2 Comparing State Performance Standards

Each state has set either one or several *standards* for performance in each grade on its mathematics assessment. We endeavored to select the primary standard for each state as the standard it uses for reporting *adequate yearly progress* to the public. However, we cannot be certain of success in all cases because in some states policies for reporting adequate yearly progress have changed. Short versions of the states' names for the standards are shown in table 2, with the primary standard listed as *standard 3*. NAEP has set three such standards, labeled *basic*, *proficient*, and *advanced*.

These standards are described in words, and they are operationalized as test scores above a corresponding *cutpoint*. This is possible for NAEP, even though the design of NAEP does not support reporting individual scores—NAEP is only intended to provide reliably reportable statistics for broad demographic groups (e.g., gender and racial/ethnic) at the state level or for very large districts.

Because each state's standards are set independently, the standards in different states can be quite different, even though they are named identically. Thus, a score in the *proficient* range in one state may not be in the *proficient* range in another state. Because NAEP is administered to a representative sample of public school students in each state, NAEP can provide the link needed to estimate the difference between two states' achievement standards.

The objective of this comparison is to place all states' mathematics performance standards for grades 4 and 8, or adjacent grades, on a common scale, along with the NAEP achievement levels. This comparison is valuable for two reasons. First, it sheds light on the variations between states in the percentages of students reported to be *proficient*, *meeting the standard*, or *making satisfactory progress*. Second, for comparisons between NAEP and state assessment trends and gaps, it makes possible the removal of one important source of bias: a difference between two years or between two subpopulations in percentages achieving a standard is affected as much by the choice of where that standard is set on the achievement scale as by instructional reform.

Table 2. Short names of state mathematics achievement performance standards, by state: 2003

State/ jurisdiction	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Alabama			Percentile Rank[1]		
Alaska		Below Proficient	Proficient	Advanced	
Arizona		Approaching	Meeting	Exceeding	
Arkansas		Basic	Proficient	Advanced	
California	Below Basic	Basic	Proficient	Advanced	
Colorado			Partially Proficient	Proficient	Advanced
Connecticut	Basic	Proficient	Goal	Advanced	
Delaware		Below	Meeting	Exceeding	Distinguished
District of Columbia		Basic	Proficient	Advanced	
Florida		Limited Success	Partial Success	Some Success	Success
Georgia			Meeting	Exceeding	
Hawaii		Approaching	Meeting	Exceeding	
Idaho		Basic	Proficient	Advanced	
Illinois	Starting	Approaching	Meeting	Exceeding	
Indiana			Pass	Pass Plus	
Iowa			Proficient		
Kansas		Basic	Proficient	Advanced	Exemplary
Kentucky		Apprentice	Proficient	Distinguished	
Louisiana	Approaching Basic	Basic	Mastery	Advanced	
Maine		Partially Meeting	Meeting	Exceeding	
Maryland			Proficient	Advanced	
Massachusetts	Warning	Needs Improvement	Proficient	Advanced	
Michigan		Basic	Meeting	Exceeding	
Minnesota	Partial Knowledge	Satisfactory	Proficient	Superior	
Mississippi		Basic	Proficient		
Missouri	Progressing	Nearing Proficient	Proficient	Advanced	
Montana		Nearing Proficient	Proficient	Advanced	
Nebraska			Meeting		
Nevada		Approaching	Meeting	Exceeding	
New Hampshire			Basic	Proficient	Advanced
New Jersey			Proficient	Advanced	
New Mexico		Top 75%	Top half	Top 25%	
New York		Need Help	Meeting	Exceeding	
North Carolina		Inconsistent Mastery	Consistent Mastery	Superior	
North Dakota			Meeting		
Ohio		Basic	Proficient	Advanced	
Oklahoma		Little Knowledge	Satisfactory	Advanced	
Oregon			Meeting	Exceeding	
Pennsylvania		Basic	Proficient	Advanced	
Rhode Island			Proficient		
South Carolina		Basic	Proficient	Advanced	
South Dakota		Basic	Proficient		
Tennessee			Percentile Rank		
Texas			Passing		
Utah			Percentile Rank		
Vermont	Below	Nearly	Achieved	Honors	
Virginia			Proficient	Advanced	
Washington		Below	Met	Above	
West Virginia		Top 75%	Top half	Top 25%	
Wisconsin		Basic	Proficient	Advanced	
Wyoming		Partially Proficient	Proficient	Advanced	

