AMERICAN INSTITUTES FOR RESEARCH®

Counting on the Future:

International Benchmarks in Mathematics for American School Districts

Dr. Gary W. Phillips
Chief Scientist
American Institutes for Research

John A. Dossey
Distinguished Professor of Mathematics
Illinois State University

With contributions by

Tao Jiang
Senior Psychometrician
American Institutes for Research

Steve Leinwand
Principal Research Analyst in Mathematics
American Institutes for Research

October 2008

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 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 district policymakers with international benchmarks against which they can compare and monitor the educational performance of their students.

In a highly interconnected world, the students served by urban school systems—the subject of this report—will require strong mathematic skills to compete against their peers around the globe. Reports such as *Counting on the Future* help policymakers and educators to know how well they are doing in meeting this challenge and to track progress over time.

Future reports in this series will update and expand on the types of international comparisons presented in this study.

About AIR

Established in 1946, with headquarters in Washington, D.C. and with nearly 30 offices in the United States and around the world, the 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

Executive Summary 1

Why This Report Matters 3

How the Study Was Conducted 5

Introduction 7

Context for the Study: Mathematics Literacy in American Society 11

Major Findings 13

International Benchmarks for the United States in Grade 4 and Grade 8 Mathematics 14

International Benchmarks for 11 U.S. Urban Cities in Grade 4 and Grade 8 Mathematics 17

Comparison of Each District With the Overall International Average 31

Comparison of Each District With the OECD International Average 31

Comparison of Public School Students From Large Central Cities With the Overall International Average 32

Comparison of Public School Students From Large Central Cities With the OECD International Average 32

Criterion-Referenced Interpretations 32

Discussion and Conclusions 43

References 45

Technical Appendix A: Statistical Linking 2003 TIMSS to 2003 NAEP 49

Technical Appendix B: Significance Testing and Multiple Comparisons 67

List of Tables

Table 1: NAEP Grade 4 Mathematics Achievement Levels Projected on the TIMSS Scale for Countries 37

Table 2: NAEP Grade 4 Mathematics Achievement Levels for States 38

Table 3: NAEP Grade 4 Mathematics Achievement Levels for Districts 39

Table 4: NAEP Grade 8 Mathematics Achievement Levels Projected on the TIMSS Scale for Countries 40

Table 5: NAEP Grade 8 Mathematics Achievement Levels for States 41

Table 6: NAEP Grade 8 Mathematics Achievement Levels for Districts 42

Table 7: Statistically Linking Test X and Test Y 50

Table 8: Means and Standard Deviations for National Samples of Grade 4 U.S. Public School Students, 2003 TIMSS and 2003 NAEP Mathematics 54

Table 9: Means and Standard Deviations for National Samples of Grade 8 U.S. Public School Students, 2003 TIMSS and 2003 NAEP Mathematics 54

Table 10: Estimating 2003 NAEP Mathematics Scores From 2003 TIMSS Mathematics Using Statistical Moderation With U.S. Grade 4 Public School National Samples 54

Table 11: Estimating 2003 NAEP Mathematics Scores From 2003 TIMSS Mathematics Using Statistical Moderation With U.S. Grade 8 Public School National Samples 54

Table 12: Grade 4 2003 NAEP Mathematics Achievement Levels Linked to Grade 4 2003 TIMSS 55

Table 13: Grade 8 2003 NAEP Mathematics Achievement Levels Linked to Grade 8 2003 TIMSS 55

Table 14: Percentage of Total Error Variance due to Linking and Sampling for TIMSS Projected onto the NAEP Grade 4 Mathematics Scale. 56

Table 15: Percentage of Total Error Variance due to Linking and Sampling for TIMSS Projected onto the NAEP Grade 8 Mathematics Scale. 56

Table 16: Validity of the Linkage Between NAEP and TIMSS for Grade 8 Mathematics in 2003 56

Table 17: Percentage at and Above Projected Proficient Achievement Level for 2003 TIMSS in Grade 8 Mathematics Based on Two Different Linking Studies 57

Table 18: Projected Mean Versus Actual NAEP Mean for Indiana, 2003, Grade 4, Mathematics 58

Table 19: Percentage at and Above Projected From TIMSS Versus the Actual Percentage at and Above for NAEP in Indiana, 2003, Grade 4, Mathematics. 58

Copies of this paper can be downloaded by searching http://www.air.org, and questions can be addressed to the first author at gwphillips@air.org. Proper citation is as follows: Phillips, G., & Dossey, J. (2008). *Counting on the future: International benchmarks in mathematics for American school districts*. Washington, DC: American Institutes for Research.

List of Tables

Table 20: Projected Mean Versus Actual NAEP Mean for Indiana, 2003, Grade 8, Mathematics 58

Table 21: Percentage at and Above Projected From TIMSS Versus the Actual Percentage at and Above for NAEP in Indiana, 2003, Grade 8, Mathematics. 59

Table 22: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the TIMSS Distribution for Each Country for Grade 4 60

Table 23: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the Projected NAEP Distribution for Each Country for Grade 4 60

Table 24: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the TIMSS Distribution for Each Country for Grade 8 61

Table 25: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the Projected NAEP Distribution for Each Country for Grade 8 61

Table 26: Percentage of Students at and Above the Proficient Achievement Level Projected on 2003 TIMSS, Grade 4, Mathematics. 62

Table 27: Percentage of Students at and Above the Proficient Achievement Level Projected on 2003 TIMSS, Grade 8, Mathematics. 63

Table 28: Each City in the Grade 4 2007 NAEP TUDA in Mathematics Compared to Each Country in the Grade 4 2003 TIMSS in Mathematics for the Percentage of Students at and Above the Proficient Level Based on NAEP Achievement Levels Projected onto the TIMSS Scale 64

Table 29: Each City in Grade 8 2007 NAEP TUDA in Mathematics Compared to Each Country in the Grade 8 2003 TIMSS in Mathematics for the Percentage of Students at and Above the Proficient Level Based on the NAEP Achievement Levels Projected onto the TIMSS Scale 65

List of Figures

Figure 1: United States 16
Figure 2: Atlanta 20
Figure 3: Austin 21
Figure 4: Boston 22
Figure 5: Charlotte 23
Figure 6: Chicago 24
Figure 7: Cleveland 25
Figure 8: District of Columbia 26
Figure 9: Houston 27
Figure 10: Los Angeles 28
Figure 11: New York City 29
Figure 12: San Diego 30
Figure 13: Comparing the School District Performance to the TIMSS Average and OECD Average for Grade 4 Mathematics 33
Figure 14: Comparing the School District Performance to the TIMSS Average and OECD Average for Grade 8 Mathematics 34
Figure 15: Standard Error of Linking Curve for Mathematics Grade 4 55
Figure 16: Standard Error of Linking Curve for Mathematics Grade 8 55

Executive Summary

> *"Globalization is not something we can hold off or turn off…it is the economic equivalent of a force of nature… like the wind and water" (Bill Clinton)*

If you are a student today competing for jobs in a global economy, the good jobs will not go to the best in your graduating class—the jobs will go to the best students in the world. Large urban cities are intimately connected to the nations of the world. Large corporations locate their businesses in U.S. cities; foreign students attend U.S. schools; and U.S. businesses export goods and services to foreign nations. Large urban cities need to know how their students stack up against peers in the nations with which the U.S. does business. This is especially important for students in the fields of science, technology, engineering, and mathematics. The students in these fields will allow our future generation to remain technologically innovative and economically competitive.

This report provides a comparison of the number of mathematically Proficient students in Grades 4 and 8 in 11 large cities in the United States with their international peers. This comparison is made possible by statistically linking the National Assessment of Educational Progress (NAEP) in 2003 and the Trends in International Mathematics and Science Study (TIMSS) in 2003 when both assessments were conducted in the United States in the same year and in the same grades. After the statistical linking was completed, it was possible to compare the most recent NAEP results (from 2007) to the most recent TIMSS results (from 2003).

How the United States compares to the overall international average.

- *At Grade 4, five countries (Singapore, Hong Kong SAR, Chinese Taipei, Japan, and the Flemish portion of Belgium) performed significantly better than the United States (Figure 1). However, the United States (at 39% proficiency) performed better than the international average (27% proficiency) of all 24 countries (Figure 13).*
- *At Grade 8, eight countries (Singapore, Hong Kong SAR, Republic of Korea, Chinese Taipei, Japan, Belgium (Flemish), Netherlands, and Hungary) performed significantly better than the United States (Figure 1). However, the United States (at 31% proficiency) performed better than the international average (21% proficiency) of all 44 countries (Figure 14).*

How the United States compares with the Organization for Economic Cooperation and Development international average.

- *At Grade 4, the report also compared the U.S. average to the average of the 10 countries that are members of the Organization for Economic Cooperation and Development (OECD). These are the more industrialized countries that are likely to be economic competitors of the United States. At Grade 4, the United States (at 39% proficiency) performed significantly higher than the OECD countries (at 30% proficiency) (Figure 13).*
- *At Grade 8, the United States (at 31% proficiency) performed about the same as the OECD countries (at 33% proficiency) (Figure 14).*

How the districts and large central cities compare to the overall international average.

- *At Grade 4, four districts (Charlotte, Austin, San Diego and New York City) performed significantly better than the overall international average of 24 countries. Five districts (Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed below the international average. The remaining two districts (Houston and Boston) performed statistically similar to the overall international average. Across the United States, the average student in large central cities (cities with populations greater than 250,000) performed similarly to the overall international average of 24 countries (Figure 13).*
- *At Grade 8, three districts (Charlotte, Austin, and Boston) performed above the international average of 44 countries. As with Grade 4, the same five districts (Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed below the international average. Three districts (San Diego, New York City, and Houston) performed comparably to the international average. The average U.S. student in large central cities performed similarly to the overall international average of 44 countries (Figure 14).*

How the districts and large central cities compare to the OECD international average.

- *At Grade 4, three districts (Charlotte, Austin, and San Diego) performed above the OECD average. Similar to other results, the same five districts (Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed below the OECD average. Three districts (New York City, Houston, and Boston) performed similarly to the OECD average. The average U.S. student in large central cities performed similarly to the OECD average (Figure 13).*
- *At Grade 8, two districts (Charlotte and Austin) performed similarly to the average of the 12 OECD countries that participated in the international study. The remaining districts performed significantly below the OECD average. In addition, the average U.S. student in large central cities performed statistically below the average of the 12 OECD countries (Figure 14).*

Why This Report Matters

If you are a student today competing for jobs in a global economy, the good jobs will not go to the best in your graduating class—the jobs will go to the best students in the world. Large urban cities are intimately connected to the nations of the world. Large corporations locate their businesses in U.S. cities; foreign students attend U.S. schools; and U.S. businesses export goods and services to foreign nations. Large urban cities need to know how their students stack up against peers in the nations with which the U.S. does business. This is especially important for students in the fields of science, technology, engineering, and mathematics. The students in these fields will allow our future generation to remain technologically innovative and economically competitive.

The National Assessment of Educational Progress (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 underway to include even more districts in the future). The urban school chiefs in these 11 large school districts, which voluntarily participated in the 2007 NAEP, recognized the global nature of educational expectations and the importance of having reliable external data against which to judge the performance of their students and to hold themselves accountable. They should be commended for their visionary goal of trying to benchmark their local performance against tough national standards. National standards provide a broad context and an external compass with which to steer educational policy to benefit local systems. The purpose of this report is to further help those systems navigate by providing international benchmarks.

How the Study Was Conducted

This paper analyzes indicators of performance in mathematics among students in Grades 4 and 8 in 11 U.S. urban school districts. The analysis is based on statistically linking the data collected by the NAEP 2007 Trial Urban District Assessment (TUDA) in Mathematics and the 2003 Trends in International Mathematics and Science Study (TIMSS). The TUDA, conducted by NCES at the U.S. Department of Education, compared the achievement results of 11 urban districts in the United States at Grades 4 and 8 (Lutkus, Grigg, & Dion, 2007). The TIMSS, the other assessment, was conducted by the International Association for the Evaluation of International Achievement (IEA) and involved 25 nations at Grade 4 and 45 nations at Grade 8 (Mullis, Martin, Gonzalez, & Chrostowski, 2004). In Technical Appendix A, data from the two studies are expressed in the same metric on the TIMSS scale through statistical linking.

This process allows for the direct comparison of the percentages of students in the 11 U.S. districts and their international peers who achieved or exceeded the achievement level associated with Proficient or above performance in mathematics, as defined for the NAEP assessments by the NAGB. These comparisons are possible after placing the performance of each of the 11 districts (along with the United States as a whole) on the same metric as the international participants on the TIMSS scale.

Introduction

This report illustrates the policy benefits of linking results from large-scale assessments of student performances in major curriculum areas. In the present case, the focus is on the relative performances of students at Grades 4 and 8 in 11 urban school districts across the United States that participated in the 2007 TUDA of the NAEP compared with students at the same educational levels in countries that participated in the 2003 TIMSS. Statistical linking methods provide the opportunity to examine the outcomes from different studies, such as these, where a common link exists between studies.

The present paper focuses on the comparison of the percentage of students in Grades 4 and 8 in the urban U.S. public school districts and in the TIMSS countries classified as performing at or above the NAEP Proficient achievement level in mathematics. This percentage of students by district and nation provides for policymakers a statistical indicator that helps compare and track student performance over time. Historically, results from NAEP and TIMSS have been used to track student performance over time within the United States and internationally. Linking methods provide a methodology for putting the scores from these two assessments on the same scale, allowing, in the present case, for urban district leaders and policy analysts to compare their cities with the performance of other nations. Educationally, this provides international benchmarks for district policymakers against which they can compare and monitor their own educational performance.

The importance of mathematics in today's society provides a context for the importance of such comparisons. Recent works—such as *The World Is Flat—Brief History of the Twenty-First Century* (Friedman, 2005), *The Singularity Is Near: When Humans Transcend Biology* (Kurzweil, 2006), *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart* (Ayres, 2007), and *The Logic of Life: The Rational Economics of an Irrational World* (Harford, 2008)—all reflect the increased role that mathematical knowledge plays in modeling and decision making in business, biology, finance, and everyday settings. However, the failure of the general population to understand the critical role of mathematics in science, engineering, technology, and business has led to a failure to establish mathematics education as a clear national priority.

This lack of recognizing and understanding the importance of mathematics is perhaps rooted in the lack of overall mathematical literacy in the United States. Until the general public understands the role that mathematics, and more broadly the applications of mathematics, play in shaping and supporting the nation's economy and global presence, little progress can be made in changing the status of mathematics education at the school level. Given the decentralized level of decision making concerning the K–12 school curriculum in the United States, the role that mathematics play in society has yet to capture the attention of school board members and concerned citizens nationwide. Indicators, such as those developed in linking local school assessment outcomes to international benchmarks, provide one way of energizing this process and lifting up the importance of mathematics and mathematics education as tools to prepare the nation's youth to address societal and technological and scientific issues affecting present and future opportunities of students.

Observing student progress toward reaching the levels of mathematical literacy commensurate with the achievement expectation levels set at national and international levels starts when students begin their education. Although the middle grades are often discussed as the launching pad for sending students into the critical core of studies in high school, in mathematics this is too late (National Mathematics Advisory Panel, 2008; Wimberly & Noeth, 2005). Students who do not develop a sound basis of numbers and operations, measurement, geometry, intuitive algebraic reasoning, data analysis, and probability in the elementary grade years are unable to connect important concepts that form the launching pad of the middle grade years.

Examining student progress at Grade 4 provides an excellent opportunity to examine students' grasp of whole numbers and the degree to which they have developed computational fluency with whole number operations. In addition, students should have an emerging understanding of decimals and common fractions. Geometrically, students at this level should have developed the basic concepts of linear measurement and be developing an understanding of area and a knowledge of the basic shapes and their basic properties. Algebraically, these students should have a grasp of arithmetic and geometric sequences based on whole numbers and their operations and should be capable of extending patterns defined by them. In a like manner, students should be able to solve simple number sentences that have whole number solutions. In data analysis, students should be able to construct and interpret frequency tables, bar graphs, line plots, and pictographs and organize and summarize data related to a question of interest using measures of center and spread, both verbally and with simple summary statistics (NAGB, 2006; National Mathematics Advisory Panel, 2008).

Examining student achievement at Grade 8 provides a picture of students' increasing grasp of numbers and operations, with the focus now on fractions and real numbers and their use in interpreting ratios and solving proportional relationships. Students' understanding of geometry and measurement should expand to include relationships from two- to three-dimensional settings. For example, by the end of Grade 8, students should have a solid grasp of standard solid figures and their defining properties and measures of area and volume. In data analysis, students should have transitioned from questions about simple data values to situations that call for the comparison of related data sets or situations that call for several responses from each subject of interest. Algebraically, students in Grade 8 should move from the study of simple whole number sequence patterns to writing and manipulating expressions involving integer and fraction coefficients and graphing and solving linear equations and inequalities. The transition from Grade 4 to Grade 8 also spans the related growth from reasoning in concrete situations to reasoning in semiformal settings. As such, students should be handling questions that have moved from concrete objects to comparisons based on variable quantities (NAGB, 2006; National Mathematics Advisory Panel, 2008).

The monitoring of student growth in mathematical knowledge and skills and the capability to apply them in solving problems is critical to assuring that students are ready to move forward to successfully study mathematics at the high school level. Sustained monitoring and consistently acting on results contributes to the probability of increasing the rate at which students study mathematics beyond the required minimums. The more educators and policymakers focus on mathematics, the more likelihood that students will not miss the opportunity to see the importance of mathematics as a body of knowledge that supports their studies in other areas related to future career and vocational choices.

This paper presents one example of how extant data from two assessments—NAEP and TIMSS—can be used to construct indicators at Grades 4 and 8 that can be used to monitor student progress relative to national and international benchmarks. In doing so, the paper demonstrates how statistical linking of the NAEP data to TIMSS data allows NAEP achievement level performances of the urban districts to be interpreted in an international context by means of comparing the performance of the nations that participated at the two grade levels in TIMSS. This allows both U.S. and international policy and educational leaders to examine and compare the performance of the urban districts and national performances relative to the NAEP achievement level of Proficient or above.

Context for the Study:

Mathematics Literacy in American Society

Today, mathematics and reading are recognized as the core areas of study that students need as they move toward their adult lives (Adelman, 2004; Barton, 2006). Historically, the discussion of mathematics performance has focused on (a) how the best prepared or most advantaged students have scored on college entrance examinations or (b) the number of doctorate degrees that a nation's colleges and universities have produced in a given amount of time in the areas of science, engineering, and mathematics. These indicators are no longer sufficient nor should they be emphasized today. Although recent studies show positive effects from education, only modest effects attach themselves to children from low-income families. Education sometimes provides a boost for these students from low-income families; however, the long-term average effect reinforces differences rooted in the family backgrounds that students bring to their schooling (Haskins, 2008). As a result, mathematics achievement of low-income students on state, national, and international comparative tests consistently show that U.S. students from Grades 4–12 are not grasping important concepts and do not demonstrate expected mathematical skills and understanding (Braswell, Dion, Daane, & Jin, 2005; Mullis et al. 2004; Organization for Economic Cooperation and Development, 2004a, 2004b). Analysis of the data shows that these gaps are not just due to a lack of basic facts, but rather to a serious deficit in the student's grasp of fundamental concepts and skills at the depth required to solve problems, both in school mathematics and in everyday life. Furthermore, these gaps occur with increasing severity as students get older and in students with limited access to economic and educational resources.

Evidence also suggests that students who take curriculum sequences in high school that include Algebra I, Geometry, Algebra II, and one (or more) upper-level mathematics course are best prepared to meet the demands of college readiness. In addition, the completion of upper-level mathematics courses is associated with future success beyond just mathematics courses (ACT, 2005; Adelman, 2004). Although many question the usefulness of such courses for students who directly enter the workforce or armed forces, studies show that preparation for a well-paying, entry-level job with opportunity for advancement requires similar knowledge and skills in mathematics (Achieve, 2004; ACT, 2006; Carnevale & Desrochers, 2004; Education Trust, 1999; Meeder & Solaris, 2006;

Moss & Tilly, 2001; Wilson, 1996). The technical, quantitative reasoning, and data analysis requirements of positions in today's workplace require advanced mathematics knowledge and skills to compete, retrain for new options, and make decisions. Readiness to compete in today's workplace rests largely on being mathematically literate.

The Organization for Economic Cooperation and Development's (OECD's) Program for International Student Assessment (PISA) focuses on the development of literacy as a societal goal, rather than on the mastery of a fixed set of knowledge and procedures that would result from a specified curricular plan of study. In doing so, OECD 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" (OECD, 2003). This definition places its emphasis on capability to reflectively identify and apply mathematics in an informed and productive fashion to solve problems that are important to the individual involved. Furthermore, OECD defines *problem solving competence* as "an individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the content areas or curricular areas that might be applicable are not within a single subject area of mathematics, science, or reading" (2003). As such, OECD sees that problem solving is not solely the application of mathematics, but the blending of mathematical knowledge from other disciplines and from life experiences in confronting and solving problems.

