3.8 | High | B+ |
| Hong Kong SAR | 572 | 5.8 | High | B |
| Japan | 570 | 2.4 | High | B |
| Hungary | 517 | 3.5 | Intermediate | C+ |
| England | 513 | 4.8 | Intermediate | C+ |
| Russian Federation | 512 | 4.1 | Intermediate | C |
| **OECD Mean** | **511** | **1.5** | **Intermediate** | **C** |
| **United States** | **508** | **2.8** | **Intermediate** | **C** |
| Lithuania | 506 | 2.3 | Intermediate | C |
| Czech Republic | 504 | 2.4 | Intermediate | C |
| Slovenia | 501 | 2.1 | Intermediate | C |
| Armenia | 499 | 3.5 | Intermediate | C |
| Australia | 496 | 3.9 | Intermediate | C |
| Sweden | 491 | 2.3 | Intermediate | C |
| Malta | 488 | 1.2 | Intermediate | C |
| Scotland | 487 | 3.7 | Intermediate | C |
| Serbia | 486 | 3.3 | Intermediate | C |
| Italy | 480 | 3.0 | Intermediate | C |
| Malaysia | 474 | 5.0 | Low | C- |
| Norway | 469 | 2.0 | Low | D+ |
| Cyprus | 465 | 1.6 | Low | D+ |
| Bulgaria | 464 | 5.0 | Low | D+ |
| Israel | 463 | 3.9 | Low | D+ |
| Ukraine | 462 | 3.6 | Low | D+ |
| Romania | 461 | 4.1 | Low | D+ |
| **International Mean** | **461** | **1.0** | **Low** | **D+** |
| Bosnia-Herzegovina | 456 | 2.7 | Low | D+ |
| Lebanon | 449 | 4.0 | Low | D+ |
| Thailand | 441 | 5.0 | Low | D+ |
| Turkey | 432 | 4.8 | Low | D |
| Jordan | 427 | 4.1 | Low | D |
| Tunisia | 420 | 2.4 | Low | D |
| Georgia | 410 | 6.0 | Low | D |
| Iran, Islamic Republic of | 403 | 4.1 | Low | D |
| Bahrain | 398 | 1.6 | Below Low | D- |
| Indonesia | 397 | 3.8 | Below Low | D- |
| Syrian Arab Republic | 395 | 3.8 | Below Low | D- |
| Egypt | 391 | 3.6 | Below Low | BD |
| Algeria | 387 | 2.1 | Below Low | BD |
| Colombia | 380 | 3.6 | Below Low | BD |
| Oman | 372 | 3.4 | Below Low | BD |
| Palestinian Nat'l Auth. | 367 | 3.5 | Below Low | BD |
| Botswana | 364 | 2.3 | Below Low | BD |
| Kuwait | 354 | 2.3 | Below Low | BD |
| El Salvador | 340 | 2.8 | Below Low | BD |
| Saudi Arabia | 329 | 2.9 | Below Low | BD |
| Ghana | 309 | 4.4 | Below Low | BD |
| Qatar | 307 | 1.4 | Below Low | BD |

Note: The above table reports on the TIMSS international benchmark level of the typical student in the country (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next-highest achievement level is within the 95% confidence interval of the country average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). *Source of data: Mullis, Martin, & Foy, 2008.*