1. Percentile rank while not a standard, is needed for comparisons in Alabama, Tennessee, and Utah. Similarly, for New Mexico and West Virginia quartiles are used for comparisons.

NOTE: Standard 3 represents the primary standard for every state. In most cases, it is the criterion for Adequate Yearly Progress (AYP). The state standards listed above are those for which assessment data exist in the NLSLSASD.

SOURCE: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

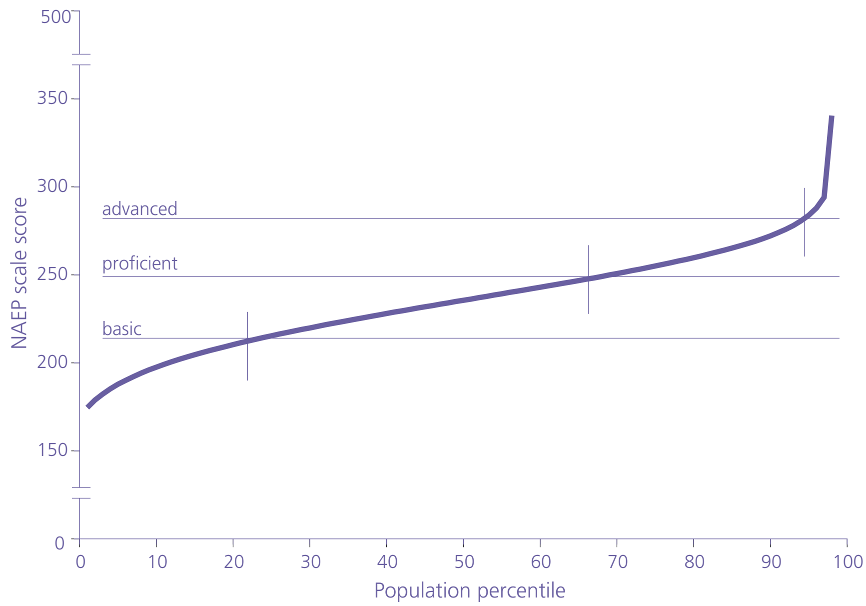
NAEP ACHIEVEMENT DISTRIBUTION

To understand the second point, we introduce the concept of a *population profile* of NAEP achievement. Achievement is a continuous process, and each individual student progresses at his or her own rate. When they are tested, these students demonstrate levels of achievement all along the continuum of mathematics skills, and these are translated by the testing into numerical scale values. Summarizing the achievement of a population as the percentage of students who meet a standard conveys some information, but it hides the profile of achievement in the population - how large the variation in achievement is, whether high-achieving students are few, with extreme achievement, or many, with more moderate achievement, and whether there are few or many students who lag behind the mainstream of achievement. A population profile is the display of the achievement of each percentile of the population, from the lowest to the highest, and by overlaying two population profiles, one can display comparisons of achievement gains and achievement gaps at each percentile. More important for the comparison of standards across states, a population profile can show how placement of a standard makes a difference in how an achievement gain translates into a gain in the percentage of students meeting that standard.

Figure 1 displays a population profile of mathematics achievement in grade 4, as measured by NAEP in 2003. To read the graph, imagine students lined up along the horizontal axis, sorted from the lowest performers on a mathematics achievement test at the left to the highest performers at the right. The graph gives the achievement score associated with each of these students. For reference, figure 1 also includes the NAEP scale scores that are thresholds for the achievement levels. The percentage of student scores at or above the basic threshold score of 214, for example (i.e., students who have achieved the basic level), is represented as the part of the distribution to the right of the point where the population profile crosses the basic threshold. For example, the curve crosses the basic achievement level at about the 24th percentile, which means that 24 percent of the student population scores below the basic level, while 76 percent scores at or above the basic level. Similarly, 32 percent of the population meets the proficient standard (scores at or above 249), and 5 percent of the population meets the advanced standard (scores at or above 282).