Many students do not achieve this level of literacy in mathematics. Even students who have achieved mastery of mathematical facts and skills often are unable to solve problems that call for them to apply this knowledge in realistic settings (Mullis et al. 2004; OECD, 2004a, 2004b). The failure to integrate and connect mathematical facts, concepts, and skills in productive ways to resolve problems and reason productively in life situations leaves students unprepared to profit from postsecondary education, enter the workforce, or function productively in today's society. This lack of preparation also propels the rate at which students opt to discontinue their study of mathematics and, in some cases, drop out of school (U.S. Department of Education, 2007).

The Council of the Great City Schools also monitors the performance of urban U.S. schools. *Beating the Odds* analyzes the achievement growth in mathematics among districts in the Council. The report asks two critical questions: "Are urban schools improving academically?" and "Are urban schools closing achievement gaps?" (Snipes, Williams, Horwitz, Soga, & Casserly, 2007).

Data from the 2007 TUDA and the 2003 TIMSS provide a window through which these and similar questions can be viewed. TUDA focuses on the achievement of central city urban students in the U.S. relative to the mathematics framework established by the NAGB (NAGB, 2006). The 2007 TUDA examined the knowledge and skills possessed by students in Grades 4 and 8 in 11 U.S. urban districts and jurisdictions: Atlanta City School District, GA; Austin Independent School District, TX; Charlotte-Mecklenburg Schools, NC; Boston School District, MA; City of Chicago School District 299, IL; Cleveland Municipal School District, OH; Houston Independent School District, TX; Los Angeles Unified School District, CA; New York City Public Schools, NY; San Diego Unified School District, CA; and the District of Columbia Public Schools, Washington, DC.

Given today's citizenship and workplace realities, all students need increasingly high levels of mathematics knowledge and skills to succeed in whatever their chosen life path. Moreover, the increased role of technology and science in everyday life requires students graduating from high school to have a broad understanding of the collection, analysis, and interpretation of data in order to participate effectively in civic life and on the job.

Major Findings

The comparisons of percentages of students at or above the Proficient level are shown in Figure 1– Figure 12 and numerical comparisons are presented in Table 28 and Table 29 in Technical Appendix A.

Figures 1–12 graphically display the international benchmarks for the United States and the 11 U.S. districts in relation to the performances of their international peers participating in the TIMSS. These contrasts are presented with the U.S. overall NAEP 2007 comparison, with the TIMSS countries for Grades 4 and 8 presented first (Figure 1). In each grade, the percentage of students in the United States at or above the Proficient level in mathematics is compared to the percentages obtained by students in each of the nations participating in the 2003 TIMSS study of mathematics. In Figure 2–Figure 12, these same comparisons are made for each of the 11 urban districts. These U.S.-by-nation and district-by-nation comparisons are made possible by the NAEP–TIMSS linking study (Technical Appendix A).

A Note About Understanding the Figures

Each figure includes two graphs: The upper graph shows the results for students in Grade 4, and the lower graph shows the results for students in Grade 8. The figures present the countries in descending order, from left to right, of the percentages of students reaching the U.S. Proficient standard projected onto the TIMSS scale. The figures also indicate whether a nation’s percentage is statistically above (the taller blue bars to the left), statistically similar to (the white bars in the middle), or statistically below (the shorter blue bars on the right) that of the United States or the 11 urban school districts. The bar indicating the U.S. or district percentage is colored gray.

International Benchmarks for the United States in Grade 4 and Grade 8 Mathematics[1]

Figure 1 depicts the results for the comparison of the percentages of U.S. students at or above the Proficient level on the NAEP mathematics scale with the percentages associated with Proficient or above in the other countries. The graphs indicate which nations' percentages are statistically above, similar to, and statistically below that of the United States.

Per Figure 1, five nations, as indicated by the taller black bars, are judged to have performed significantly higher in terms of the percentage of Grade 4 students reaching or exceeding the Proficient level of achievement in mathematics.

1. Singapore
2. Hong Kong SAR
3. Chinese Taipei
4. Japan
5. Belgium (Flemish)

Six nations are judged to have performed similarly to (i.e., not statistically different from) the United States in terms of the percentage of Grade 4 students reaching or exceeding the Proficient level of achievement in mathematics.

1. England
2. Latvia
3. Lithuania
4. Netherlands
5. Russian Federation
6. Hungary

Thirteen nations are judged to have performed statistically lower than the United States in terms of the percentage of Grade 4 students reaching or exceeding the Proficient level of achievement in mathematics.

1. Cyprus
2. Republic of Moldova
3. Italy
4. Australia
5. New Zealand
6. Scotland
7. Slovenia
8. Armenia
9. Norway
10. Philippines
11. Islamic Republic of Iran
12. Tunisia
13. Morocco

As one can see, some of these latter nations have percentages of students reaching the Proficient or above criterion in the single digit range.

The analysis for international benchmarks for the comparison of Grade 8 mathematics percentages follows in a similar fashion. This is shown graphically in the second graph in Figure 1.

Eight nations are judged to have performed significantly higher in terms of the percentage of Grade 8 students reaching or exceeding the Proficient level of achievement in mathematics.

[1] Note that some of the projected percentages (at and above the Proficient level) for each nation in TIMSS in this report are slightly different from those reported by Phillips (2007a). This is because the projected achievement levels in the Phillips (2007a) report were based on a linking study that compared 2000 NAEP with 1999 TIMSS in Grade 8 mathematics and science. This current report uses the latest data available and is based on a more recent linking study using 2003 NAEP and 2003 TIMSS in Grades 4 and 8 mathematics. The more recent linking study is described in Technical Appendix A. The results of the two studies for Grade 8 mathematics are almost identical; the minor differences are displayed in Table 17.

1. Singapore
2. Hong Kong SAR
3. Republic of Korea
4. Chinese Taipei
5. Japan
6. Belgium (Flemish)
7. Netherlands
8. Hungary
9. Armenia
10. Italy
11. Bulgaria
12. Republic of Moldova
13. Cyprus
14. Norway
15. Republic of Macedonia
16. Jordan
17. Egypt
18. Indonesia
19. Palestinian National Authority
20. Lebanon
21. Islamic Republic of Iran
22. Chile
23. Bahrain
24. Philippines
25. Tunisia
26. Morocco
27. Botswana
28. South Africa
29. Saudi Arabia
30. Ghana

Six nations are judged to have performed similarly to (i.e., not statistically different from) the United States in terms of the percentage of Grade 8 students reaching or exceeding the Proficient level of achievement in mathematics.

1. Estonia
2. Slovak Republic
3. Australia
4. Russian Federation
5. Malaysia
6. Latvia

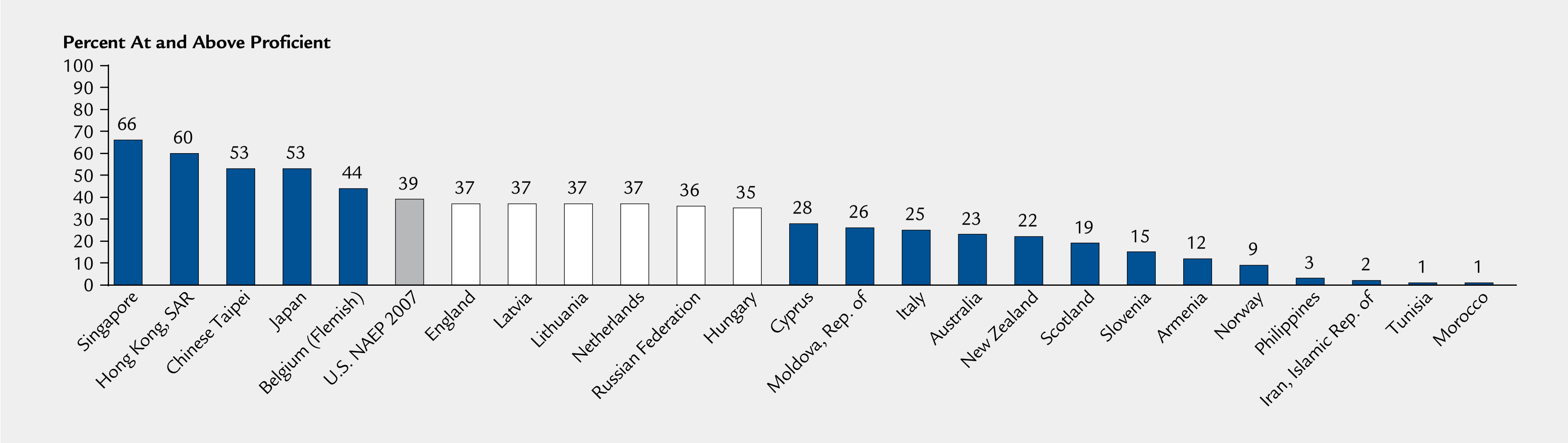
Thirty nations are judged to have performed statistically lower than the United States in terms of the percentage of Grade 8 students reaching or exceeding the Proficient level of achievement in mathematics.

1. Lithuania
2. Israel
3. Scotland
4. New Zealand
5. Sweden
6. Serbia
7. Slovenia
8. Romania

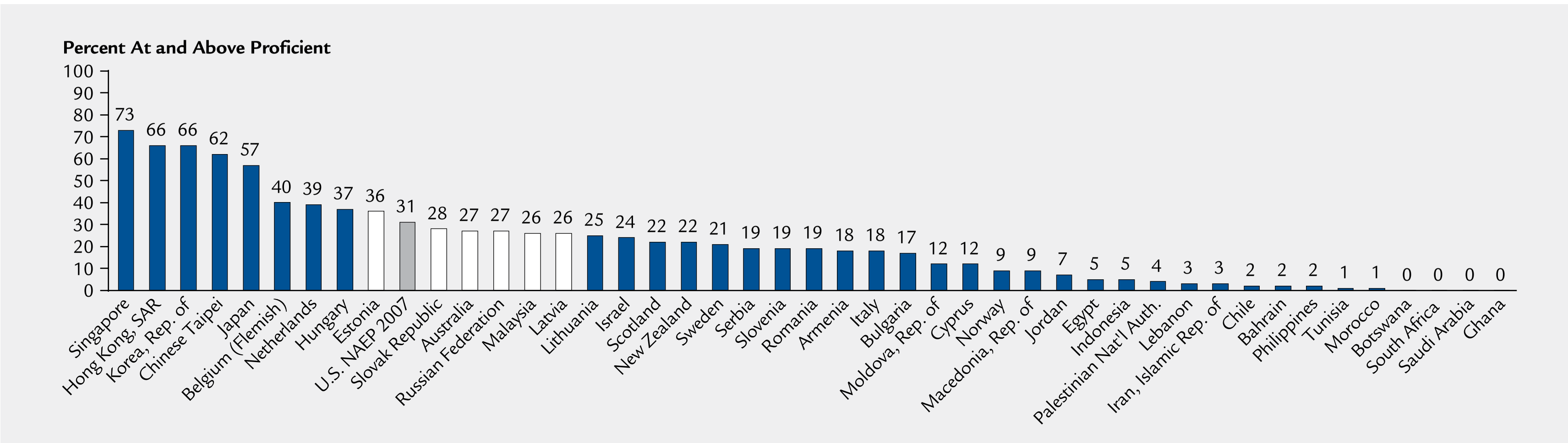
The low performance of many of these nations is indicated by single-digit percentages of students reaching the level of Proficient or above. Many of these nations are developing countries with meager resources. For many, taking part in TIMSS 2003 was the nation's first participation in an international large-scale assessment. In four of the countries, less than one-half of 1% of students met the criterion of Proficient or above (these were rounded off to zero in the graphs).

FIGURE 1: United States

Comparisons between 2007 grade 4 NAEP in mathematics for the United States and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



Comparisons between 2007 grade 8 NAEP in mathematics for the United States and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

International Benchmarks for 11 U.S. Urban Cities in Grade 4 and Grade 8 Mathematics

The mathematical proficiency of students in the 11 urban U.S. school districts can be compared with the proficiency of students in other nations that participated in TIMSS 2003. The results of the study are graphically presented in Figure 2–Figure 12. To illustrate how to read the figures, we will use Atlanta. In Figure 2, the percentages of students in Grades 4 and 8 in Atlanta who are achieving at or above the Proficient level are contrasted with the percentages of international students meeting the same criteria.

Figure 2 contains two graphs. The first graph displays the results for Grade 4 mathematics in Atlanta compared with Grade 4 mathematics in each nation participating in TIMSS 2003. The second graph displays similar comparative data for Grade 8. For each district and nation, the graph displays the rounded percentage of students estimated to be at or above the Proficient level. The nations have been rank-ordered, with the highest performing nations in terms of percentages on the left and the lowest performing nations on the right. Embedded within the graph is a lightly shaded bar for Atlanta and the percentage at which students in Atlanta are performing at and above the Proficient level.

Illustration of International Benchmarks Using the Atlanta District as an Example: Grade 4 Mathematics

Per Table 28 and Figure 2, 14 nations, including the United States, have a significantly greater percentage of their Grade 4 students achieving the Proficient level for mathematics than students in Atlanta.

1. Singapore
2. Hong Kong SAR
3. Chinese Taipei
4. Japan
5. Belgium (Flemish)
6. U.S. NAEP 2007
7. England
8. Latvia
9. Lithuania
10. Netherlands
11. Russian Federation
12. Hungary
13. Cyprus
14. Republic of Moldova

In six nations, the percentage of Grade 4 students achieving at or above the Proficient level in mathematics is similar to the percentage of students observed in Atlanta.

1. Italy
2. Australia
3. New Zealand
4. Scotland
5. Slovenia
6. Armenia

Five nations have a significantly lower percentage of their Grade 4 students achieving at the Proficient level for mathematics than students in Atlanta.

1. Norway
2. Philippines
3. Islamic Republic of Iran
4. Tunisia
5. Morocco

Illustration of International Benchmarks Using the Atlanta District as an Example: Grade 8 Mathematics

Per Table 29 and Figure 2, 25 nations, including the United States, have a significantly greater percentage of their Grade 8 students achieving the Proficient level for mathematics than students in Atlanta.

1. Singapore
2. Hong Kong SAR
3. Republic of Korea
4. Chinese Taipei
5. Japan
6. Belgium (Flemish)
7. Netherlands
8. Hungary
9. Estonia
10. U.S. NAEP 2007
11. Slovak Republic
12. Australia
13. Russian Federation
14. Malaysia
15. Latvia
16. Lithuania
17. Israel
18. Scotland
19. New Zealand
20. Sweden
21. Serbia
22. Slovenia
23. Romania
24. Armenia
25. Italy

In six nations, the percentage of Grade 8 students achieving at or above the Proficient level is similar to the percentage of students observed in Atlanta.

1. Bulgaria
2. Republic of Moldova
3. Cyprus
4. Norway
5. Republic of Macedonia
6. Jordan

Fourteen nations have a significantly lower percentage of their Grade 8 students achieving at the Proficient level for mathematics than students in Atlanta.

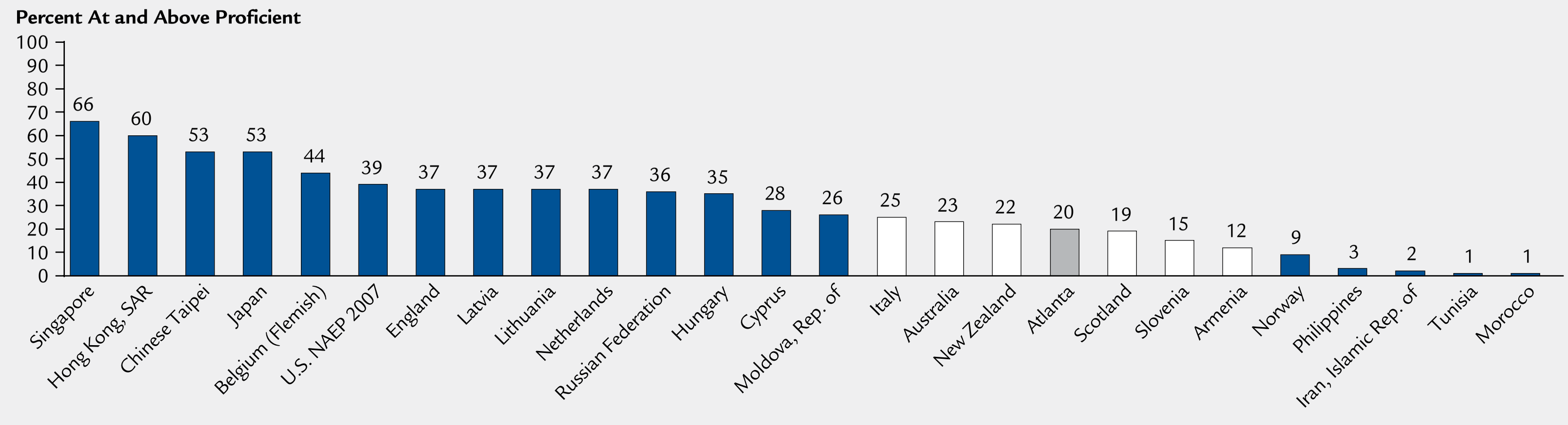
1. Egypt
2. Indonesia
3. Palestinian National Authority
4. Lebanon
5. Islamic Republic of Iran
6. Chile
7. Bahrain
8. Philippines
9. Tunisia
10. Morocco
11. Botswana
12. South Africa
13. Saudi Arabia
14. Ghana

Of note, the mathematics results for Atlanta in Figure 2 are 2007 TUDA results from the publicly available data at *http://www.nces.ed.gov* and in a report by Lutkus et al. (2007). The U.S. national percentages in the tables and figures (those labeled U.S. NAEP 2007) refer to the performance of U.S. public school students on the 2007 NAEP examination (Lee, Grigg, & Dion, 2007), which is available on NCES's Web site *(http://www.nces.ed.gov)*. The 2007 U.S. NAEP percentage of Grade 4 public school students meeting or exceeding the Proficient level was 39%, and the percentage of U.S. Grade 8 public school students meeting or exceeding the Proficient level was 31%. International data related to TIMSS 2003 is available at the international study's Web site *(http://timss.bc.edu)* and through the international report (Mullis et al. 2004).

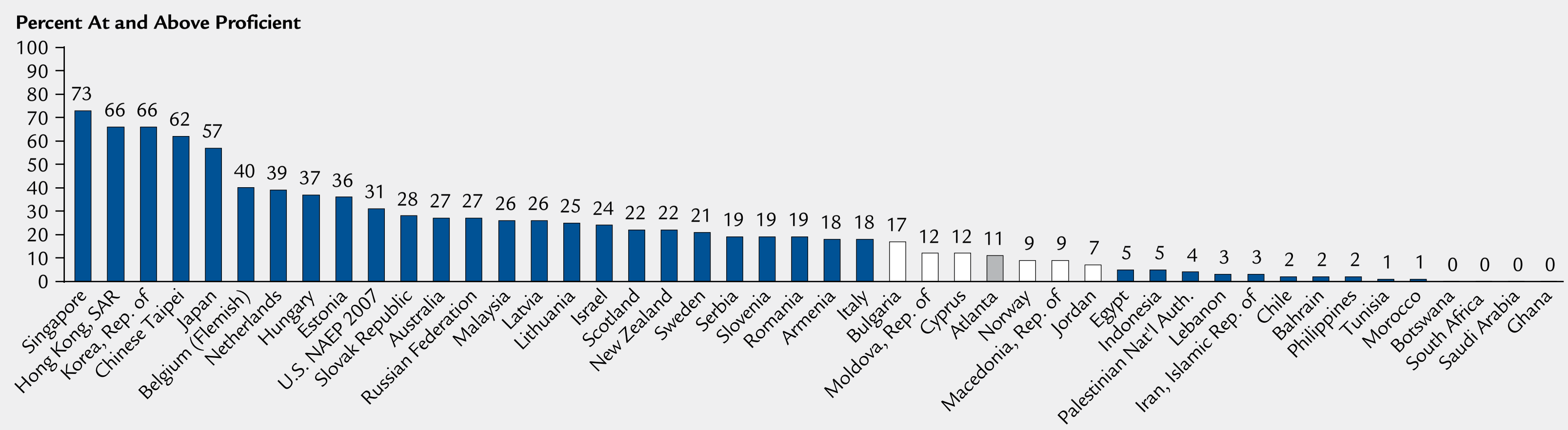
The analyses for Atlanta can be repeated for each of the remaining 10 urban districts that participated in TUDA. Each comparison tells a different story. Which countries are important as international benchmarks for one urban district may be different for each urban district.