**Table 17: Relationship Between International Grades and International Benchmarks in Mathematics for States, Grade 4**

| State | Estimated TIMSS Mean | Standard Error of Estimated TIMSS Mean | International Benchmark Level of Estimated TIMSS Mean | International Grade |
|---|---|---|---|---|
| Massachusetts | 572 | 3.4 | High | B |
| Minnesota | 554 | 3.8 | High | B |
| New Jersey | 552 | 3.8 | High | B |
| New Hampshire | 552 | 3.3 | High | B |
| Kansas | 551 | 3.5 | High | B |
| Vermont | 546 | 2.9 | Intermediate | B- |
| North Dakota | 544 | 2.9 | Intermediate | C+ |
| Indiana | 543 | 3.3 | Intermediate | C+ |
| Ohio | 542 | 3.7 | Intermediate | C+ |
| Wisconsin | 541 | 3.5 | Intermediate | C+ |
| Pennsylvania | 540 | 3.3 | Intermediate | C+ |
| Wyoming | 540 | 2.8 | Intermediate | C+ |
| Montana | 539 | 3.2 | Intermediate | C+ |
| Virginia | 539 | 3.4 | Intermediate | C+ |
| Iowa | 537 | 3.3 | Intermediate | C+ |
| Connecticut | 537 | 3.8 | Intermediate | C+ |
| New York | 536 | 3.3 | Intermediate | C+ |
| Washington | 536 | 3.6 | Intermediate | C+ |
| Maine | 536 | 3.3 | Intermediate | C+ |
| Texas | 536 | 3.1 | Intermediate | C+ |
| Florida | 535 | 3.3 | Intermediate | C+ |
| Delaware | 534 | 2.7 | Intermediate | C+ |
| North Carolina | 534 | 3.2 | Intermediate | C+ |
| South Dakota | 533 | 3.1 | Intermediate | C+ |
| Idaho | 532 | 3.1 | Intermediate | C+ |
| **OECD Mean** | **531** | **1.3** | **Intermediate** | **C+** |
| Maryland | 531 | 3.5 | Intermediate | C+ |
| Colorado | 530 | 3.7 | Intermediate | C+ |
| DoDEA | 530 | 2.8 | Intermediate | C+ |
| **United States** | **529** | **2.4** | **Intermediate** | **C+** |
| Missouri | 528 | 3.5 | Intermediate | C+ |
| Utah | 528 | 3.5 | Intermediate | C+ |
| Nebraska | 525 | 3.9 | Intermediate | C+ |
| Arkansas | 524 | 3.8 | Intermediate | C+ |
| Michigan | 523 | 4.1 | Intermediate | C+ |
| Illinois | 523 | 3.8 | Intermediate | C+ |
| Alaska | 523 | 3.6 | Intermediate | C+ |
| South Carolina | 522 | 3.3 | Intermediate | C+ |
| Oklahoma | 521 | 3.3 | Intermediate | C+ |
| West Virginia | 520 | 3.4 | Intermediate | C+ |
| Oregon | 519 | 3.6 | Intermediate | C+ |
| Rhode Island | 519 | 3.4 | Intermediate | C+ |
| Georgia | 517 | 3.3 | Intermediate | C+ |
| Kentucky | 517 | 3.5 | Intermediate | C+ |
| Hawaii | 515 | 3.3 | Intermediate | C+ |
| Tennessee | 511 | 3.4 | Intermediate | C |
| Arizona | 509 | 3.7 | Intermediate | C |
| Nevada | 508 | 3.5 | Intermediate | C |
| Louisiana | 504 | 3.6 | Intermediate | C |
| California | 504 | 3.2 | Intermediate | C |
| Alabama | 500 | 4.3 | Intermediate | C |
| New Mexico | 498 | 3.6 | Intermediate | C |
| Mississippi | 497 | 3.7 | Intermediate | C |
| **International Mean** | **482** | **1.0** | **Intermediate** | **C** |
| Washington, DC | 461 | 3.7 | Low | D+ |

Note: The above table reports on the TIMSS international benchmark level of the typical student in the state (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next-highest achievement level is within the 95% confidence interval of the state average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). DoDEA = Department of Defense Education Activity. *Source of data: Lee, Grigg, &Dion, 2007.*