- The scale of achievement is the NAEP scale, ranging from 0 to 500; achievement ranges from less than 200 in the lowest 10 percent of the population to above 275, in the top 10 percent of the population.
- In the middle range of the population, from the 20th percentile to the 80th percentile, each percent of the population averages about 1 point on the NAEP scale higher than the next lower percent. At the extremes, where the slopes of the curve are steeper, the variation in achievement between adjacent percentages of the population is much greater.

Figure 1. Distribution of NAEP mathematics scale scores for the nation's public school students at grade 4, with NAEP basic, proficient, and advanced thresholds: 2003

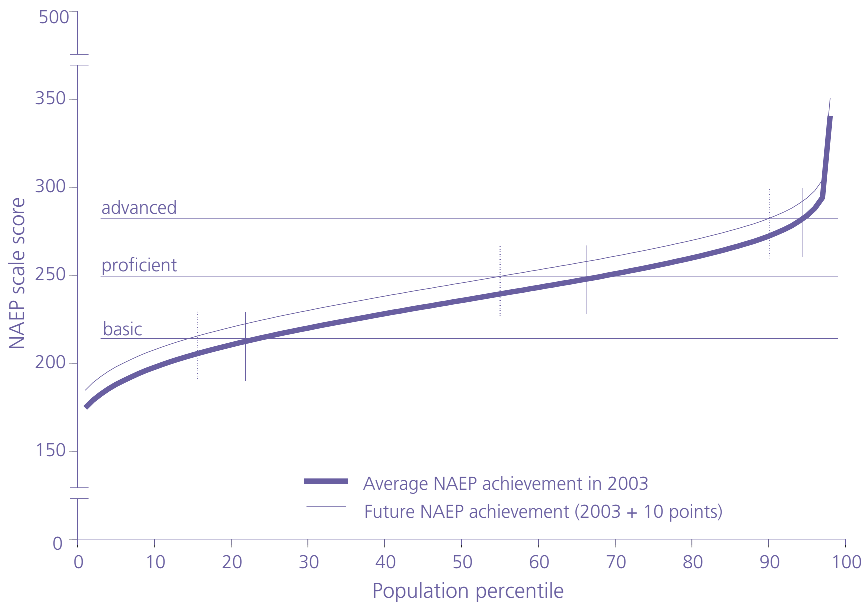


NOTE: Each point on the curve is the expected scale score for the specified percentile of the student population.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates.

Next we suppose that as a result of educational reform, everybody's mathematics achievement improves by 10 points on the NAEP scale. We can superimpose this hypothetical result on the population profile in figure 1, creating the comparison profile in figure 2. At each percentile of the population, the score in the hypothetical future is 10 points higher than in 2003. In the middle of the distribution, this is equivalent to a gain of about 13 percentile points (e.g., a student at the median in the future would be achieving at a level achieved by the 63rd percentile of students in 2003, or in other words, 50 percent of the future population would be achieving at levels only reached by 37 percent of students in 2003). Again, the NAEP basic, proficient, and advanced achievement thresholds are superimposed on the population profile.

As expected, the hypothetical profile of future achievement crosses the achievement thresholds at different points on the achievement continuum. In terms of percentages of students meeting standards, an additional 9 percent are above the basic cutpoint and an additional 13 percent are above the proficient cutpoint, but only 5 percent more are above the advanced cutpoint. Where the standard is set determines the gain in the percentage of the population reported to be achieving the standard. Percentage gains would appear to be twice as large for standards set in the middle of the distribution as for standards set in the tails of the distribution.