FIGURE 2: Atlanta

Comparisons between 2007 grade 4 NAEP in mathematics for Atlanta and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



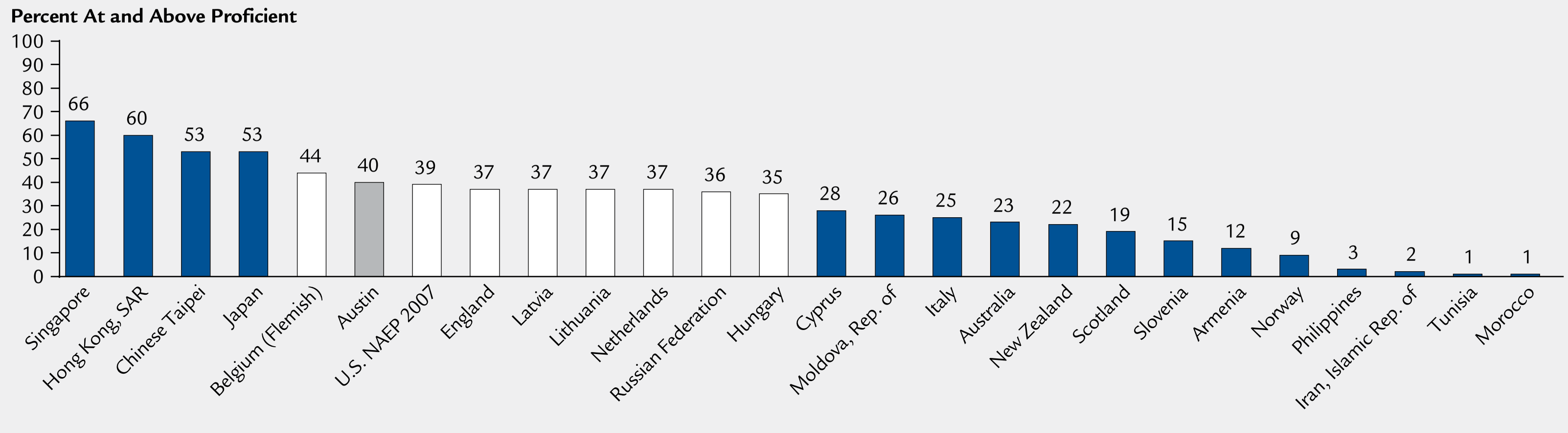
Comparisons between 2007 grade 8 NAEP in mathematics for Atlanta and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



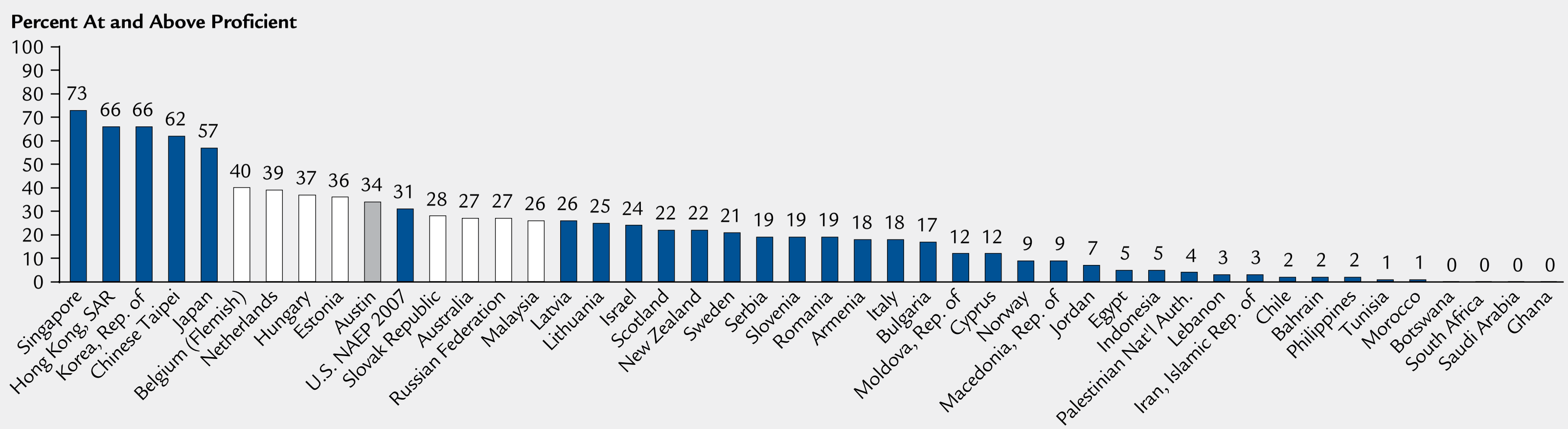
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

FIGURE 3: Austin

Comparisons between 2007 grade 4 NAEP in mathematics for Austin and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



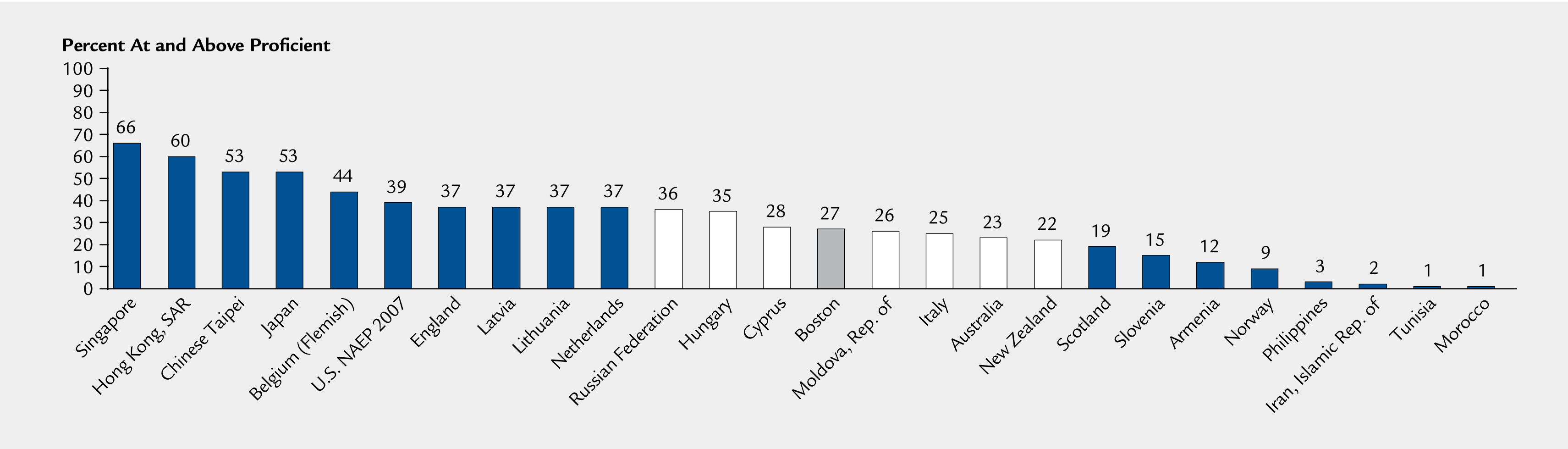
Comparisons between 2007 grade 8 NAEP in mathematics for Austin and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



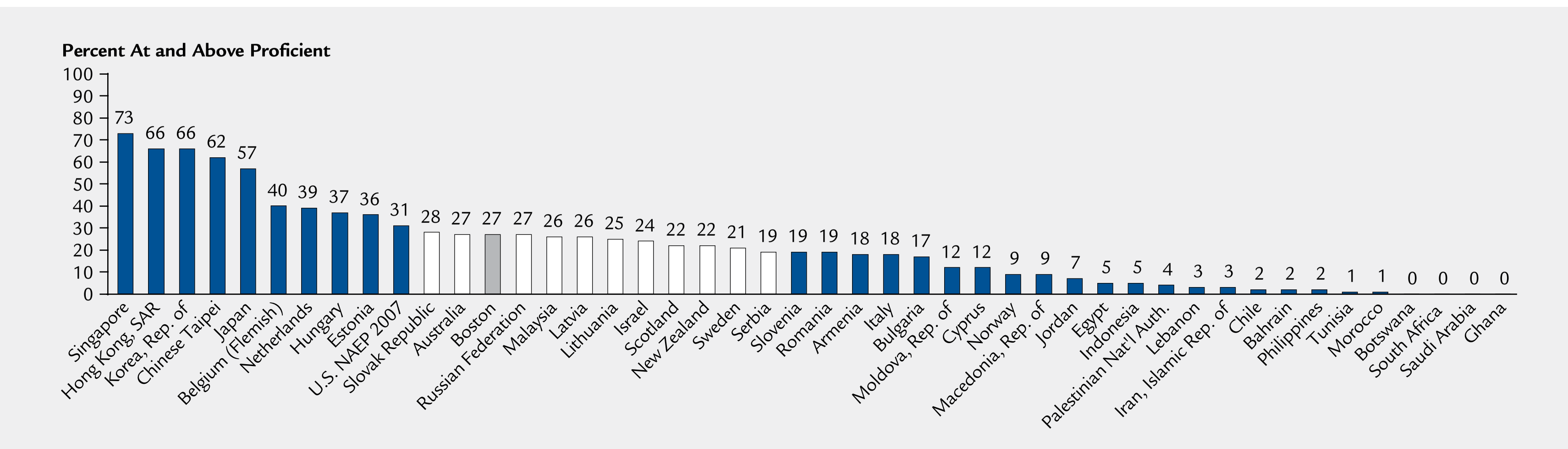
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 4: Boston

Comparisons between 2007 grade 4 NAEP in mathematics for Boston and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



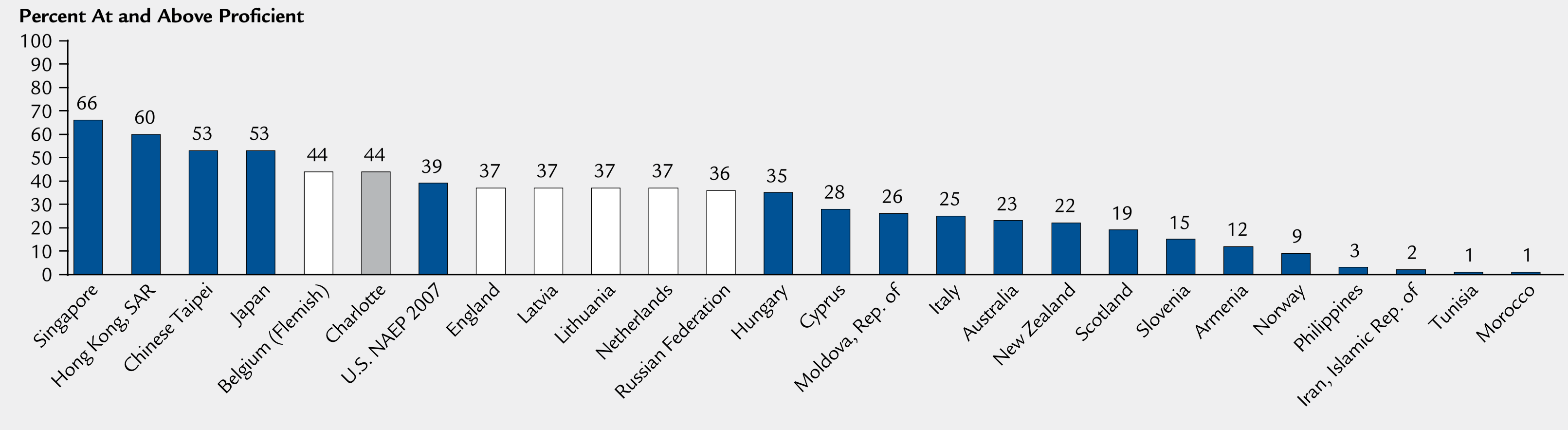
Comparisons between 2007 grade 8 NAEP in mathematics for Boston and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



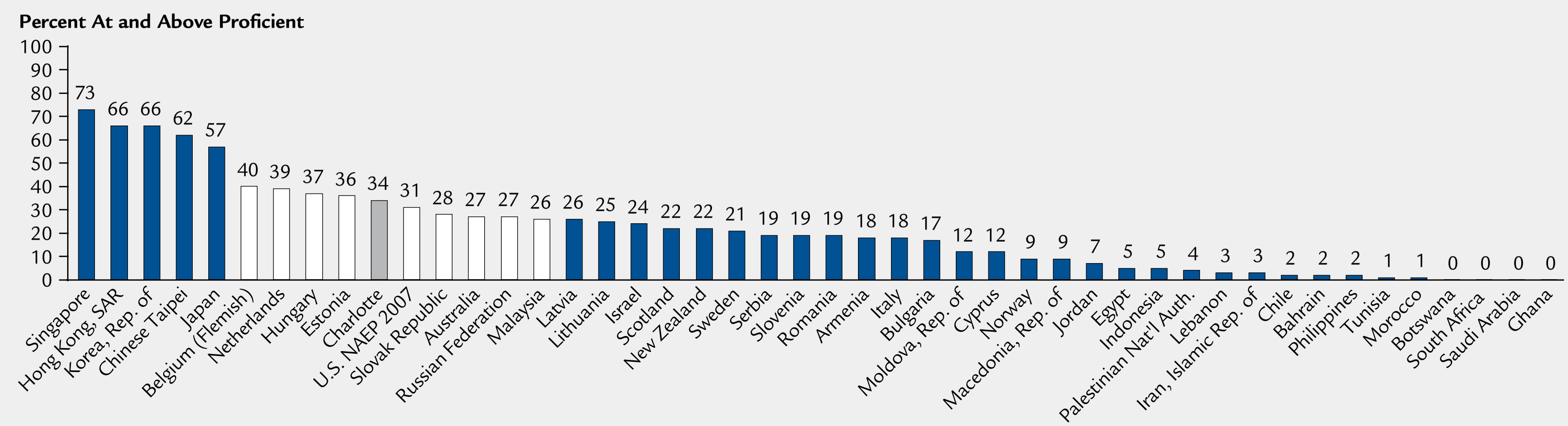
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 5: Charlotte

Comparisons between 2007 grade 4 NAEP in mathematics for Charlotte and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



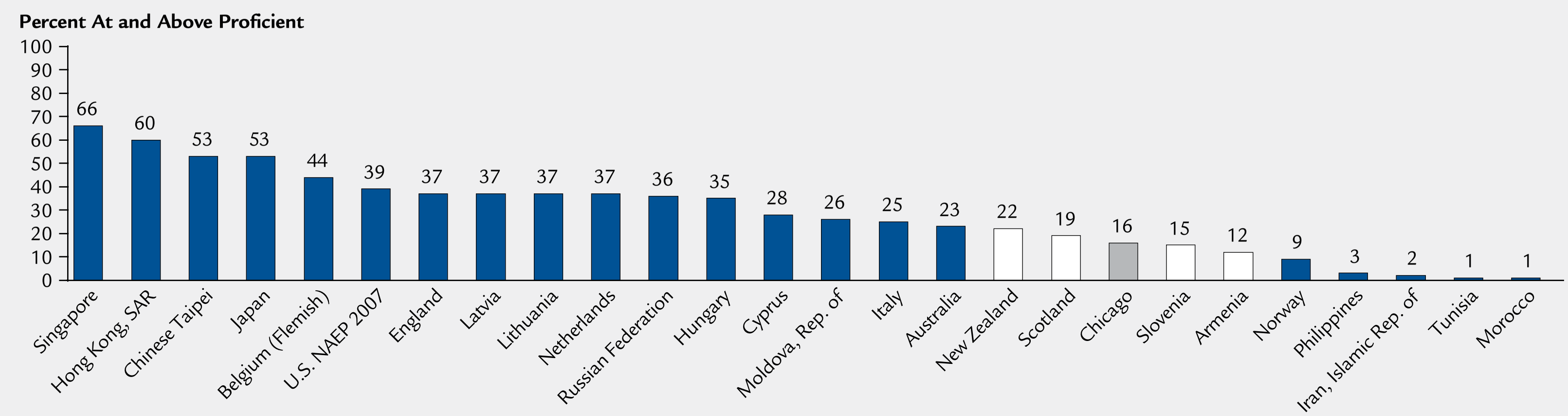
Comparisons between 2007 grade 8 NAEP in mathematics for Charlotte and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



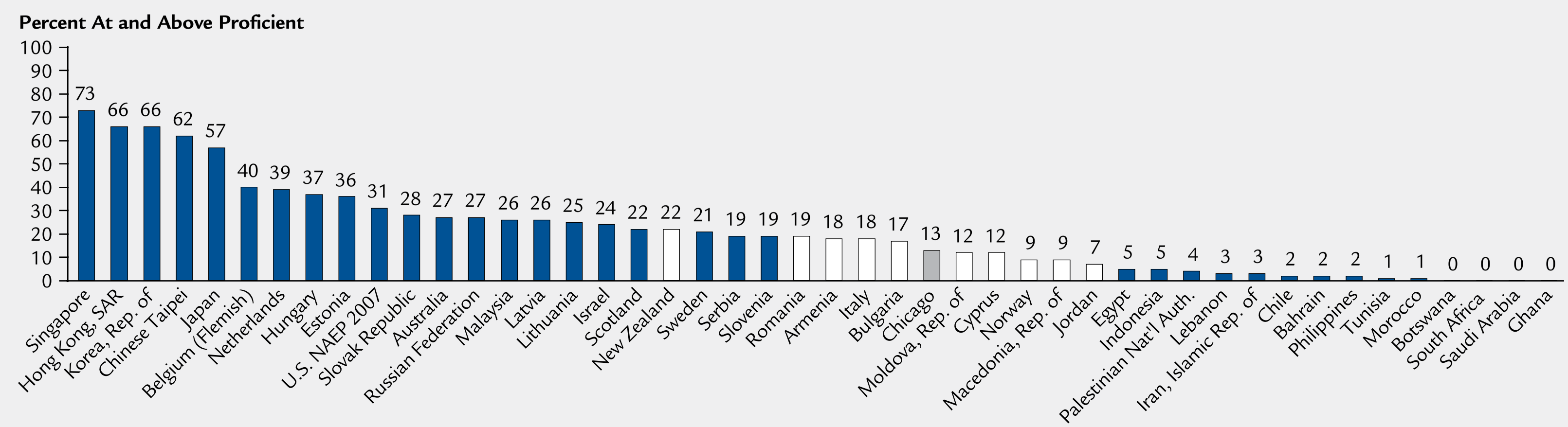
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 6: Chicago

Comparisons between 2007 grade 4 NAEP in mathematics for Chicago and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



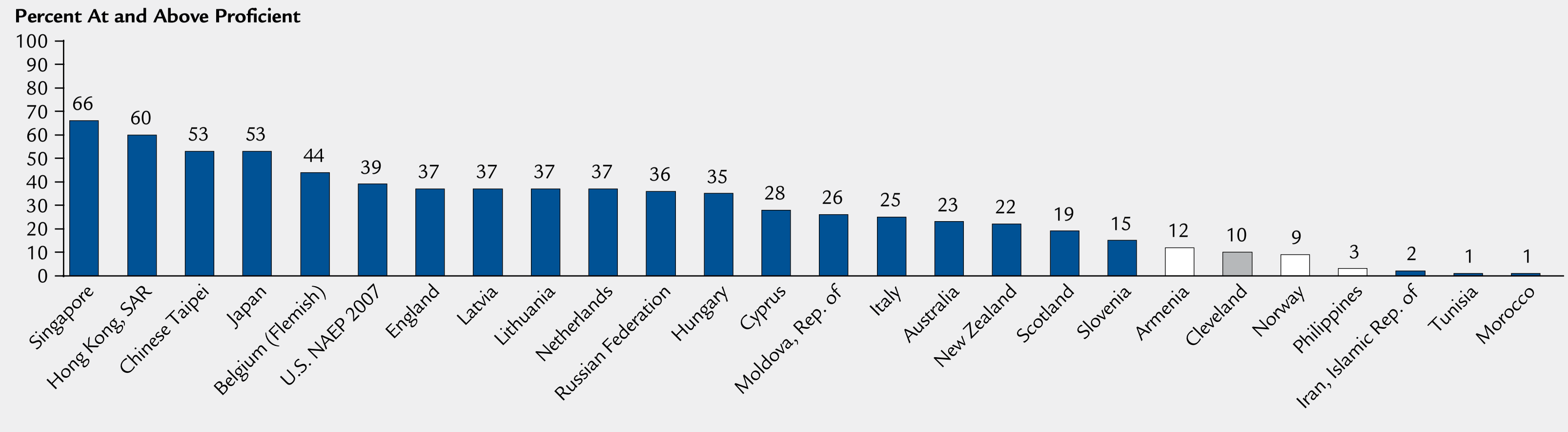
Comparisons between 2007 grade 8 NAEP in mathematics for Chicago and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