**Table 18: Relationship Between International Grades and International Benchmarks in Mathematics for States, Grade 8**

| State | Estimated TIMSS Mean | Standard Error of Estimated TIMSS Mean | International Benchmark Level of Estimated TIMSS Mean | International Grade |
|---|---|---|---|---|
| Massachusetts | 547 | 4.1 | Intermediate | B- |
| Minnesota | 532 | 3.7 | Intermediate | C+ |
| North Dakota | 530 | 3.3 | Intermediate | C+ |
| Vermont | 529 | 3.4 | Intermediate | C+ |
| Kansas | 527 | 3.8 | Intermediate | C+ |
| New Jersey | 524 | 3.9 | Intermediate | C+ |
| South Dakota | 524 | 3.5 | Intermediate | C+ |
| Virginia | 522 | 3.8 | Intermediate | C+ |
| New Hampshire | 522 | 3.3 | Intermediate | C+ |
| Montana | 521 | 3.3 | Intermediate | C+ |
| Wyoming | 520 | 3.4 | Intermediate | C+ |
| Maine | 519 | 3.4 | Intermediate | C+ |
| Colorado | 519 | 3.6 | Intermediate | C+ |
| Pennsylvania | 519 | 3.8 | Intermediate | C+ |
| Texas | 518 | 3.6 | Intermediate | C+ |
| Maryland | 518 | 3.9 | Intermediate | C+ |
| Wisconsin | 518 | 3.7 | Intermediate | C+ |
| Iowa | 517 | 3.5 | Intermediate | C+ |
| DoDEA | 516 | 3.4 | Intermediate | C+ |
| Indiana | 516 | 3.8 | Intermediate | C+ |
| Washington | 516 | 3.6 | Intermediate | C+ |
| Ohio | 516 | 3.9 | Intermediate | C+ |
| North Carolina | 514 | 3.7 | Intermediate | C+ |
| Oregon | 514 | 3.8 | Intermediate | C+ |
| Nebraska | 513 | 3.6 | Intermediate | C+ |
| Idaho | 513 | 3.5 | Intermediate | C+ |
| Delaware | 512 | 3.2 | Intermediate | C |
| Alaska | 511 | 3.7 | Intermediate | C |
| **OECD Mean** | **511** | **1.5** | **Intermediate** | **C** |
| Connecticut | 511 | 4.3 | Intermediate | C |
| South Carolina | 509 | 3.6 | Intermediate | C |
| **United States** | **508** | **2.8** | **Intermediate** | **C** |
| Utah | 508 | 3.5 | Intermediate | C |
| Missouri | 507 | 3.6 | Intermediate | C |
| Illinois | 507 | 3.8 | Intermediate | C |
| New York | 506 | 3.9 | Intermediate | C |
| Kentucky | 503 | 3.8 | Intermediate | C |
| Florida | 500 | 4.0 | Intermediate | C |
| Michigan | 499 | 4.1 | Intermediate | C |
| Arizona | 496 | 4.0 | Intermediate | C |
| Rhode Island | 496 | 3.3 | Intermediate | C |
| Georgia | 494 | 3.6 | Intermediate | C |
| Oklahoma | 494 | 3.6 | Intermediate | C |
| Tennessee | 493 | 3.8 | Intermediate | C |
| Arkansas | 493 | 3.7 | Intermediate | C |
| Louisiana | 489 | 3.8 | Intermediate | C |
| Nevada | 486 | 3.5 | Intermediate | C |
| California | 485 | 3.5 | Intermediate | C |
| West Virginia | 484 | 3.7 | Intermediate | C |
| Hawaii | 482 | 3.4 | Intermediate | C |
| New Mexico | 479 | 3.6 | Intermediate | C |
| Alabama | 476 | 4.4 | Intermediate | C |
| Mississippi | 473 | 3.5 | Low | C- |
| **International Mean** | **461** | **1.0** | **Low** | **D+** |
| Washington, DC | 438 | 4.0 | Low | D+ |

Note: The above table reports on the TIMSS international benchmark level of the typical student in the state (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next-highest achievement level is within the 95% confidence interval of the state average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). DoDEA = Department of Defense Education Activity. *Source of data: Lee, Grigg, &Dion, 2007*.

**Table 19: Relationship Between International Grades and International Benchmarks in Mathematics for School Districts, Grade 4**

| District | Estimated TIMSS Mean | Standard Error of Estimated TIMSS Mean | International Benchmark Level of Estimated TIMSS Mean | International Grade |
|---|---|---|---|---|
| Charlotte | 540 | 3.8 | Intermediate | C+ |
| Austin | 532 | 4.0 | Intermediate | C+ |
| **OECD Mean** | **531** | **1.3** | **Intermediate** | **C+** |
| **United States** | **529** | **2.4** | **Intermediate** | **C+** |
| New York City | 519 | 4.3 | Intermediate | C+ |
| Houston | 515 | 3.9 | Intermediate | C+ |
| San Diego | 514 | 4.5 | Intermediate | C+ |
| Boston | 511 | 3.9 | Intermediate | C |
| Large Central Cities | 503 | 3.0 | Intermediate | C |
| Atlanta | 487 | 3.6 | Intermediate | C |
| **International Mean** | **482** | **1.0** | **Intermediate** | **C** |
| Los Angeles | 480 | 3.7 | Intermediate | C |
| Chicago | 476 | 3.9 | Intermediate | C |
| Cleveland | 465 | 5.1 | Low | C- |
| Washington, DC | 461 | 3.7 | Low | D+ |

Note: The above table reports on the TIMSS international benchmark level of the typical student in the district (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next-highest achievement level is within the 95% confidence interval of the district average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). *Source of data: Lutkus, Grigg, &Dion, 2007*.