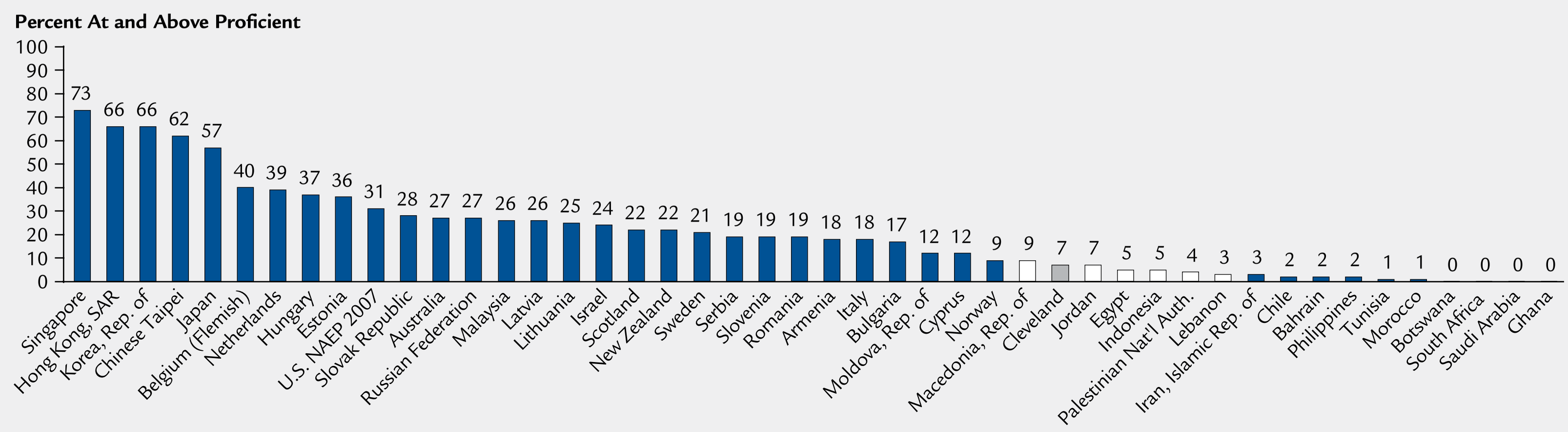
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 7: Cleveland

Comparisons between 2007 grade 4 NAEP in mathematics for Cleveland and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



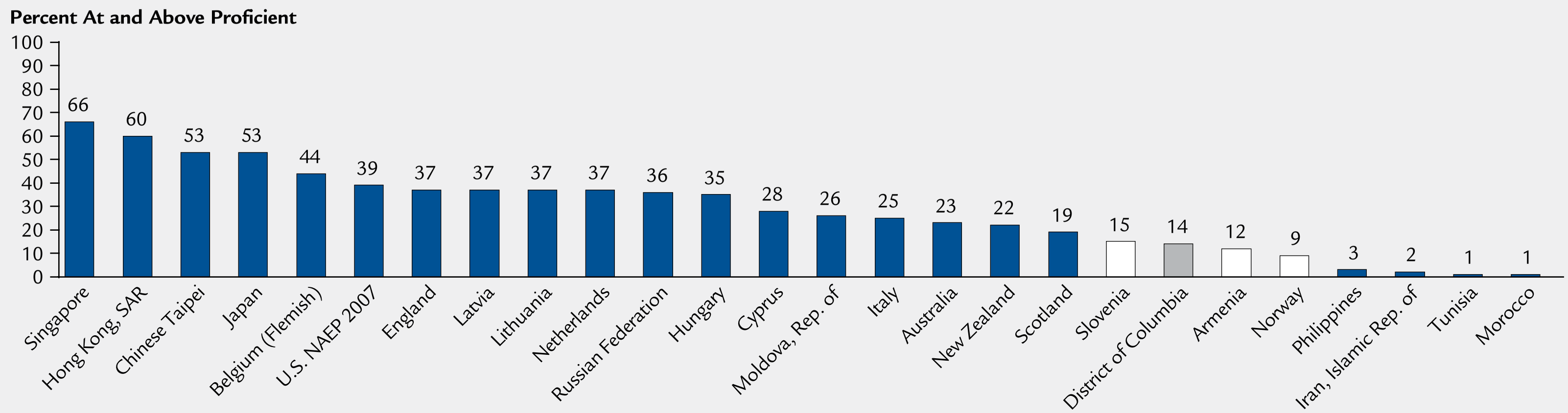
Comparisons between 2007 grade 8 NAEP in mathematics for Cleveland and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



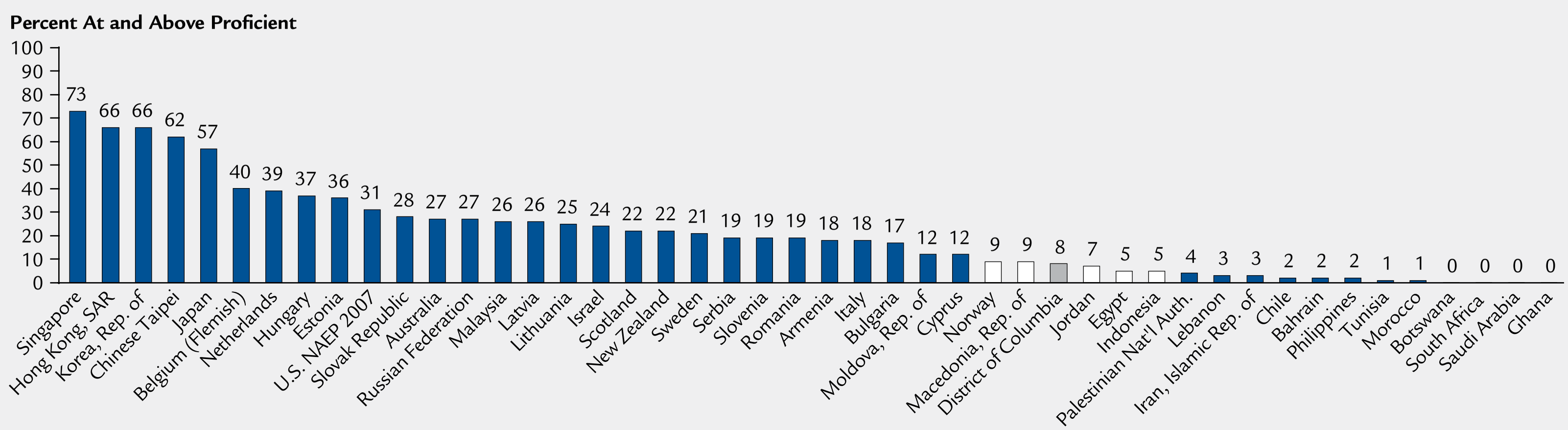
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 8: District of Columbia

Comparisons between 2007 grade 4 NAEP in mathematics for District of Columbiaand grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



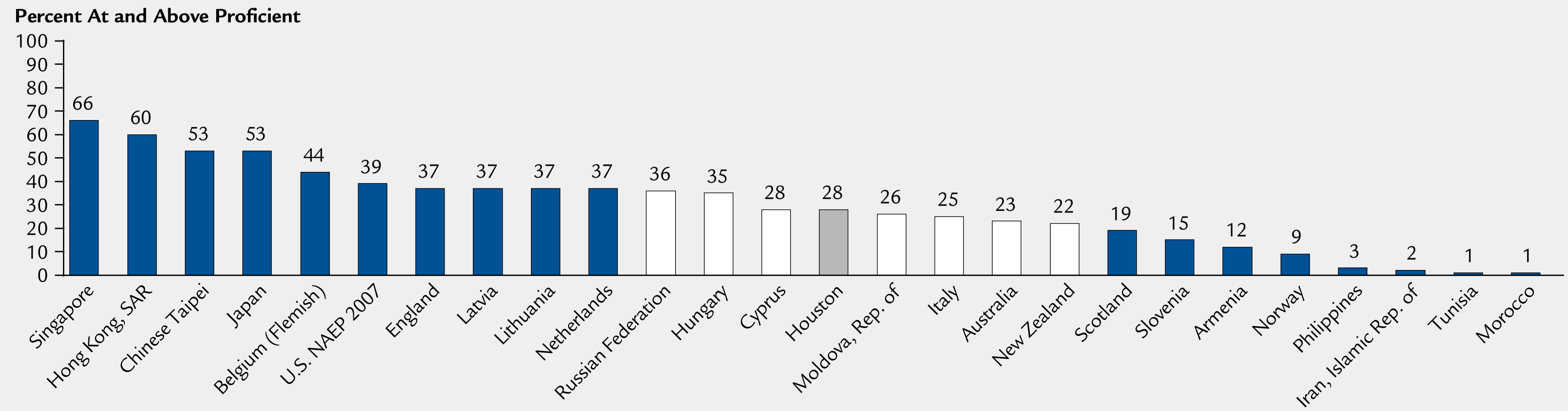
Comparisons between 2007 grade 8 NAEP in mathematics for District of Columbiaand grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



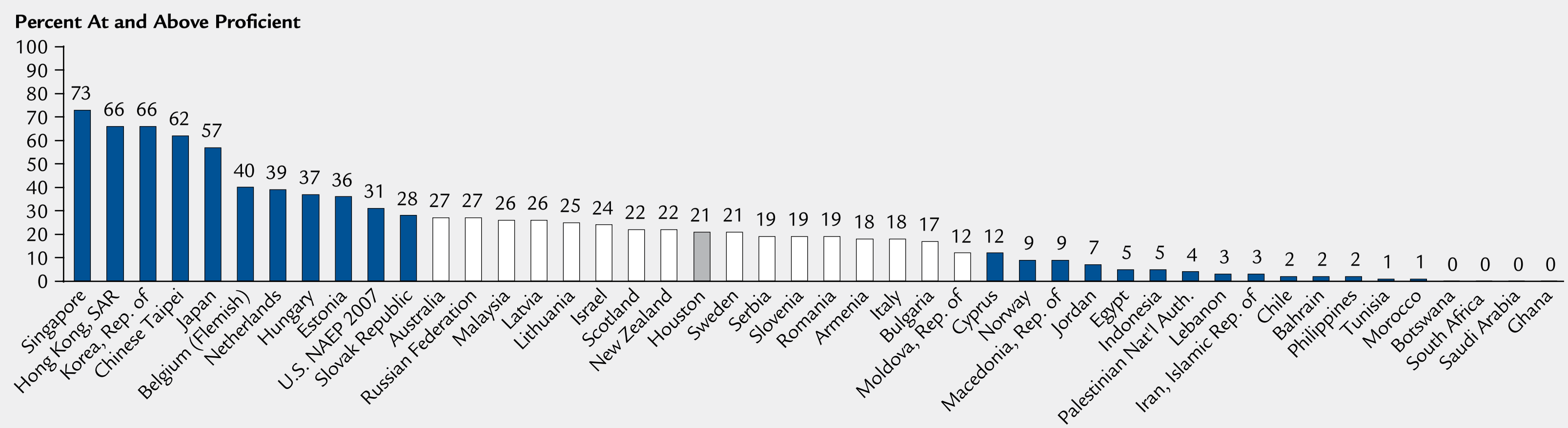
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 9: Houston

Comparisons between 2007 grade 4 NAEP in mathematics for Houston and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



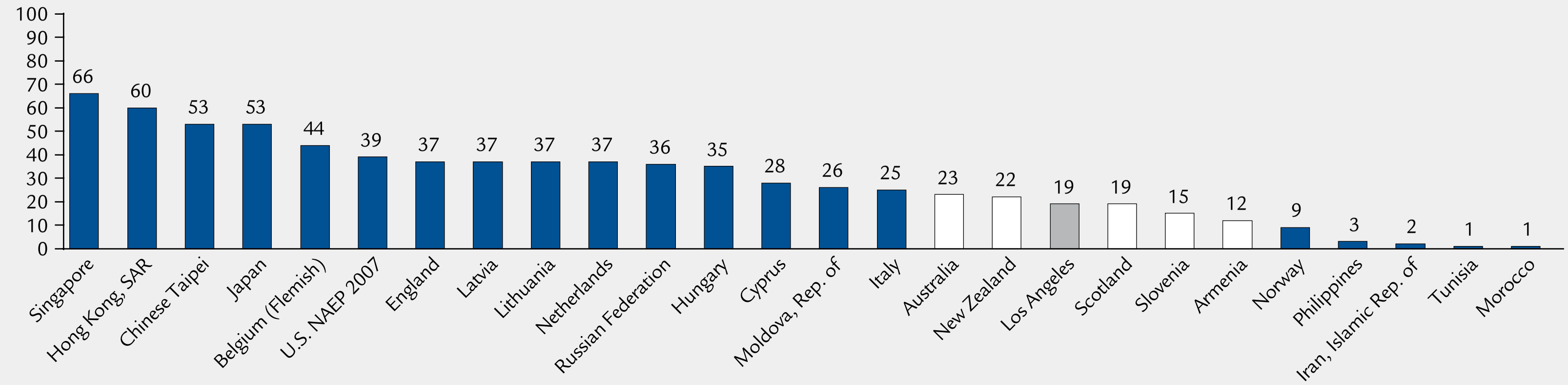
Comparisons between 2007 grade 8 NAEP in mathematics for Houston and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



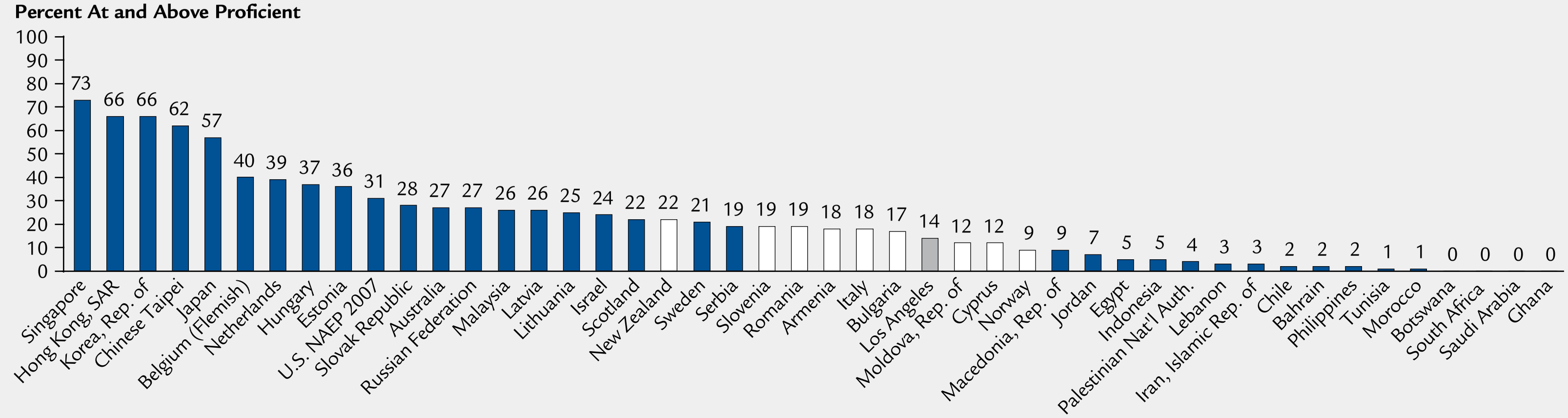
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007

Figure 10: Los Angeles

Comparisons between 2007 grade 4 NAEP in mathematics for Los Angeles and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



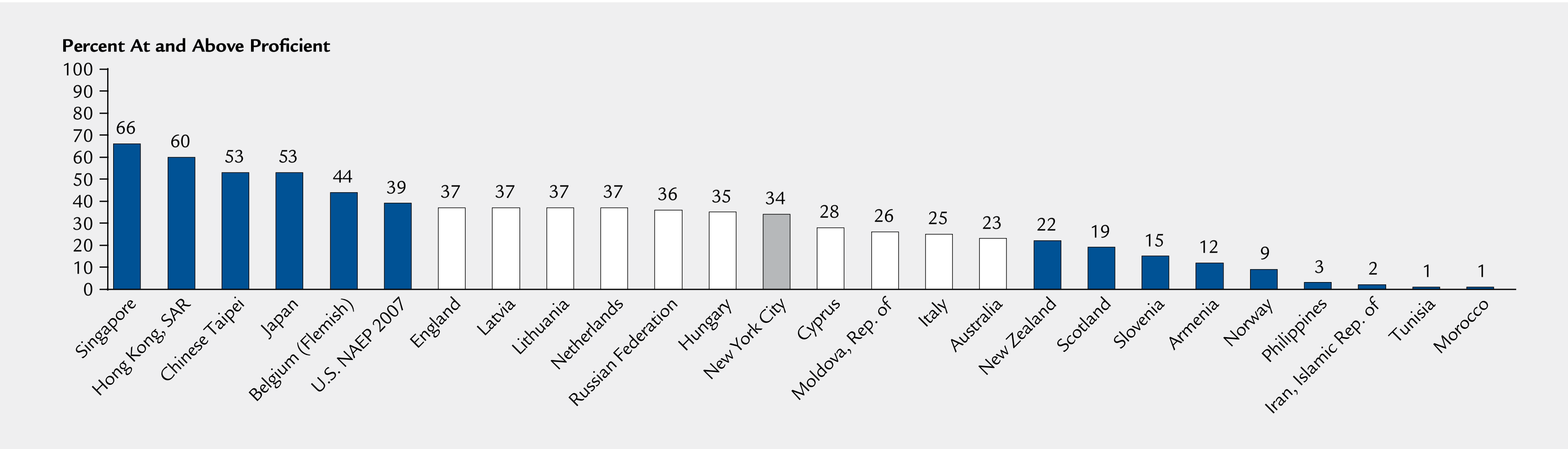
Comparisons between 2007 grade 8 NAEP in mathematics for Los Angeles and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



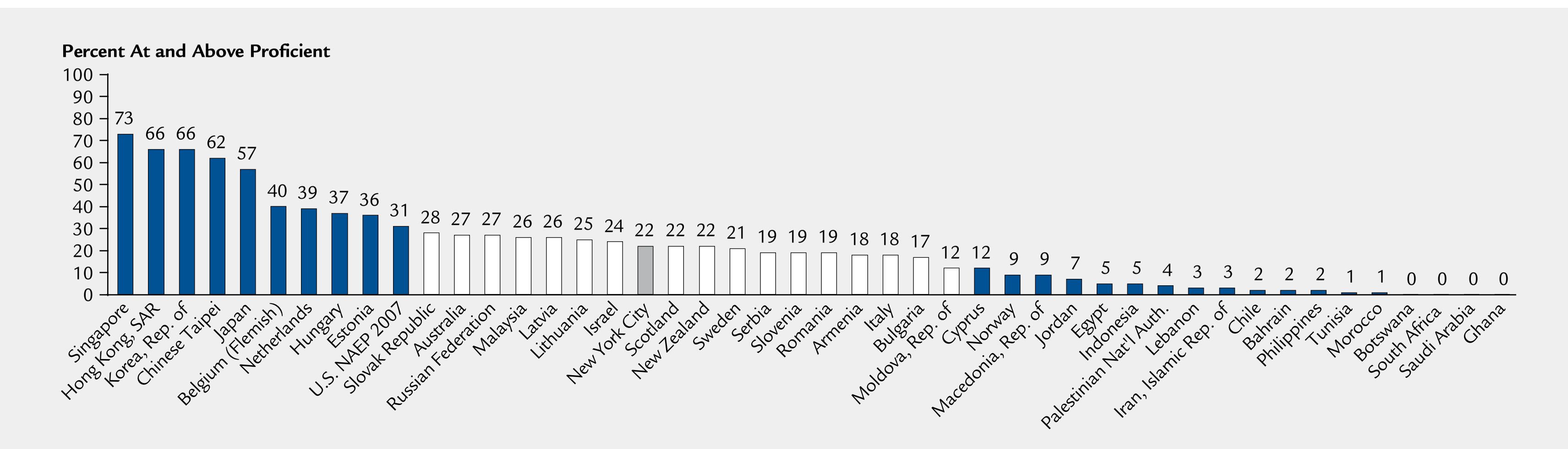
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 11: New York City

Comparisons between 2007 grade 4 NAEP in mathematics for New York and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



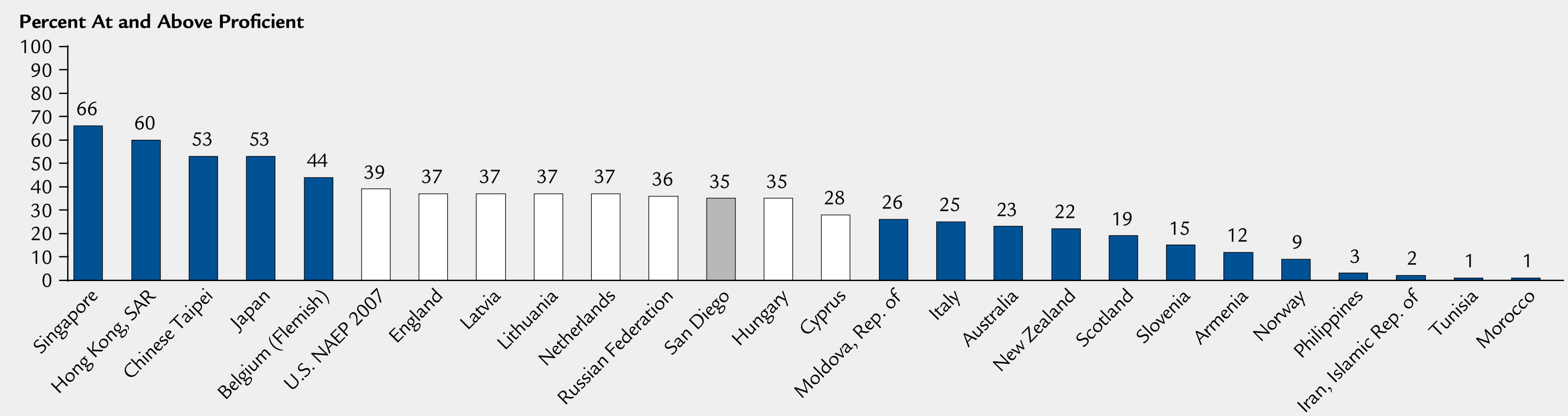
Comparisons between 2007 grade 8 NAEP in mathematics for New York and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



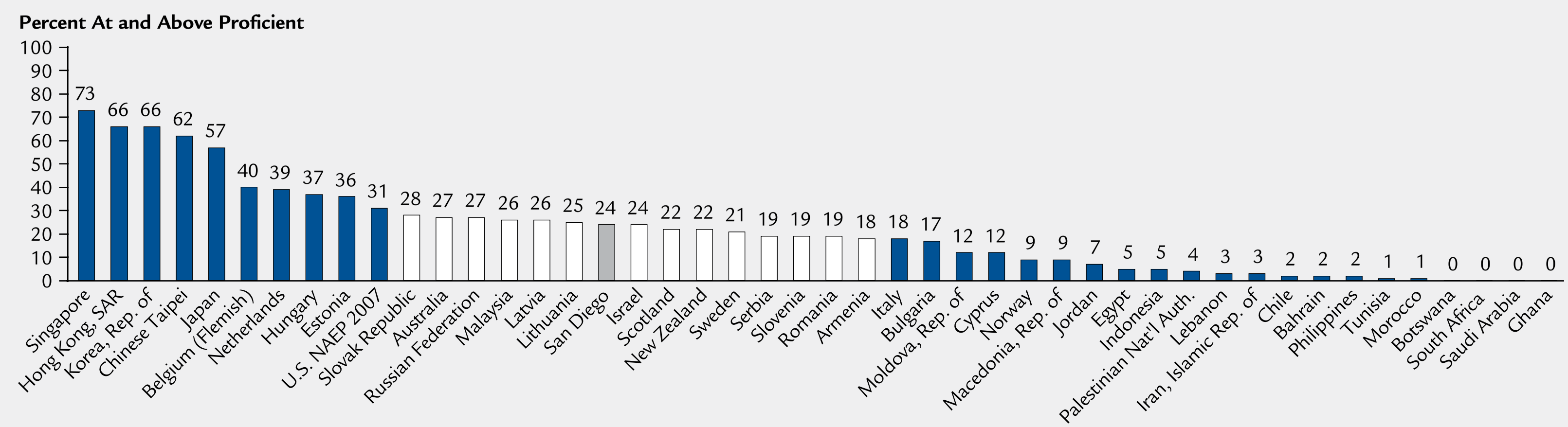
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 12: San Diego

Comparisons between 2007 grade 4 NAEP in mathematics for San Diego and grade 4 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



Comparisons between 2007 grade 8 NAEP in mathematics for San Diego and grade 8 2003 TIMSS international mathematics results for the percent at and above proficient based on NAEP achievement levels projected onto the TIMSS scale



Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Comparison of Each District With the Overall International Average

In addition to comparing each school district in NAEP to each country in TIMSS, each school district can also be compared with the overall TIMSS average. This is done in the first graphs in Figure 13 and Figure 14.

For Grade 4, four districts (Charlotte, Austin, San Diego, and New York City) performed significantly better than the TIMSS overall average of 27% proficient.[2] Five districts (Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed below the international average. The remaining two districts (Houston and Boston) performed statistically similar to the overall TIMSS international average.

For Grade 8, three school districts (Charlotte, Austin, and Boston) performed above the international TIMSS average of 21%. As with Grade 4, the five school districts (Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed below the international average. The remaining three districts (San Diego, New York City, and Houston) performed comparably to the TIMSS international average.

Comparison of Each District With the OECD International Average

The TIMSS overall international average is based on the aggregate performance of all 24 countries in Grade 4 and all 44 countries in Grade 8 that participated in TIMSS. (As noted previously, the United States was excluded from these averages.) Although the *overall* international average is a useful benchmark, an even more important benchmark is the international average of all countries in the OECD that participated in the 2003 TIMSS. This is a set of more industrialized countries that are more likely to be economic competitors of the United States. For Grade 4, the 2003 TIMSS included 10 OECD countries (in addition to the United States): Japan, Belgium (Flemish), Netherlands, England, Hungary, Italy, Australia, New Zealand, Scotland, and Norway. For Grade 8, the 2003 TIMSS included 12 OECD countries (in addition to the United States): Republic of Korea, Japan, Belgium (Flemish), Netherlands, Hungary, Slovak Republic, Australia, Sweden, Scotland, New Zealand, Italy, and Norway.[3]

The OECD average in this report is based on the aggregate performance of all 10 OECD countries in Grade 4 and all 12 OECD countries in Grade 8 that participated in the 2003 TIMSS. (Again, the United States was excluded from these averages.) The results are presented in the second graphs in Figure 13 and Figure 14. As shown, the OECD average is higher than the overall TIMSS average in both grades. Consequently, the OECD average is a more challenging international benchmark against which to compare the school districts that participated in the 2007 NAEP.

For Grade 4, three districts (Charlotte, Austin, and San Diego) performed significantly better than the international OECD average of 30% Proficient. Five districts (Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed below the OECD average. The remaining three districts (New York City, Houston, and Boston) performed statistically similar to the OECD average.

For Grade 8, zero school districts performed significantly above the international OECD average of 33% Proficient. Nine districts (Boston, San Diego, New York

---

[2] The overall international average for TIMSS and the international average for participating OECD countries were determined by excluding the United States from the average. For that reason, the international averages reported here may be different from those in the 2003 TIMSS report. To be consistent with the international report, the international averages are unweighted. Significance tests are based on simple comparisons using z-tests with 95% confidence intervals.

[3] At Grade 8, England participated in the 2003 TIMSS but did not meet the minimum sampling requirements for this study. Thus, England is excluded from this report.

City, Houston, Atlanta, Los Angeles, Chicago, District of Columbia, and Cleveland) performed significantly below the OECD average. The remaining two school districts (Charlotte and Austin) performed comparably to the OECD average.

Comparison of Public School Students From Large Central Cities With the Overall International Average

The 2007 NAEP provided estimates of the average performance of U.S. students from large central cities (with populations greater than 250,000)[4]—an approach that provided a fair national average against which to compare urban school districts. Linking the NAEP to TIMSS, however, allows the mathematical proficiency of students in large central cities in the United States to be compared with students in other countries. For Grade 4, the large central city average of 28% Proficient is statistically comparable to the overall international average of 27% in 24 countries (Figure 13). Similarly, for Grade 8, the large central city average of 22% Proficient is comparable to the overall international average of 21% in 44 countries (Figure 14).

Comparison of Public School Students From Large Central Cities With the OECD International Average

U.S. student performance in large central city public schools can also be compared to the OECD international average. For Grade 4, the large central city average of 28% Proficient is statistically comparable to the OECD international average of 30% Proficient in 10 countries (Figure 13). However, for Grade 8, the large central city average of 22% Proficient is significantly lower than the OECD international average of 33% Proficient in 12 countries (Figure 14).

[4] The authors would like to thank John Mazzeo and Andreas Oranje at the Educational Testing Service for providing the standard errors associated with the performance of U.S. students in the large central cities.

Criterion-Referenced Interpretations

All of the previously discussed results are essentially *norm-referenced* interpretations of the performance of students in U.S. urban districts and other countries. The comparison of the percentages reaching the Proficient level or above is informative and helps inform the comparisons among nations, states, and districts. But, in the end, the percentages do not tell how well the nations, states, and districts are really doing compared with an absolute standard (Barton & Cooley, 2008). Furthermore, the percentages do not tell what level of mathematical understanding students at Grades 4 and 8 have and do not have. For example, the overall percentage of U.S. students reaching the Proficient level in mathematics in 2007 was 39% for Grade 4 and 31% for Grade 8. But how good is this? Is it good enough?

One *criterion-referenced* strategy for answering these questions is to examine the achievement level associated with a national, state, or district average. This strategy can help provide answers to the following question: Is the average student in a nation, state, or urban district at least Proficient in mathematics at Grade 4 or at Grade 8 or is the average student achieving at a Basic or Below Basic level?

The criterion-referenced description of what it means for the average student to be Proficient, Basic, or Below Basic at these respective grade levels can be obtained from the definitions of the NAGB's achievement levels for mathematics (2006).

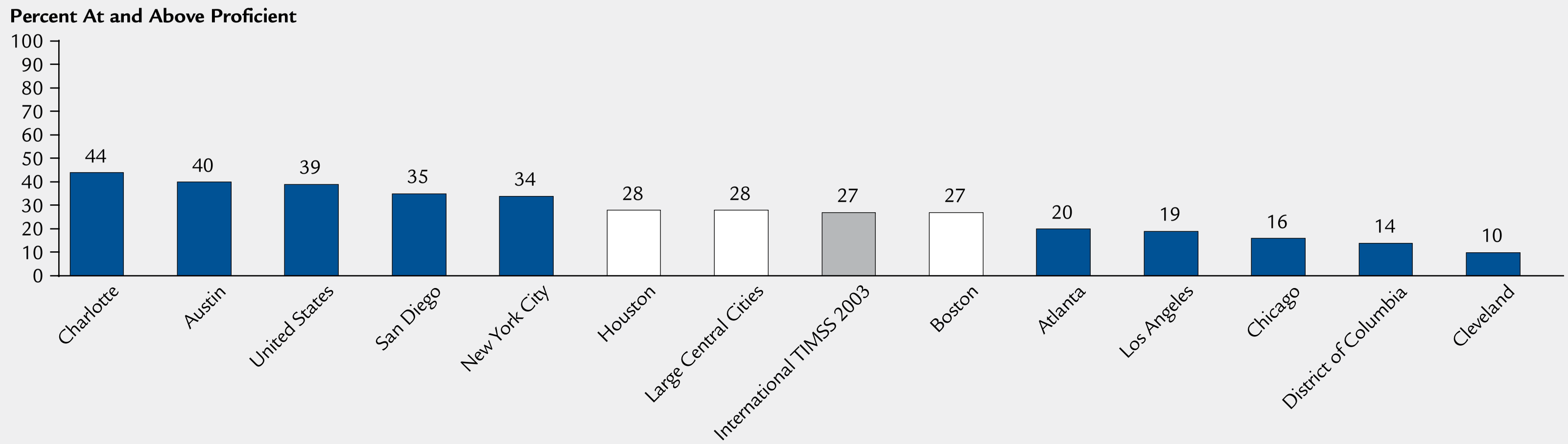
The NAEP framework for the 2007 assessment defined performance in mathematics for Grades 4 and 8 as follows:

Grade 4 Achievement Levels

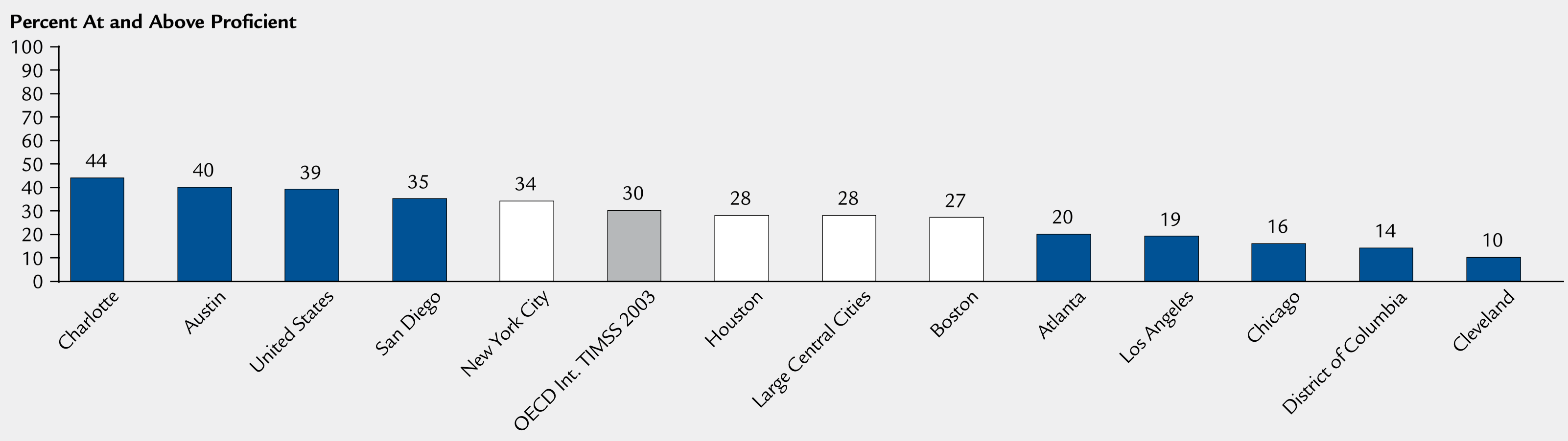
Basic—*Fourth-grade students performing at the* ***Basic*** *level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content areas.*

Figure 13: Comparing the School District Performance to the TIMSS Average and OECD Average for Grade 4 Mathematics

2007 NAEP Grade 4 Mathematics Compared to 2003 TIMSS International Average



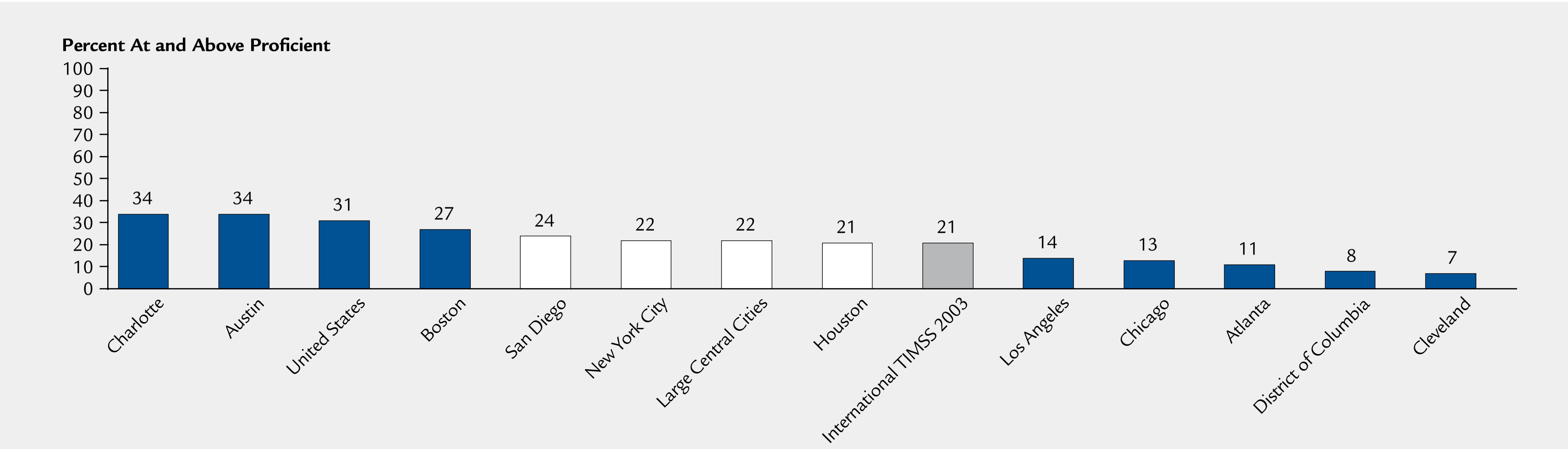
2007 NAEP Grade 4 Mathematics Compared to 2003 OECD International Average for Countries Participating in 2003 TIMSS



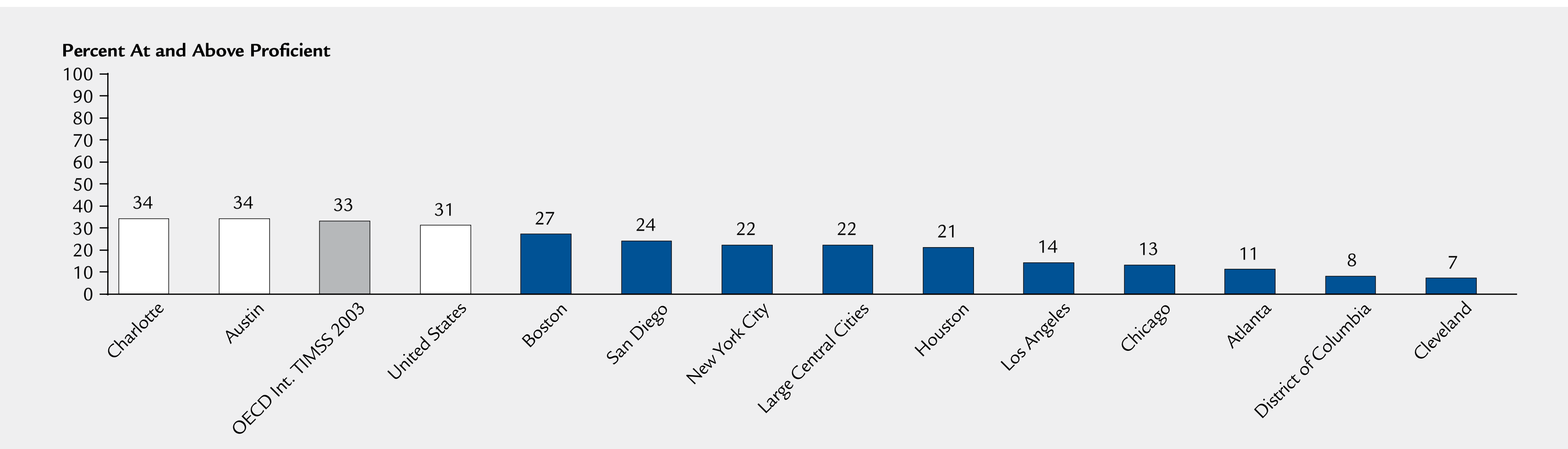
Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Figure 14: Comparing the School District Performance to the TIMSS Average and OECD Average for Grade 8 Mathematics

2007 NAEP Grade 8 Mathematics Compared to 2003 TIMSS International Average



2007 NAEP Grade 8 Mathematics Compared to 2003 OECD International Average for Countries Participating in 2003 TIMSS



Source: Phillips, G., & Dossey, J. (2008). *Counting on the future*. Washington, DC: American Institutes for Research (based on a secondary analysis of 2003 TIMSS and 2007 NAEP).

Fourth graders performing at the *Basic* level should be able to estimate and use basic facts to perform simple computations with whole numbers; show some understanding of fractions and decimals; and solve some simple real-world problems in all NAEP content areas. Students at this level should be able to use—though not always accurately—four-function calculators, rulers, and geometric shapes. Their written responses are often minimal and presented without supporting information.

Proficient*—Fourth-grade students performing at the* ***Proficient*** *level should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.*

Fourth graders performing at the *Proficient* level should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content areas; and use four-function calculators, rulers, and geometric shapes appropriately. Students performing at the *Proficient* level should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented both with supporting information and explanations of how they were achieved.

Advanced*—Fourth-grade students performing at the* ***Advanced*** *level should apply integrated procedural knowledge and conceptual understanding to complex and non-routine real-world problem solving in the five NAEP content areas.*

Fourth graders performing at the *Advanced* level should be able to solve complex non-routine real-world problems in all NAEP content areas. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. These students are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely (NAGB, 2006, p. 53).

Grade 8 Achievement Levels

Basic*—Eighth-grade students performing at the* ***Basic*** *level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.*

Eighth graders performing at the *Basic* level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools, including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving. As they approach the *Proficient* level, students at the *Basic* level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these eighth graders show limited skill in communicating mathematically.

Proficient*—Eighth-grade students performing at the* ***Proficient*** *level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.*

Eighth graders performing at the *Proficient* level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections among fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of *Basic* level arithmetic operations—an understanding sufficient for problem solving in practical situations. Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs, apply properties of informal geometry, and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

Advanced—*Eighth-grade students performing at the* ***Advanced*** *level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content areas.*

Eighth graders performing at the *Advanced* level should be able to probe examples and counterexamples in order to shape generalizations from which they can develop models. Eighth graders performing at the *Advanced* level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions (NAGB, 2006, pp. 54-55).