**Table 20: Relationship Between International Grades and International Benchmarks in Mathematics for School Districts, Grade 8**

| District | Estimated TIMSS Mean | Standard Error of Estimated TIMSS Mean | International Benchmark Level of Estimated TIMSS Mean | International Grade |
|---|---|---|---|---|
| Charlotte | 511 | 3.9 | Intermediate | C |
| Austin | 511 | 3.8 | Intermediate | C |
| **OECD Mean** | **511** | **1.5** | **Intermediate** | **C** |
| **United States** | **508** | **2.8** | **Intermediate** | **C** |
| Boston | 498 | 3.6 | Intermediate | C |
| Houston | 492 | 3.9 | Intermediate | C |
| San Diego | 489 | 4.2 | Intermediate | C |
| New York City | 483 | 4.9 | Intermediate | C |
| Large Central Cities | 482 | 3.4 | Intermediate | C |
| Chicago | 464 | 5.1 | Low | D+ |
| **International Mean** | **461** | **1.0** | **Low** | **D+** |
| Los Angeles | 457 | 4.0 | Low | D+ |
| Cleveland | 456 | 4.9 | Low | D+ |
| Atlanta | 455 | 4.6 | Low | D+ |
| Washington, DC | 438 | 4.0 | Low | D+ |

Note: The above table reports on the TIMSS international benchmark level of the typical student in the district (i.e., the mean student). The grade is based on A = Advanced (625), B = High (550), C = Intermediate (475), D = Low (400), and BD = below a D. A grade with a minus (e.g., B-) occurs when the next-highest achievement level is within the 95% confidence interval of the district average. A grade with a plus (e.g., C+) occurs when the mean is more than halfway between international benchmarks. The international averages have been weighted by the student population size of each country. The shaded cells indicate the mean is significantly above or below the OECD international average (using a 95% confidence interval). *Source of data: Lutkus, Grigg, &Dion, 2007*.

## Validity Evidence for the Statistical Linking

The linking study reported here involved re-expressing the NAEP results in the national sample in terms of the metric of the TIMSS scale. The linking parameter estimates obtained in the national sample are then applied to the state-NAEP samples where NAEP was administered but TIMSS was not administered. In 2007 two states, Massachusetts and Minnesota, participated in both the state-NAEP assessment as well as a state-TIMSS assessment. This provides an opportunity to validate the TIMSS estimates obtained from the linking against the actual TIMSS results obtained in the two states. These comparisons are provided in Table 21 and Table 22.

From Table 21 we see that the national linking study provided an underestimate of the Massachusetts mean at Grade 4. However, for Minnesota the estimate of the Grade 4 TIMSS mean was not significantly different from the actual TIMSS mean. Furthermore, from Table 22 we see that the linking study provided accurate estimates of the state means in Grade 8 for both Massachusetts and Minnesota. It is a typical frustration in most linking studies that only limited validity information is available. For example, in the linking study reported here it was only possible to validate the linkage in two states but it would have been desirable to do so in all 50 states. This limited validity evidence suggests that the 2007 linking between NAEP and TIMSS was generally successful.

A second check on the validity of the linking in this report can be provided by comparing the estimates in this study to similar estimates in two earlier studies conducted by the author. The three studies used are

1. Eighth-grade mathematics and science with TIMSS 1999 and NAEP 2000,
2. Fourth- and eighth-grade mathematics with TIMSS 2003 and NAEP 2003, and
3. Fourth- and eighth-grade mathematics with TIMSS 2007 and NAEP 2007 (which is the current study).

In the tables below the estimates of the *TIMSS-equivalent* achievement levels are determined from all three linking studies. These are the NAEP achievement levels on the NAEP converted to the metric of TIMSS. If the three separate linking studies conducted over a 10-year period give consistent results, then the estimates of the NAEP achievement levels on the TIMSS scale should be consistent across the three studies. Linking has been done for Grade 3 mathematics twice and the comparison is provided in Table 23. The estimates of all three NAEP achievement levels are consistent across both studies. Linking has been done for Grade 8 mathematics three times and they are presented in Table 24. The estimates of all three achievement levels are again consistent across the three studies, with the single exception of the Advanced level, when the TIMSS 1999–NAEP 2000 study is compared to the TIMSS 2007–NAEP 2007. Overall, this represents a remarkable degree of consistency across 10 years and three separate studies and provides evidence of the validity of the current study.