Using these concepts about what students should know and be able to do can help illuminate what the typical student in the United States, in the 11 urban districts, and in the TIMSS 2003 nations knows and is able to do in mathematics. That information is presented in Table 1–Table 6. Table 1 and Table 4[5] show the projected NAEP achievement levels for mathematics performance that is associated with the typical or average student in each participating TIMSS 2003 country. For example, the average score of 594 for Singapore on the Grade 4 TIMSS scale is associated with the Proficient level of performance. Table 2 and Table 5 shows the same thing for the United States and Table 3 and Table 6 shows the results for the 11 urban districts (except now the achievement levels are based on the actual NAEP assessment rather than those projected on the TIMSS scale). Per Table 3, the typical U.S. student in Grade 4 performs at the Basic level of proficiency. The same is true for 10 of the 11 urban school districts. Of the 11 districts, only the District of Columbia performs at the Below Basic level.

Together, these data create a criterion-referenced picture, graded against the NAEP achievement levels, of the typical student for each country, for the United States as a whole, for each state, and for each urban district.

[5] Note: In Grade 8 mathematics, the projected achievement levels (Basic-473, Proficient–555, and Advanced-631) for each nation in TIMSS in this report are slightly different from those reported by Phillips (2007a). This is because the projected achievement levels in the Phillips (2007a) report were based on a linking study that compared 2000 NAEP to 1999 TIMSS in Grade 8 mathematics and science. This current report uses the latest data available and conducts a more recent linking study using 2003 NAEP and 2003 TIMSS in Grades 4 and 8 mathematics. This more recent linking study is described in Technical Appendix A. The results of the two studies for mathematics Grade 8 are almost identical; the minor differences are displayed in Table 16 and Table 17 of Technical Appendix A.

Table 1: NAEP Grade 4 Mathematics Achievement Levels Projected on the TIMSS Scale for Countries (Basic-464, Proficient-559, Advanced-649)

	2003 TIMSS Grade 4 Mathematics	
Country	Mean	Achievement Level of Mean
Singapore	594	Proficient
Hong Kong SAR	575	Proficient
Japan	565	Proficient
Chinese Taipei	564	Proficient
Belgium (Flemish)	551	Basic
Netherlands	540	Basic
Latvia	536	Basic
Lithuania	534	Basic
Russian Federation	532	Basic
England	531	Basic
Hungary	529	Basic
United States TIMSS	**518**	**Basic**
Cyprus	510	Basic
Rep. of Moldova	504	Basic
Italy	503	Basic
Australia	499	Basic
New Zealand	493	Basic
Scotland	490	Basic
Slovenia	479	Basic
Armenia	456	Below Basic
Norway	451	Below Basic
Islamic Rep. of Iran	389	Below Basic
Philippines	358	Below Basic
Morocco	347	Below Basic
Tunisia	339	Below Basic

Table 2: NAEP Grade 4 Mathematics Achievement Levels for States (Basic-214, Proficient-249, Advanced-282)

	2007 State Grade 4 Mathematics	
State	**Mean**	**Achievement Level of Mean**
Massachusetts	252	Proficient
New Jersey	249	Basic
New Hampshire	249	Basic
Kansas	248	Basic
Minnesota	247	Basic
Vermont	246	Basic
North Dakota	245	Basic
Indiana	245	Basic
Ohio	245	Basic
Wisconsin	244	Basic
Pennsylvania	244	Basic
Wyoming	244	Basic
Montana	244	Basic
Virginia	244	Basic
Iowa	243	Basic
Connecticut	243	Basic
New York	243	Basic
Washington	243	Basic
Maine	242	Basic
Texas	242	Basic
Florida	242	Basic
Delaware	242	Basic
North Carolina	242	Basic
South Dakota	241	Basic
Idaho	241	Basic
Maryland	240	Basic
Colorado	240	Basic
DoDEA[6]	240	Basic
Missouri	239	Basic
Utah	239	Basic
United States	**239**	**Basic**
Nebraska	238	Basic
Arkansas	238	Basic
Michigan	238	Basic
Illinois	237	Basic
Alaska	237	Basic
South Carolina	237	Basic
Oklahoma	237	Basic
West Virginia	236	Basic
Oregon	236	Basic
Rhode Island	236	Basic
Georgia	235	Basic
Kentucky	235	Basic
Hawaii	234	Basic
Tennessee	233	Basic
Arizona	232	Basic
Nevada	232	Basic
Louisiana	230	Basic
California	230	Basic
Alabama	229	Basic
New Mexico	228	Basic
Mississippi	228	Basic
District of Columbia	213.7	Below Basic

[6] Department of Defense Education Activity (overseas and domestic schools)

Table 3: NAEP Grade 4 Mathematics Achievement Levels for Districts (Basic-214, Proficient-249, Advanced-282)

	2007 TUDA Grade 4 Mathematics	
District	**Mean**	**Achievement Level of Mean**
Charlotte	244	Basic
Austin	241	Basic
United States	**239**	**Basic**
New York City	236	Basic
Houston	234	Basic
San Diego	234	Basic
Boston	233	Basic
Atlanta	224	Basic
Los Angeles	221	Basic
Chicago	220	Basic
Cleveland	215	Basic
District of Columbia	213.7	Below Basic

Table 4: NAEP Grade 8 Mathematics Achievement Levels Projected on the TIMSS Scale for Countries (Basic-473, Proficient-555, Advanced-631)

	2003 TIMSS Grade 8 Mathematics	
Country	**Mean**	**Achievement Level of Mean**
Singapore	605	Proficient
Rep. of Korea	589	Proficient
Hong Kong SAR	586	Proficient
Chinese Taipei	585	Proficient
Japan	570	Proficient
Belgium (Flemish)	537	Basic
Netherlands	536	Basic
Estonia	531	Basic
Hungary	529	Basic
Slovak Republic	508	Basic
Russian Federation	508	Basic
Malaysia	508	Basic
Latvia	508	Basic
Australia	505	Basic
United States TIMSS	**504**	**Basic**
Lithuania	502	Basic
Sweden	499	Basic
Scotland	498	Basic
Israel	496	Basic
New Zealand	494	Basic
Slovenia	493	Basic
Italy	484	Basic
Armenia	478	Basic
Serbia	477	Basic
Bulgaria	476	Basic
Romania	475	Basic
Norway	461	Below Basic
Rep. of Moldova	460	Below Basic
Cyprus	459	Below Basic
Rep. of Macedonia	435	Below Basic
Lebanon	433	Below Basic
Jordan	424	Below Basic
Indonesia	411	Below Basic
Islamic Rep. of Iran	411	Below Basic
Tunisia	410	Below Basic
Egypt	406	Below Basic
Bahrain	401	Below Basic
Palestinian Nat'l Auth.	390	Below Basic
Chile	387	Below Basic
Morocco	387	Below Basic
Philippines	378	Below Basic
Botswana	366	Below Basic
Saudi Arabia	332	Below Basic
Ghana	276	Below Basic
South Africa	264	Below Basic

Table 5: NAEP Grade 8 Mathematics Achievement Levels for States (Basic-262, Proficient-299, Advanced-333)

	2007 State Grade 8 Mathematics	
State	**Mean**	**Achievement Level of Mean**
Massachusetts	298	Basic
Minnesota	292	Basic
North Dakota	292	Basic
Vermont	291	Basic
Kansas	290	Basic
New Jersey	289	Basic
South Dakota	288	Basic
Virginia	288	Basic
New Hampshire	288	Basic
Montana	287	Basic
Wyoming	287	Basic
Maine	286	Basic
Colorado	286	Basic
Pennsylvania	286	Basic
Texas	286	Basic
Maryland	286	Basic
Wisconsin	286	Basic
Iowa	285	Basic
DoDEA	285	Basic
Indiana	285	Basic
Washington	285	Basic
Ohio	285	Basic
North Carolina	284	Basic
Oregon	284	Basic
Nebraska	284	Basic
Idaho	284	Basic
Delaware	283	Basic
Alaska	283	Basic
Connecticut	282	Basic
South Carolina	282	Basic
Utah	281	Basic
Missouri	281	Basic
Illinois	280	Basic
United States	**280**	**Basic**
New York	280	Basic
Kentucky	279	Basic
Florida	277	Basic
Michigan	277	Basic
Arizona	276	Basic
Rhode Island	275	Basic
Georgia	275	Basic
Oklahoma	275	Basic
Tennessee	274	Basic
Arkansas	274	Basic
Louisiana	272	Basic
Nevada	271	Basic
California	270	Basic
West Virginia	270	Basic
Hawaii	269	Basic
New Mexico	268	Basic
Alabama	266	Basic
Mississippi	265	Basic
District of Columbia	248	Below Basic

Table 6: NAEP Grade 8 Mathematics Achievement Levels for Districts (Basic-262, Proficient-299, Advanced-333)

	2007 TUDA Grade 8 Mathematics	
District	**Mean**	**Achievement Level of Mean**
Charlotte	283	Basic
Austin	283	Basic
United States	**280**	**Basic**
Boston	276	Basic
Houston	273	Basic
San Diego	272	Basic
New York City	270	Basic
Chicago	260	Below Basic
Los Angeles	257	Below Basic
Cleveland	257	Below Basic
Atlanta	256	Below Basic
District of Columbia	248	Below Basic

Discussion and Conclusions

This paper demonstrates that a statistical linking strategy can be used to combine results from NAEP and TIMSS to create a meaningful index of comparisons of student performance in Grades 4 and 8 mathematics at the urban district, state, national, and international levels. The index presents the percentage of students at or above the Proficient level, as defined by the NAEP achievement levels. By statistically linking the NAEP scores of districts and nations to the TIMSS scale, these same achievement levels can be located on the TIMSS scale, permitting an index to be calculated across all nations that participate in TIMSS. Districts, states, and nations can use this type of information to monitor performance and know how much progress is needed, as measured against international benchmarks.

By using the indicator of the percentage of students at or above the Proficient level, we have identified an index, which NAEP provides in the form of a single number for each state that participates in NAEP and each urban district that participates in TUDA, that is easy to understand and that represents a high, yet appropriate, level of expected performance. Moreover, the indicator is a direct measure of what students are learning in their mathematics curricula at Grades 4 and 8. In addition, the indicator is external to the districts, states, and nations that participate in the surveys: The districts, states, and nations cannot bias the selected samples, alter the test administration, or select the test items in an advantageous way. Consequently, they cannot, through their own actions, beat the system or corrupt the indicator.

The analysis reveals a wide variance among states and among the 11 urban districts, ranging from Charlotte and Austin at the upper end to the District of Columbia and Cleveland at the lower end. The findings firmly highlight the widely reported Basic performance of U.S. students nationally and the Basic and Below Basic performance of students in the 11 urban districts.

The recent report by the National Mathematics Advisory Panel (2008) pretty much sums up the importance of mathematics to the United States:

> *The eminence, safety, and well-being of nations have been entwined for centuries with the ability of their people to deal with sophisticated quantitative ideas. Leading societies have commanded mathematical skills that have brought them advantages in medicine and health, in technology and commerce, in navigation and exploration, in defense and finance, and in the ability to understand past failures and to forecast future developments. (p. xi)*

Given that the increasing mathematical demands of the workplace, especially in positions that have even the smallest room for continual advancement, are at the Proficient level at the minimum, the average student in the 11 U.S. urban districts is facing a quantitative headwind in his or her adult life.

The findings in this report reinforce the fact that neither the typical student in the United States or in any of the 11 urban districts has achieved the Proficient level of performance found in Singapore, Hong Kong SAR, Chinese Taipei, and Japan. If the United States is counting on today's mathematics education to seed the future technology and science needed to carry our cities and our nation forward, then we are already at a competitive disadvantage.

References

Achieve. (2004). *Ready or not: Creating a high school diploma that counts*. Washington, DC.

ACT. (2005). *Courses count: Preparing students for postsecondary success*. Iowa City, IA.

ACT. (2006). *Ready for college and ready for work: Same or different?* Iowa City, IA.

Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972–2000*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Ayres, I. (2007). *Super crunchers: Why thinking-by-numbers is the new way to be smart*. New York: Bantam.

Barton, P. (2006). *High school reform and work: Facing labor market realities*. Princeton, NJ: Educational Testing Service.

Barton, P., & Cooley, R. J. (2008). *Windows on achievement and inequality*. Princeton, NJ: Educational Testing Service.

Beaton, A. E., & Gonzales, E. J. (1993). *Comparing the NAEP Trial State Assessment results with the IAEP international results* (Report prepared for the National Academy of Education Panel on the NAEP Trial State Assessment). Stanford, CA: National Academy of Education.

Benjamini, Y., & Hochberg, Y. (1994). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Serial B*. 57(1), 289–300.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8*, 3–62.

Braswell, J. S., Dion, G. S., Daane, M. C., & Jin, Y. (2005). *The nation's report card: Mathematics 2003*. Washington, DC: National Center for Education Statistics.

Carnevale, A., & Desrochers, D. (2004). *Standards for what? The economic roots of K–16 reform*. Washington, DC: Educational Testing Service.

Education Trust. (1999). Ticket to nowhere: The gap between leaving high school and entering college and high-performance jobs. *Thinking K-16, 3*(2), 3–9.

Friedman, T. (2005). *The world is flat: Brief history of the twenty-first century*. New York: Farrar, Straus, and Giroux.

Garofano, A., & Sable, J. (2008). *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2004–05*. Washington, DC: National Center for Education Statistics.

Harford, T. (2008). *The logic of life: The rational economics of an irrational world*. New York: Random House.

Haskins, R. (2008). Education and economic mobility. In R. Haskins, J. Isaacs, & I. Sawhill (Eds.), *Getting ahead or losing ground: Economic mobility in America*. Washington, DC: Economic Mobility Project, The Brookings Institution.

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.

Johnson, E. G., & Siengondorf, A. (1998). *Linking the National Assessment of Educational Progress and the Third International Mathematics and Science Study: Eighth grade results*. Washington, DC: National Center for Education Statistics.

Kurzweil, R. (2006). *The singularity is near: When humans transcend biology*. New York: Penguin.

LaPointe, A. E., Mead, N. A., & Askew, J. M. (1992). *Learning mathematics*. Princeton, NJ: Educational Testing Service,

LaPointe, A. E., Mead, N. A., & Phillips, G. W. (1989). *A world of differences: An international assessment of mathematics and science*. Princeton, NJ: Educational Testing Service.

Lee, J., Grigg, W., & Dion, G. (2007). *The nation's report card: Mathematics 2007*. Washington, DC: National Center for Education Statistics.

Linn, R. L. (1993). Linking results of district assessments. *Applied Measurement in Education, 6*, 83–102.

Lutkus, A., Grigg, W., & Dion, G. (2007). *The nation's report card: Trial urban district assessment in mathematics 2007*. Washington, DC: National Center for Education Statistics.

Meeder, H., & Solares, G. (2006). *Perspectives on English and mathematics skills needed for work readiness*. Washington, DC: Achieve, Inc.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods and prospects*. Princeton, NJ: Policy Information Center, Educational Testing Service.

Moss, P., & Tilly, C. (2001). *Stories employers tell: Race, skill, and hiring in America*. New York: Russell Sage Foundation.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Boston: TIMSS International Study Center, Boston College.

National Assessment Governing Board (NAGB). (2006). *Mathematics framework for the 2007 National Assessment of Educational Progress*. Washington, DC.

National Center for Education Statistics. (2007). *Mapping 2005 state proficiency standards onto the NAEP scales*. Washington, DC: National Center for Education Statistics.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Nohara, D. (2001). *A comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Program for International Student Assessment (PISA)*. Washington, DC: National Center for Education Statistics.

Organization for Economic Cooperation and Development (OECD). (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris.

Organization for Economic Cooperation and Development (OECD). (2004a). *Learning for tomorrow's world: First results from PISA 2003*. Paris.

Organization for Economic Cooperation and Development (OECD). (2004b). *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003*. Paris.

Pashley, P. J., & Phillips, G. W. (1993). *Toward world class standards: A research study linking national and international assessments*. Princeton, NJ: Educational Testing Service.

Phillips, G. W. (2007a). *Chance favors the prepared mind: Mathematics and science indicators for comparing states and nations*. Washington, DC: American Institutes for Research.

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.

Snipes, J., Williams, A., Horwitz, A, Soga, K., & Casserly M. (2007). *Beating the odds: A city-by-city analysis of student performance and achievement gaps on state assessments: Results from the 2005–2006 school year*. Washington, DC: Council of the Great City Schools.

Stephens, M., & Coleman, M. (2007). *Comparing PIRLS and PISA with NAEP in reading, mathematics, and science* (Working Paper). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved September 24, 2008, from http://nces.ed.gov/Surveys/PISA/pdf/comppaper12082004.pdf

U.S. Department of Education. *The nation's report card: Trial Urban District Assessment—mathematics highlights 2003*. Washington, DC.

U.S. Department of Education. (2007, April). *Course credit accrual and dropping out of high school* (Issue Brief) (NCES 2007-018). Washington, DC: National Center for Education Statistics.

Wilson, W. J. (1996). *When work disappears: The world of the new urban poor*. New York: Knopf.

Wimberly, G. L., & Noeth, R. J. (2005). *College readiness begins in middle school*. Iowa City, IA: ACT.

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

Technical Appendix A

Statistical Linking
2003 TIMSS to 2003 NAEP

Linking

This technical appendix describes how and why the statistical linking between NAEP and TIMSS was done. Most of this technical appendix is reproduced and adapted from Phillips (2007a).

Educators, researchers, and policymakers have considerable interest in how the American educational system compares with those in other countries. One major index for comparison is student academic achievement. Unfortunately, a lack of common metrics and different definitions of performance standards make it difficult to compare measures of student achievement. The difficulty is similar to trying to compare the U.S. poverty level to that of other countries in the world. To do this, we first need a common metric. For example, we need to convert currencies of different countries to a common currency, such as dollars. Then we need a common definition and standard of poverty. That means either using a U.S. definition and standard and applying them to the rest of the world or using a common world definition and standard and applying those to the United States. No matter what common metric, definition, and standard are used, some people will argue it should have been done differently or not at all. Such comparisons are not perfect, always require more research, and should be done with caution. However, such cross-country comparisons result in the cross-fertilization of information and help inform debate. In general, comparisons are useful in providing information to policymakers and the general public to help them achieve broad understandings that they otherwise would not have.

This technical appendix shows how to link the scale of the Trends in International Mathematics and Science Study (TIMSS) to the scale of the National Assessment of Educational Progress (NAEP). The purpose of this linking is to locate the NAEP achievement levels on the TIMSS scale. The linking is done for mathematics in Grades 4 and 8.

- For Grade 4, the TIMSS 2003 scale in mathematics was projected on to the Grade 4 NAEP 2003 assessment in mathematics. Once the link was established, it was then applied to the Grade 4 TIMSS 2003 international results.
- For Grade 8, the TIMSS 2003 scale in mathematics was projected on to the Grade 8 NAEP 2003 assessment in mathematics. Again, once the link was established, it was then applied to the Grade 8 TIMSS 2003 international results.

The goal is to project TIMSS onto the NAEP scale and thereby estimate the percentage of Basic, Proficient, and Advanced students in each country that participated in the 2003 TIMSS. The three achievement levels used were *Basic*, *Proficient*, and *Advanced* for mathematics as defined in *The Nation's Report Card: Trial Urban District Assessment—Mathematics Highlights 2003* (U.S. Department of Education). The TIMSS results can be found in *TIMSS 2003: International Mathematics Report* (Mullis et al. 2004).

Linking Methods

Mislevy (1992) and Linn (1993) have described many of the conceptual and statistical issues associated with linking assessments. They have outlined four forms of statistical linking: equating, calibration, projection, and statistical moderation.

The three assumptions that distinguish the different forms of statistical linking are that two tests (call them X and Y) have scores that are highly correlated, measure the same content, and are equally reliable. These assumptions are displayed in Table 7.

In *equating*, both tests, X and Y, have been designed and developed to be equally reliable, and each measures the same content. Equating is used when the goal is to relate two alternate forms of the same test, such as alternate forms of the ACT or the SAT. Under these conditions, the only difference between the two tests is the metric, such as expressing temperature in terms of Fahrenheit or Celsius. In equating, the distributions of tests X and Y are aligned or matched up directly. The matching can be done with equipercentile equating or linear equating, and the distributions can be either observed score distributions or estimates of the true score distributions. When the three assumptions (high correlation, same content, and equal reliability) are met:

- the linking function should be the same for X expressed in terms of Y, and for Y expressed in terms of X, and
- the linking function should be the same for different subgroups, across contexts and time.

In *calibration* (e.g., with the use of item-response theory), two tests are assumed to measure the same content, but they are not equally reliable. For example, one test X might be a long test, whereas the other test Y is short. The two versions of the test are not equated, but they are indirectly comparable because they have been calibrated to a common scale θ. This type of linking is done across years in NAEP, TIMSS, PIRLS, PISA, most state criterion-referenced tests, and most nationally standardized norm-referenced tests. Calibration procedures provide unbiased estimates for individual students and means, but additional statistical techniques are needed to accurately estimate group characteristics, such as the variance or the percentage at and above performance standards. When the two assumptions (high correlation and same content) are met:

- the linking function between X and θ (e.g., the test characteristic curve) is different from the linking function between Y and θ,
- both X and Y can be used to get unbiased estimates of θ for individual students (although the error in the estimates will be higher for Y), and

Table 7: Statistically linking Test X and Test Y

	Equating	Calibration	Projection	Moderation
High true score correlation[7]	x	x	x	
Same content	x	x		
Equal reliability	x			

[7] The true-score correlation between X and Y is assumed to equal 1.0 .

- the observed score distributions of X for groups do not match the observed score distributions for Y.

In *projection*, a regression equation uses the correlation between the two tests to predict the scores on one test Y from those of another test X. There is no assumption that the two tests measure the same content or that they are equally reliable. With projection, there is no longer a symmetric relationship between one test and the other. The conversion table for predicting the first test from the second is different from the table predicting the second test from the first. When the assumption of high correlation is met:

- the linking function for X expressed in terms of Y (e.g., regression equation) will be different from the linking function for Y expressed in terms of X, and
- the linking function will likely be different for different subgroups, across contexts and time.

In *statistical moderation*, the scores on the first test X are adjusted to have the same distributional characteristics as the scores on the second test Y. In this case, X is linked to Y. This is typically done by matching the means and standard deviations of X and Y or matching their percentile ranks. The usual assumption is that both X and Y have been administered to comparable populations of students (e.g., the student populations taking both tests are randomly equivalent). Statistical moderation typically does not use the correlation between the two tests. When statistical moderation is used:

- the linking function for X expressed in terms of Y (e.g., a z-score equivalency) will be different from the linking function for Y expressed in terms of X,
- the linking function will likely be different for different subgroups, across contexts and time, and
- the degree of the relationship between X and Y is typically unknown.

Linking is essentially a process that provides a concordance table that expresses scores on one test (e.g., TIMSS) in terms of the metric of another test (e.g., NAEP). This paper uses *statistical moderation* to link the NAEP achievement levels to TIMSS by extending the process used in the *2000 NAEP–1999 TIMSS Linking Report* (Johnson, Cohen, Chen, Jiang, & Zhang, 2005). The main goal of this report was to use the link between NAEP and TIMSS to estimate how the students in the states of the United States would have performed if they had taken the TIMSS test, based on the fact that they took the NAEP test. This same linking process also can be used to answer the following question: How would students in other countries perform if their TIMSS results could be expressed in terms of NAEP achievement levels? In other words, we can use the findings in the 2003 TIMSS to project the NAEP achievement levels onto the TIMSS scale as a way to interpret how each country performed on the 2003 TIMSS assessment in terms of U.S. performance standards.

Linking NAEP to International Assessments

Several major attempts have been made to link NAEP statistically to international assessments.

The first attempt involved linking the 1991 International Assessment of Educational Progress (IAEP) to the 1992 NAEP in mathematics (Pashley & Phillips, 1993). The IAEP was first conducted in February 1988 in five countries (Ireland, Korea, Spain, United Kingdom, and United States) and four provinces in Canada (LaPointe, Mead, & Phillips, 1989) using representative samples of 13-year-old students assessed in mathematics and science. The IAEP was expanded and repeated again in 1991 (LaPointe, Mead, & Askew, 1992) in 20 countries in which representative samples of 9- and 13-year-old students were assessed in mathematics and science. Pashley & Phillips (1993) conducted the IAEP–NAEP linking study in mathematics using *projection* methodology. To establish the link between the IAEP and NAEP, a nationally representative linking sample of 1,609 students was administered for both the IAEP and NAEP in 1992. The linking study used samples of eighth-grade students who took NAEP versus 13-year-old students who took the IAEP. (NAEP was based on grade, whereas the IAEP was based on age.) The direction of the link was to predict NAEP performance from IAEP results in other countries. The purpose of the study was to estimate how other countries stacked up against the NAEP achievement levels. The IAEP–NAEP linkage was done within the context of the policy environment at the time. The nation's governors and President held the National Education Summit and adopted six broad national goals. The fourth goal was that, by the year 2000, "U.S. students would be the first in the world in science and mathematics achievement." The IAEP–NAEP linking study was the first effort to address directly the need for a common metric and common standard in international comparisons (i.e., predict how other countries would do on NAEP based on their performance on IAEP). Once the predicted NAEP scores were obtained, then the NAEP achievement levels were used to report the performance of different countries. The IAEP was not repeated; however, it had many design features (such as linking studies) that were incorporated into subsequent international assessments of TIMSS.

A second attempt to link NAEP to an international study was done by Beaton and Gonzales (1993). They used *statistical moderation* to link the 1991 IAEP to the 1990 NAEP scale in mathematics. The results of the Beaton and Gonzales (1993) study were similar to the Pashley and Phillips (1993) study only for countries with performance similar to the U.S. average.

The third study used *statistical moderation* to link Grades 4 and 8 1996 NAEP to 1995 TIMSS, Grades 4 and 8, mathematics and science (Johnson & Siengondorf, 1998). Based on the validation analyses (in two states that took both NAEP and TIMSS), the NAEP–TIMSS link appeared to work at Grade 8 but not at Grade 4.[8]

The fourth study (Johnson et al. 2005) used *projection* methods (similar to Pashley and Phillips, 1993) for Grade 8 mathematics and science to link NAEP to TIMSS. The TIMSS assessment in mathematics and science was conducted in 1999, and the NAEP assessment in mathematics and science was conducted in 2000. In addition to projection methods, the study also used *statistical moderation* as a secondary method of linking. Based on a validation study in which 12 states took both NAEP and TIMSS, the general finding was that, for the U.S. national linking sample, the projection method did not work. However, the statistical moderation method (which used the national samples of both NAEP and TIMSS instead of the linking sample) performed well in the validation study.

Although statistical moderation provided an acceptable link, the estimates provided by statistical moderation should be considered rough, ballpark estimates and should be used for only broad policy understandings.

Caveats Associated With Linking NAEP Achievement Levels to TIMSS

Several important caveats are associated with these analyses. First, the standard errors and the validation analyses are based on data collected within only

[8] The link worked at Grade 8 based on the validation sample. The predicted TIMSS results for Minnesota (the only state that administered the eighth grade TIMSS) were comparable to the actual TIMSS results. The link did not work at Grade 4. The predicted TIMSS results for the two states (Colorado and Minnesota) that administered fourth-grade TIMSS were considerably higher than the actual TIMSS results. The study was not able to determine why this result occurred in the Grade 4 link.

the United States. In the United States, students took both NAEP and TIMSS; in all other countries, however, students only took TIMSS. Whether the linking parameters are stable in other countries is an empirical question that cannot currently be answered. In fact, no international linking study has been designed to answer this question. There is no guarantee that linking parameters estimated from one group (e.g., the United States) will be the same in other groups.

The second caveat is that the percentage at or above Basic, Proficient, and Advanced levels in the tables below is based on the assumption of a normal distribution of performance within each country. In most cases, this assumption should be approximately true.

The third caveat is that the achievement levels developed for the NAEP were based on the content of the NAEP. Although content similarities between the eighth-grade NAEP and TIMSS (Nohara, 2001) are substantial, the NAEP achievement levels do not strictly apply to TIMSS. The problem is similar to the poverty-level analogy used above. Definitions and standards of poverty in the United States do not strictly apply to other countries in the world; however, the definitions and standards can be used to estimate approximately how the rest of the world relates to U.S. expectations of a decent standard of living. For a thoughtful and thorough discussion of similarities and differences in several international assessments, review *Comparing PIRLS and PISA with NAEP in reading, mathematics, and science* (Stephens & Coleman, 2007).

All of these caveats reinforce what was said previously about the limits of inference from these data. At best, these concordance tables should be used for rough approximations to give policymakers a general idea of how the 11 school districts stack up against the rest of the world.

Linking Using Statistical Moderation

This report uses the procedures outlined in Johnson et al. (2005), in which NAEP was linked to TIMSS by using statistical moderation. One major difference is that this report uses extant statistics from the 2003 NAEP and TIMSS published reports rather than recalculating them from the public-use data files available from the NAEP and TIMSS assessments.

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

$$y = \hat{A} + \hat{B}(x). \tag{1.1}$$

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

$$\hat{A} = \hat{\mu}_W - \hat{B}\hat{\mu}_X$$
$$\hat{B} = \frac{\hat{\sigma}_W}{\hat{\sigma}_X}. \tag{1.2}$$

In equation (1.2), $\hat{\mu}_X$ and $\hat{\mu}_W$ are the national means of the U.S. TIMSS and U.S. NAEP (for public school students), respectively, while $\hat{\sigma}_X$ and $\hat{\sigma}_W$ are the standard deviations of the tests.

Linking Error Variance in the Scaled Score Metric

The linking procedure described in this paper is straightforward and easy to accomplish. The intermediate calculations of the error variance, however, are complex and tedious. This technical appendix describes the details of how the error variances reported in the paper were determined.

With statistical moderation, the estimated y is a linear transformation of x. Therefore, the linking error variance in y is

$$\hat{\sigma}_y^2 = \hat{B}^2\hat{\sigma}_x^2 + \hat{\sigma}_A^2 + 2(x)\hat{\sigma}_{AB} + (x)^2\hat{\sigma}_B^2. \tag{1.3}$$

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

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

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

$$\hat{\sigma}^2_B = \hat{B}^2\left[\frac{Var(\hat{\sigma}_W)}{\hat{\sigma}^2_W} + \frac{Var(\hat{\sigma}_X)}{\hat{\sigma}^2_X}\right]. \qquad (1.4)$$

Equations (1.3) and (1.4) were used with data in the United States linking sample to derive the estimates of linking error variance in this paper.

The means and standard deviations in equation (1.2) are reported in Table 8 and Table 9. The resulting estimates of the linking parameters $\hat{A}$ and $\hat{B}$ are reported in Table 10 and Table 11.

Table 8: Means and Standard Deviations for National Samples of Grade 4 U.S. Public School Students, 2003 TIMSS and 2003 NAEP Mathematics

	Mean	Error of Mean	Standard Deviation	Error of Standard Deviation
TIMSS	518.24	2.44	76.27	1.75
NAEP	234.00	0.20	28.00	0.14

Sources. Mullis et al. 2004; Braswell et al. 2005

Table 9: Means and Standard Deviations for National Samples of Grade 8 U.S. Public School Students, 2003 TIMSS and 2003 NAEP Mathematics

	Mean	Error of Mean	Standard Deviation	Error of Standard Deviation
TIMSS	504.37	3.31	79.99	2.38
NAEP	276.00	0.30	36.00	0.22

Sources: Mullis et al. 2004; Braswell et al. 2005

Table 10: Estimating 2003 NAEP Mathematics Scores From 2003 TIMSS Mathematics Using Statistical Moderation With U.S. Grade 4 Public School National Samples

	Estimates of Linking Parameters A and B	
	A	B
Parameter	43.73	0.37
Standard error	4.58	0.01
Covariance	-0.04	

Table 11: Estimating 2003 NAEP Mathematics Scores From 2003 TIMSS Mathematics Using Statistical Moderation With U.S. Grade 8 Public School National Samples

	Estimates of Linking Parameters A and B	
	A	B
Parameter	49.02	0.45
Standard error	7.06	0.01
Covariance	-0.09	

The TIMSS score that is associated with the NAEP achievement levels are presented in Table 12 and Table 13. The standard errors of linking reported in Table 12 and Table 13 are the square root of equation (1.3).

It is instructive to compare the standard error of linking for the projected NAEP mean to the standard error of linking for the projected NAEP achievement levels. Because the linking error is smaller at the mean, the standard error of linking for the NAEP projected achievement levels should be larger than for the mean. In fact, this is the case. The standard error of linking curves is presented in Figures 15 and 16. In both cases, the standard error of linking for the mean is smaller than the standard error of linking for the achievement levels reported and in Figure 15 and Figure 16.

Table 12: Grade 4 2003 NAEP Mathematics Achievement Levels Linked to Grade 4 2003 TIMSS

	TIMSS Score Associated With NAEP Achievement Level	NAEP Achievement Level	Standard Error of Linking
Basic	464	214	1.74
Proficient	559	249	1.65
Advanced	649	282	2.96

Table 13: Grade 8 2003 NAEP Mathematics Achievement Levels Linked to Grade 8 2003 TIMSS

	TIMSS Score Associated With NAEP Achievement Level	NAEP Achievement Level	Standard Error of Linking
Basic	473	262	3.48
Proficient	555	299	3.70
Advanced	631	333	6.04

Figure 15: Standard Error of Linking Curve for Mathematics Grade 4

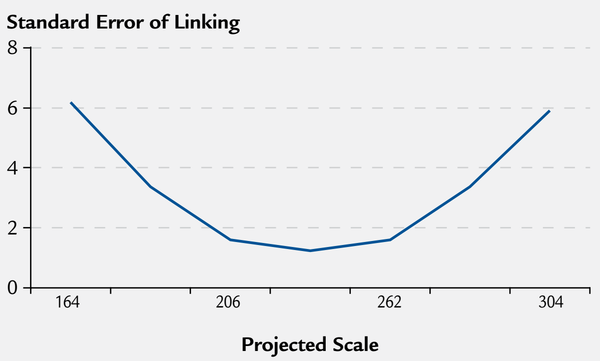
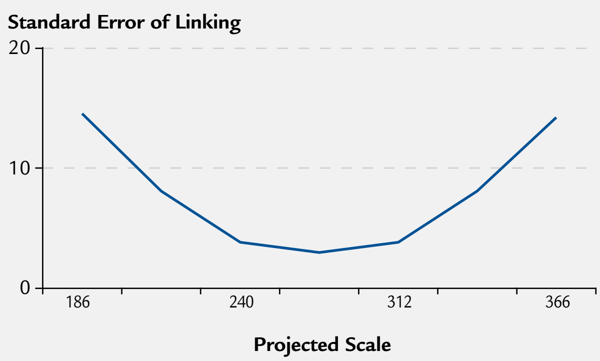


Figure 16: Standard Error of Linking Curve for Mathematics Grade 8



Percentage of Total Survey Error due to Linking

One interesting question in a linking study is how much of the total survey error is due to linking error and how much is due to sampling error. The answer varies by country. Table 14 and Table 15 show the breakdown for the United States 2003 TIMSS distribution projected on the NAEP scale.

In Table 14 and Table 15, the linking error is always larger than the sampling error for all three achievement levels. For the Advanced level, the linking error is substantially larger than the size of the sampling error. In other words, the dominate source of error was due to linking, not sampling. Another way of saying this is that the error variance in this report is greater than the error variance in the 2003 TIMSS report. This is because the 2003 TIMSS does not have linking as a component of error, whereas linking is the major source of error in this report. The moral of this story is that there is substantial error in linking studies and that is why linking error should always be calculated, reported, and taken into account in significance testing.

Validity Evidence for the Linking

This report uses linking between 2003 NAEP and 2003 TIMSS based on fourth- and eighth-grade public school samples in the United States. Fortunately, there is some evidence that supports the validity of this link.

For Grade 8 mathematics, the 2003 NAEP–2003 TIMSS linking was a replication of what was done in the 2000 NAEP–1999 TIMSS (Phillips, 2007b). It makes sense to compare the 2003 linking with the earlier one. This comparison is provided in Table 16.

Table 14: Percentage of Total Error Variance due to Linking and Sampling for TIMSS Projected onto the NAEP Grade 4 Mathematics Scale

	Linking		Sampling		
	Error Variance due to Linking	Percentage of Error Variance due to Linking	Error Variance due to Sampling	Percentage of Error Variance due to Sampling	Total Error Variance
Basic	3.04	79.15	0.80	20.85	3.83
Proficient	2.73	77.34	0.80	22.66	3.53
Advanced	8.76	91.64	0.80	8.36	9.56

Table 15: Percentage of Total Error Variance due to Linking and Sampling for TIMSS Projected onto the NAEP Grade 8 Mathematics Scale

	Linking		Sampling		
	Error Variance due to Linking	Percentage of Error Variance due to Linking	Error Variance due to Sampling	Percentage of Error Variance due to Sampling	Total Error Variance
Basic	12.08	84.49	2.22	15.51	14.30
Proficient	13.70	86.07	2.22	13.93	15.92
Advanced	36.46	94.27	2.22	5.73	38.67

Table 16: Validity of the Linkage Between NAEP and TIMSS for Grade 8 Mathematics in 2003

	Performance Standards Projections for TIMSS 1999–NAEP 2000 Versus TIMSS 2003–NAEP 2003, Grade 8						
	TIMSS 1999	NAEP 2000	1999 SE Linking	TIMSS 2003	NAEP 2003	2003 SE Linking	z-test
Basic	***469***	262	4.83247	***473***	262	3.50530	0.73691
Proficient	***556***	299	5.13256	***555***	299	3.71548	-0.14738
Advanced	***637***	333	6.71745	***631***	333	5.11872	-0.69084

In Table 16, the projected achievement levels in the 1999 TIMSS are compared with the projected achievement levels in the 2003 TIMSS (based on the 2003 NAEP–TIMSS linkage). A z-test shows that the projected achievement levels are not significantly different. This implies that the 2003 linkage between NAEP and TIMSS (Grade 8, mathematics) was stable over the 4 years since the 1999 linkage.

This fact is further supported in Table 17 by comparing the percentage at or above Proficient on 2003 TIMSS

Table 17: Percentage at and Above Projected Proficient Achievement Level for 2003 TIMSS in Grade 8 Mathematics Based on Two Different Linking Studies

Country	2000 NAEP–1999 TIMSS Link Percent	2003 NAEP–2003 TIMSS Link Percent
Armenia	18	18
Australia	27	27
Bahrain	2	2
Belgium (Flemish)	40	40
Botswana	0	0
Bulgaria	17	17
Chile	2	2
Chinese Taipei	61	62
Cyprus	11	12
Egypt	5	5
Estonia	36	36
Ghana	0	0
Hong Kong SAR	66	66
Hungary	37	37
Indonesia	5	5
Islamic Rep. of Iran	2	3
Israel	24	24
Italy	17	18
Japan	57	57
Jordan	7	7
Rep. of Korea	65	66
Latvia	25	26
Lebanon	3	3
Lithuania	24	25
Rep. of Macedonia	8	9
Malaysia	26	26
Rep. of Moldova	12	12
Morocco	1	1
Netherlands	38	39
New Zealand	21	22
Norway	9	9
Palestinian Nat’l Auth.	4	4
Philippines	2	2
Romania	18	19
Russian Federation	26	27
Saudi Arabia	0	0
Scotland	22	22
Serbia	19	19
Singapore	73	73
Slovak Republic	28	28
Slovenia	19	19
South Africa	0	0
Sweden	21	21
Tunisia	1	1
United States	26	26

based on estimates obtained from the earlier 2000 NAEP–1999 TIMSS linking study versus the current 2003 NAEP–2003 TIMSS linking study.

Table 17 indicates that the estimated percentages of students at and above Proficient from both linking studies are almost identical.

In addition to comparing the 2003 linkage to the 1999 linkage, a second piece of evidence helps to validate the 2003 linkage. In 2003, the state of Indiana participated in both NAEP and TIMSS. Therefore, Indiana provides a cross-validation sample that can be used to check the predictions of the linkage derived from the U.S. national sample. Per Table 18, the predicted NAEP mean for Indiana (predicted from the linking analysis) is not significantly different from the actual mean. Furthermore, Table 19 indicates that the percentage of students achieving each achievement level predicted from TIMSS in Indiana was statistically equivalent to the percentage of students actually achieving each achievement level on NAEP. These results support the validity of the 2003 NAEP–2003 TIMSS linkage for Grade 4 in mathematics.

Table 18: Projected Mean Versus Actual NAEP Mean for Indiana, 2003, Grade 4, Mathematics

	Projected Mean on NAEP 2003 Mathematics	Standard Error of Projected Mean on NAEP 2003 Mathematics	Actual Mean on NAEP 2003 Mathematics	Standard Error of Actual Mean on NAEP 2003 Mathematics	z-test
Mean	239	1.38	238	0.90	0.85

Table 19: Percentage at and Above Projected From TIMSS Versus the Actual Percentage at and Above for NAEP in Indiana, 2003, Grade 4, Mathematics

	Projected Percentage at and Above on NAEP 2003 Mathematics	Standard Error of Projected Percentage on NAEP 2003 Mathematics	Actual Percentage at and Above on NAEP 2003 Mathematics	Standard Error of Actual Percentage on NAEP 2003 Mathematics	z-test
% Basic	86	2.56	82	1.00	1.32
% Proficient	34	3.15	35	1.40	-0.18
% Advanced	4	1.00	4	0.50	-0.26

Table 20: Projected Mean Versus Actual NAEP Mean for Indiana, 2003, Grade 8, Mathematics

	Projected Mean on NAEP 2003 Mathematics	Standard Error of Projected Mean on NAEP 2003 Mathematics	Actual Mean on NAEP 2003 Mathematics	Standard Error of Actual Mean on NAEP 2003 Mathematics	z-test
Mean	278	2.79	281	1.10	-1.12

These last two analyses can be replicated for the 2003 NAEP–2003 TIMSS linkage for mathematics at Grade 8. Table 20 indicates that the predicted NAEP Grade 8 mathematics mean for Indiana is statistically equivalent to the actual NAEP mean. In Table 21, all of the percentages predicted to achieve the NAEP projected achievement levels in Indiana are statistically the same as the percentages reaching the actual achievement levels in NAEP.