## Caveats

Holland (2007) has recently outlined three broad categories of linking. These are *equating, scale alignment,* and *prediction*. This report uses the second type: scale alignment. The method of linking is statistical moderation based on the aggregate reporting of NAEP and TIMSS. It is the scales of the total aggregate distributions that are aligned, so the linking should not be used for disaggregated reporting of individual students or demographic subgroups (such as race/ethnicity or gender) or subpopulations (such as schools). Also, the reader should be aware that the concordance between NAEP and TIMSS established in this report for 2007 may not be applicable in subsequent years. However, the relationship between NAEP and TIMSS has largely remained the same over the past decade across three separate linking studies conducted by the author, as indicated by Table 23 and Table 24.

**Table 21: Validity of the NAEP 2007–TIMSS 2007 Linkage for State Estimates, Grade 4**

| | TIMSS Mean Estimated From the NAEP Mean for 2007, Grade 4 | | | | |
|---|---|---|---|---|---|
| | Estimated Mean on TIMSS 2007 | Standard Error of Estimated Mean on TIMSS 2007 | Actual Mean on TIMSS 2007 | Standard Error of Actual Mean on TIMSS 2007 | z-test |
| Massachusetts | 562 | 3.43 | 572 | 3.51 | -2.04 |
| Minnesota | 551 | 3.48 | 554 | 5.86 | -0.52 |

**Table 22: Validity of the NAEP 2007–TIMSS 2007 Linkage for State Estimates, Grade 8**

| | TIMSS Mean Estimated From the NAEP Mean for 2007, Grade 8 | | | | |
|---|---|---|---|---|---|
| | Estimated Mean on TIMSS 2007 | Standard Error of Estimated Mean on TIMSS 2007 | Actual Mean on TIMSS 2007 | Standard Error of Actual Mean on TIMSS 2007 | z-test |
| Massachusetts | 544 | 4.08 | 547 | 4.56 | -0.56 |
| Minnesota | 531 | 3.69 | 532 | 4.44 | -0.28 |

**Table 23: Validity of the NAEP 2007–TIMSS 2007 Linkage for Estimates of NAEP Achievement Levels, Grade 4**

| | Achievement Level Estimates for TIMSS2003–NAEP2003 Versus TIMSS2007–NAEP2007, Grade 4 | | | | | | |
|---|---|---|---|---|---|---|---|
| | TIMSS-equivalent 2003 | NAEP Achievement Level 2003 | SE Linking 2003 | TIMSS-equivalent 2007 | NAEP Achievement Level 2007 | SE Linking 2007 | z-test |
| Basic | 462 | 214 | 2.8 | 461 | 214 | 3.0 | -0.18 |
| Proficient | 556 | 249 | 2.7 | 553 | 249 | 2.7 | -0.71 |
| Advanced | 645 | 282 | 3.9 | 640 | 282 | 3.7 | -0.83 |

**Table 24: Validity of the NAEP 2007–TIMSS 2007 Linkage for Estimates of NAEP Achievement Levels, Grade 8**

| | Achievement Level Estimates for TIMSS1999–NAEP2000 Versus TIMSS2007–NAEP2007, Grade 8 | | | | | | |
|---|---|---|---|---|---|---|---|
| | TIMSS-equivalent 1999 | NAEP Achievement Level 2000 | SE Linking 1999 | TIMSS-equivalent 2007 | NAEP Achievement Level 2007 | SE Linking 2007 | z-test |
| Basic | 469 | 262 | 4.8 | 467 | 262 | 3.2 | -0.27 |
| Proficient | 556 | 299 | 5.1 | 546 | 299 | 3.2 | -1.72 |
| Advanced | 637 | 333 | 6.7 | 618 | 333 | 4.3 | -2.32 |
| | **Achievement Level Estimates for TIMSS2003–NAEP2003 Versus TIMSS2007–NAEP2007, Grade 8** | | | | | | |
| | **TIMSS-equivalent 2003** | **NAEP Achievement Level 2003** | **SE Linking 2003** | **TIMSS-equivalent 2007** | **NAEP Achievement Level 2007** | **SE Linking 2007** | **z-test** |
| Basic | 473 | 262 | 3.6 | 467 | 262 | 3.2 | -1.23 |
| Proficient | 555 | 299 | 3.8 | 546 | 299 | 3.2 | -1.90 |
| Advanced | 631 | 333 | 5.4 | 618 | 333 | 4.3 | -1.84 |

AIR®

Headquarters ■ 1000 Thomas Jefferson Street, NW ■ Washington, DC 20007-3835 ■ www.air.org