Descriptive Statistics Associated With the Projected Distribution in Each Country

After linking TIMSS to NAEP in the U.S. linking sample, the relationship between TIMSS and NAEP can be used to establish a projected distribution of TIMSS on the NAEP scale for each TIMSS country. Assuming the projected distribution is approximately normal, then a variety of descriptive statistics can be obtained for the projected distribution. The mean, standard deviation, and standard error of the mean of the projected distribution for each country are as follows.

$$\hat{\mu}_Y = \hat{A} + \hat{B}\left(\hat{\mu}_X\right) \tag{1.5}$$

$$\sigma_Y = B\hat{\sigma}_X \tag{1.6}$$

$$\hat{\sigma}_{\mu_Y} = \sqrt{B^2\hat{\sigma}^2_{\mu_X} + \hat{\sigma}^2_y} \tag{1.7}$$

Percentage at and Above Estimated Achievement Levels on TIMSS

To estimate the percentage of students at and above the estimated achievement levels in foreign countries, the *projected distribution* of TIMSS on the NAEP scale must be estimated for each country. Another way of saying this is we want to integrate $\Pr\left(Y \geq y_c\right)$ over the x distribution $f_N\left(x\middle|\hat{\mu}_X, \hat{\sigma}^2_X\right)$ in each country. This was done by making the assumption that the projected TIMSS distribution and the actual TIMSS distribution in each country are approximately normal. Given the normality assumption, the marginal proportion of students at and above each estimated achievement level on TIMSS, $(1 - p_c)$ is equal to

$$1 - p_c = \Pr\left(Y \geq y_c\right) = \int_{-\infty}^{\infty} \Pr\left(Y \geq y_c \middle| x\right) f_N\left(x\middle|\hat{\mu}_X, \hat{\sigma}^2_X\right) dx \tag{1.8}$$

Table 21: Percentage at and Above Projected From TIMSS Versus the Actual Percentage at and Above for NAEP in Indiana, 2003, Grade 8, Mathematics

	Projected Percentage at and Above on NAEP 2003 Mathematics	Standard Error of Projected Percentage on NAEP 2003 Mathematics	Actual Percentage at and Above on NAEP 2003 Mathematics	Standard Error of Actual Percentage on NAEP 2003 Mathematics	z-test
% Basic	69	4.65	74	1.40	-1.03
% Proficient	25	4.34	31	1.20	-1.36
% Advanced	4	1.61	5	0.40	-0.64

In equation (1.8), the integral is equal to the average $\Pr(Y \geq y_c)$ across the TIMSS sample. The integral above estimates the proportion at and above each projected achievement level on TIMSS. Multiplying the integral by 100 expresses the proportion as a percentage.

Determining the Overall International Average and the International OECD Average for the Percentage of Students at and Above the Projected Proficient Levels

To provide useful international comparisons for the school districts, two international averages were determined. The first was the percentage of students at and above the projected proficient standard for the aggregate of all countries in the TIMSS study. The U.S. data were removed from this aggregate so that it would not be counted in the international total. Therefore, the overall international distribution was based on 24 countries at Grade 4 and 44 countries at Grade 8. The second aggregate was the participating OECD countries (again excluding the United States). This aggregate was based on 10 countries at Grade 4 and 12 countries at Grade 8. Rather than weighting each country by its population size, unweighted averages were used, as was the same practice used in the 2003 international TIMSS report. Therefore, the formula used for the international mean, international standard deviation, and standard error of the international mean were as follows.

Suppose there are m TIMSS countries, the jth country has a normal distribution with estimated mean $\hat{\mu}_{Y_j}$ and standard deviation $\hat{\sigma}_{Y_j}$. The standard error of the estimated mean is $\hat{\sigma}_{\hat{\mu}_{Y_j}}$. The metric of TIMSS has been be expressed in terms of the metric of NAEP through equation . The international aggregates of these projected NAEP statistics can be estimated as follows.

$$\hat{\mu}_Y = \frac{\sum_{j=1}^{m} \hat{\mu}_{Y_j}}{m}, \tag{1.9}$$

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{j=1}^{m} \left(\hat{\sigma}_{Y_j}^2 + \hat{\mu}_{Y_j}^2\right)}{m} - \hat{\mu}_Y^2}, \tag{1.10}$$

and

$$\hat{\sigma}_{\mu_y}^2 = \frac{\sqrt{\sum_{j=1}^{m} \hat{\sigma}_{\mu_{y_j}}^2}}{m} \tag{1.11}$$

Table 22: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the TIMSS Distribution for Each Country for Grade 4

	Number of Countries Aggregated	Mean	Standard Error of the Mean	Standard Deviation
International Aggregate	24	495	0.8	107
OECD Aggregate	10	515	0.9	83

Note: The U.S. sample was excluded from these aggregates.

Table 23: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the Projected NAEP Distribution for Each Country for Grade 4

	Number of Countries Aggregated	Mean	Standard Error of the Mean	Standard Deviation
International Aggregate	24	225	0.98	39
OECD Aggregate	10	233	0.98	30

Note: The U.S. sample was excluded from these aggregates.

The resulting international averages are presented in Table 22–Table 25.

Once these statistics are calculated, then equation can be used to estimate the overall international average and the international OECD average for the percentage of students at and above the projected proficiency levels.

Linking Error Variance at and Above the Projected Achievement Levels

So far in this technical appendix, all the error variances have been calculated in the scale score metric. However, the report is really about the proportion of students at and above various estimated achievement levels in each country. Thus, the standard errors of linking in the cumulative proportion metric must be determined. Linking error variance in the cumulative proportion metric can be approximated with the Taylor series as follows:

$$\begin{aligned}\hat{\sigma}^2_{Linking} &= Var(1-\hat{p}_c) \\ &= Var\left(h(\hat{\mu}_Y,\hat{\sigma}_Y)\right) \\ &\approx \begin{pmatrix} \dfrac{\partial h(\hat{\mu}_Y,\hat{\sigma}_Y)}{\partial\hat{\mu}_Y} & \dfrac{\partial h(\hat{\mu}_Y,\hat{\sigma}_Y)}{\partial\hat{\sigma}_Y}\end{pmatrix} \\ &\times Var(\hat{\mu}_Y,\hat{\sigma}_Y) \\ &\times \begin{pmatrix} \dfrac{\partial h(\hat{\mu}_Y,\hat{\sigma}_Y)}{\partial\hat{\mu}_Y} & \dfrac{\partial h(\hat{\mu}_Y,\hat{\sigma}_Y)}{\partial\hat{\sigma}_Y}\end{pmatrix}^T\end{aligned}$$

Finding the partial derivatives and simplifying the equations results in

$$\hat{\sigma}^2_{Linking} \approx \left(\frac{\exp\left(-(y_c-\hat{\mu}_Y)^2/(2\hat{\sigma}^2_Y)\right)}{\sqrt{2\pi\hat{\sigma}_Y}}\right)^2 \mathrm{Var}(\hat{\mu}_Y) + \left(\left(\frac{y_c-\hat{\mu}_Y}{\hat{\sigma}_Y}\right)\frac{\exp\left(-(y_c-\hat{\mu}_Y)^2/(2\hat{\sigma}^2_Y)\right)}{\sqrt{2\pi\hat{\sigma}_Y}}\right)^2 \mathrm{Var}(\hat{\sigma}_Y) \tag{1.12}$$

Table 24: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the TIMSS Distribution for Each Country for Grade 8

	Number of Countries Aggregated	Mean	Standard Error of the Mean	Standard Deviation
International Aggregate	44	465	0.5	111
OECD Aggregate	12	518	1.0	84

Note: The U.S. sample was excluded from these aggregates.

Table 25: Aggregate of All TIMSS Countries and All OECD TIMSS Countries Using the Projected NAEP Distribution for Each Country for Grade 8

	Number of Countries Aggregated	Mean	Standard Error of the Mean	Standard Deviation
International Aggregate	44	258	1.6	50
OECD Aggregate	12	282	1.6	38

Note: The U.S. sample was excluded from these aggregates.

Sampling Error Variance at and Above the Projected Achievement Levels

Because TIMSS is a survey that is administered in each country, all statistics derived from it will have sampling error. Therefore, the proportion of students at and above each projected achievement level will have sampling error associated with it. The sampling error can be estimated from the published international reports by calculating the standard error of a proportion.

$$\hat{\sigma}^2_{Sampling} = \frac{p_c\left(1-p_c\right)}{effn_{\left(1-p_c\right)}} \tag{1.13}$$

The quantity $effn_{\left(1-p_c\right)}$ is the effective sample size associated with $1-p_c$ (i.e., the actual sample size of the TIMSS survey divided by the design effect for $1-p_c$).

Total Error Variance at and Above the Projected Achievement Levels

The total error variance for the percentage of students at and above each achievement level is the sum of the linking error variance (1.12) and sampling error variance (1.13).

$$\hat{\sigma}^2_{Total\ Error} = \hat{\sigma}^2_{Linking} + \hat{\sigma}^2_{Sampling} \tag{1.14}$$

The standard errors (i.e., the square root of $\hat{\sigma}^2_{Total\ Error}$) for projected Proficient achievement levels are reported in Table 26 for Grade 4 mathematics and Table 27 for Grade 8 mathematics. The standard errors have been multiplied by 100 to report percentages rather than proportions.

Table 26: Percentage of Students at and Above the Proficient Achievement Level Projected on 2003 TIMSS, Grade 4, Mathematics

Country	Percentage	Standard Error
Singapore	66	4.5
Hong Kong SAR	60	3.8
Chinese Taipei	53	2.9
Japan	53	2.5
Belgium (Flemish)	44	3.1
England	37	2.9
Latvia	37	2.9
Lithuania	37	2.8
Netherlands	37	3.4
Russian Federation	36	4.1
Hungary	35	2.8
Cyprus	28	2.1
Rep. of Moldova	26	3.2
Italy	25	2.7
Australia	23	2.8
New Zealand	22	1.9
Scotland	19	2.3
Slovenia	15	1.9
Armenia	12	1.8
Norway	9	1.4
Philippines	3	1.6
Islamic Rep. of Iran	2	1.0
Tunisia	1	0.7
Morocco	1	0.8

Table 27: Percentage of Students at and Above the Proficient Achievement Level Projected on 2003 TIMSS, Grade 8, Mathematics

Country	Percentage	Standard Error
Singapore	73	3.8
Hong Kong SAR	66	4.1
Rep. of Korea	66	3.1
Chinese Taipei	62	3.5
Japan	57	3.2
Belgium (Flemish)	40	3.5
Netherlands	39	4.3
Hungary	37	3.3
Estonia	36	3.7
Slovak Republic	28	3.0
Australia	27	4.0
Russian Federation	27	3.5
Malaysia	26	3.7
Latvia	26	3.2
Lithuania	25	2.6
Israel	24	2.8
Scotland	22	3.2
New Zealand	22	4.1
Sweden	21	2.8
Slovenia	19	2.5
Serbia	19	2.1
Romania	19	2.9
Armenia	18	2.3
Italy	18	2.7
Bulgaria	17	2.9
Rep. of Moldova	12	2.4
Cyprus	12	1.7
Norway	9	1.9
Rep. of Macedonia	9	1.8
Jordan	7	1.8
Egypt	5	1.5
Indonesia	5	1.7
Palestinian Nat'l Auth.	4	1.1
Lebanon	3	1.4
Islamic Rep. of Iran	3	1.1
Chile	2	1.2
Bahrain	2	0.7
Philippines	2	1.3
Tunisia	1	0.8
Morocco	1	0.6
Botswana	0	0.6
South Africa	0	0.4
Saudi Arabia	0	0.4
Ghana	0	0.3

Table 28: Each City in the Grade 4 2007 NAEP TUDA in Mathematics Compared to Each Country in the Grade 4 2003 TIMSS in Mathematics for the Percentage of Students at and Above the Proficient Level Based on NAEP Achievement Levels Projected onto the TIMSS Scale

		Cities in 2007 NAEP										
Nations in TIMSS	**%**	**Charlotte**	**Austin**	**San Diego**	**New York City**	**Houston**	**Boston**	**Atlanta**	**Los Angeles**	**Chicago**	**District of Columbia**	**Cleveland**
		44	40	35	34	28	27	20	19	16	14	10
Singapore	66	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Hong Kong SAR	60	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Chinese Taipei	53	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Japan	53	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Belgium (Flemish)	44			▼	▼	▼	▼	▼	▼	▼	▼	▼
U.S. NAEP 2007	**39**	▲			▼	▼	▼	▼	▼	▼	▼	▼
England	37					▼	▼	▼	▼	▼	▼	▼
Latvia	37					▼	▼	▼	▼	▼	▼	▼
Lithuania	37					▼	▼	▼	▼	▼	▼	▼
Netherlands	37					▼	▼	▼	▼	▼	▼	▼
Russian Federation	36							▼	▼	▼	▼	▼
Hungary	35	▲						▼	▼	▼	▼	▼
Cyprus	28	▲	▲					▼	▼	▼	▼	▼
Rep. of Moldova	26	▲	▲	▲				▼	▼	▼	▼	▼
Italy	25	▲	▲	▲					▼	▼	▼	▼
Australia	23	▲	▲	▲						▼	▼	▼
New Zealand	22	▲	▲	▲	▲						▼	▼
Scotland	19	▲	▲	▲	▲	▲	▲				▼	▼
Slovenia	15	▲	▲	▲	▲	▲	▲					▼
Armenia	12	▲	▲	▲	▲	▲	▲					
Norway	9	▲	▲	▲	▲	▲	▲	▲	▲	▲		
Philippines	3	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
Islamic Rep. of Iran	2	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Tunisia	1	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Morocco	1	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲

Note: Select a city at the top, and then read down the column for comparisons with the countries listed on the left. The symbol ▲ indicates that the percentage of students Proficient in the city on the left is significantly higher than the comparison country above. The symbol ▼ indicates that the percentage of students Proficient in the city at the top is significantly lower than the comparison country on the left. A blank space indicates the city and nation are statistically comparable. With a 95% confidence interval, 5% of the comparisons will be significant by chance.

Table 29: Each City in Grade 8 2007 NAEP TUDA in Mathematics Compared to Each Country in the Grade 8 2003 TIMSS in Mathematics for the Percentage of Students at and Above the Proficient Level Based on the NAEP Achievement Levels Projected onto the TIMSS Scale

		Cities in 2007 NAEP										
Nations in TIMSS	**%**	**Austin**	**Charlotte**	**Boston**	**San Diego**	**New York City**	**Houston**	**Los Angeles**	**Chicago**	**Atlanta**	**District of Columbia**	**Cleveland**
		34	34	27	24	22	21	14	13	11	8	7
Singapore	73	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Hong Kong SAR	66	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Korea, Rep. of	66	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Chinese Taipei	62	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Japan	57	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Belgium (Flemish)	40			▼	▼	▼	▼	▼	▼	▼	▼	▼
Netherlands	39			▼	▼	▼	▼	▼	▼	▼	▼	▼
Hungary	37			▼	▼	▼	▼	▼	▼	▼	▼	▼
Estonia	36			▼	▼	▼	▼	▼	▼	▼	▼	▼
U.S. NAEP 2007	**31**	▲		▼	▼	▼	▼	▼	▼	▼	▼	▼
Slovak Republic	28						▼	▼	▼	▼	▼	▼
Australia	27							▼	▼	▼	▼	▼
Russian Federation	27							▼	▼	▼	▼	▼
Malaysia	26							▼	▼	▼	▼	▼
Latvia	26	▲	▲					▼	▼	▼	▼	▼
Lithuania	25	▲	▲					▼	▼	▼	▼	▼
Israel	24	▲	▲					▼	▼	▼	▼	▼
Scotland	22	▲	▲					▼	▼	▼	▼	▼
New Zealand	22	▲	▲							▼	▼	▼
Sweden	21	▲	▲					▼	▼	▼	▼	▼
Serbia	19	▲	▲					▼	▼	▼	▼	▼
Slovenia	19	▲	▲	▲					▼	▼	▼	▼
Romania	19	▲	▲	▲						▼	▼	▼
Armenia	18	▲	▲	▲						▼	▼	▼
Italy	18	▲	▲	▲	▲					▼	▼	▼
Bulgaria	17	▲	▲	▲	▲						▼	▼
Rep. of Moldova	12	▲	▲	▲	▲						▼	▼
Cyprus	12	▲	▲	▲	▲	▲	▲				▼	▼
Norway	9	▲	▲	▲	▲	▲	▲					▼
Rep. of Macedonia	9	▲	▲	▲	▲	▲	▲	▲				
Jordan	7	▲	▲	▲	▲	▲	▲	▲				
Egypt	5	▲	▲	▲	▲	▲	▲	▲	▲	▲		
Indonesia	5	▲	▲	▲	▲	▲	▲	▲	▲	▲		
Palestinian Nat'l Auth.	4	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
Lebanon	3	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
Islamic Rep. of Iran	3	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Chile	2	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Bahrain	2	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Philippines	2	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Tunisia	1	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Morocco	1	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Botswana	0	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
South Africa	0	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Saudi Arabia	0	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Ghana	0	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲

Note: Select a city at the top, and then read down the column for comparisons with the countries listed on the left. The symbol ▲ indicates that the percentage of students Proficient in the city on the left is significantly higher than the comparison country above. The symbol ▼ indicates that the percentage of students Proficient in the city at the top is significantly lower than the comparison country on the left. A blank space indicates the city and nation are statistically comparable. With a 95% confidence interval, 5% of the comparisons will be significant by chance.

Technical Appendix B

Significance Testing and Multiple Comparisons

If we conducted only one significance test between country A and country B, then a 95% confidence interval would be 95% $CI = \pm Z_{\alpha/2}\sqrt{\hat{\sigma}^2_{P_{Error}(A)} + \hat{\sigma}^2_{P_{Error}(B)}}$. However, when conducting a large number of hypotheses testing, an adjustment for α is often used to compensate for the fact that many significance tests are being performed. If we have k independent tests, each at level α, then the probability that at least one is falsely rejected is $1-(1-\alpha)^k = \alpha_k$. For example, in the district-by-nation comparisons for mathematics, each district may wish to make 45 comparisons (i.e., 44 international comparisons plus one for the U.S. sample with each district). With each $\alpha = .025$ (i.e., $\alpha = .05$ with a 2-tailed test), the family-wise error rate is $\alpha_k = .69$, so the probability of a false positive (or type-I error) among the 45 comparisons is equal to .69. When conducting multiple hypothesis tests, we usually want to control α_k. This is referred to as controlling the family-wise error rate. The most common type of control for the family-wise error rate is the Bonferroni procedure (Bonferroni, 1936), where the α for each test would be $\alpha = \frac{\alpha_k}{45} = \frac{.025}{45} = .000555$. With this procedure, you divide the significance level for each test by the number of significance tests so that the family-wise error rate is $\alpha_k = \frac{\alpha}{k}$, therefore $\alpha_k = 1-\left(1-\frac{.025}{45}\right)^{45} = .025$. Unfortunately, the Bonferroni procedure suffers from low power properties when the number of tested hypotheses is large.

Instead of controlling for the chance of *any* false positive (like the Bonferroni procedure), the false discovery rate (FDR) controls for the proportion of false positives (Benjamini & Hochberg, 1994). The FDR is the expected proportion of true null hypotheses rejected out of the total number of null hypotheses rejected. Multiple comparison procedures controlling the FDR are more powerful than the commonly used multiple comparison procedures based on the family-wise error rate. FDR controlling procedures are especially suited to situations where a large number of hypotheses are being tested. Suppose k hypotheses are tested, and R of them are rejected. Of the rejected hypotheses, suppose that V of them are really null (i.e., V is the number of type-I errors, or false positives). The FDR is defined as $E\left(\frac{V}{R}\right)$, where E is the expected value. Let $H_1 \ldots H_k$ be the null hypotheses and $P_1 \ldots P_k$ their corresponding p values. The p values have been ordered from lowest (most significant) to highest (least significant). For each P_j, we calculate Q_j, where $Q_j = \frac{j}{k}\alpha$. Using the *FDR*, if $P_j \le Q_j$, then we reject the null hypothesis. The FDR is used in this paper for significance testing in Figures 1–12.