Figure 2. Distribution of NAEP mathematics scale scores for the nation's public school students at grade 4: 2003 and hypothetical future



NOTE: Each point on the curve is the expected scale score for the specified percentile of the student population.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates.

This is important in comparing NAEP and state assessment results.[14] If NAEP's proficiency standard is set at a point in an individual state's distribution where achievement gains have small effects on the percentage meeting the standard, and if the state's proficiency standard is set at a point in the state's distribution where the same achievement gains have a relatively large effect on the percentage meeting the standard, then a simple comparison of percentages might find a discrepancy between NAEP and state assessment gains in percentages meeting standards when there is really no discrepancy in achievement gains.

The same problem affects measurement of gaps in achievement in terms of percentages meeting a standard. NAEP might find the poverty gap in a state to be larger than the state assessment reports merely due to differences in the positions of the state's and NAEP's proficiency standards relative to the state's population profiles for students in poverty and other students. And the problem is compounded in measurement of trends in gaps, or gap reduction.[15]

14. Figure 1 is the distribution for the entire nation. The population profiles for individual states vary, although the NAEP cutpoints remain constant for all states.

15. In this report, our interest is that variations in standards can distort comparisons between NAEP and state assessment gaps and trends. However, the same problem distorts comparisons of trends in percentages meeting standards between states.

The solution for implementing comparisons between NAEP and state assessment results is to make the comparisons at the same standard. This is possible if we can determine the point on the NAEP scale corresponding to the cutpoint for the state's standard. NAEP data can easily be re-scored in terms of any specified standard's cutpoint. The percentage of NAEP scale scores (*plausible values)* greater than the cutpoint is the percentage of the population meeting the standard.

NAEP Scale Equivalents

The method for determining the NAEP scale score corresponding to a state's standard is a straightforward equipercentile mapping. In nearly every public school participating in NAEP, a percentage of students meeting the state's achievement standard on its own assessment is also available. The percentage reported in the state assessment to be meeting the standard in each NAEP school is matched to the point in the NAEP achievement scale corresponding to that percentage. For example, if the state reports that 55 percent of the students in fourth grade in a school are meeting their achievement standard and 55 percent of the estimated NAEP achievement distribution in that school lies above 230 on the NAEP scale, then the best estimate from that school's results is that the state's standard is equivalent to 230 on the NAEP scale.[16] These results are aggregated over all of the NAEP schools in a state to provide an estimate of the NAEP scale equivalent of the state's threshold for its standard. The specific methodology is described in appendix A.

A strength and weakness of this method is that it can be applied to any set of numbers, whether or not they are meaningfully related. To ensure scores are comparable, after determining the NAEP scale equivalents for each state standard, we return to the results for each NAEP school and compute the discrepancy between (a) the percentage meeting the standard reported by the state for that school and (b) the percentage of students meeting the state standard estimated by NAEP data for that school. If the mapping were error-free, these would be in complete agreement; however, some discrepancies will arise from random variation. This discrepancy should not be noticeably larger than would be accounted for by simple random sampling variation. If it is noticeably larger than would be expected if NAEP and the state assessment were parallel tests, then we note that the validity of the mapping is questionable—that is, the mapping appears to apply differently in some schools than in others. As a criterion for questioning the validity of the placement of the state standard on the NAEP scale, we determine whether the discrepancies are sufficiently large to indicate that the NAEP and state achievement scales have less than 50 percent of variance in common.[17]

On the following pages, figures 3 and 4 display the NAEP scale score equivalents of primary grade 4 and grade 8 mathematics achievement standards in 45 states and the

16. The school's range of plausible achievement scale values for fourth grade is based on results for its NAEP sample of students.

District of Columbia.[18] In both grades the NAEP equivalents of the states' primary standards ranged from well below the NAEP basic level to slightly above the NAEP proficient level. In both grades, the median state primary standard was between the NAEP basic and proficiency thresholds.

The horizontal axis in figures 3 and 4 indicates the relative error criterion–the ratio of the errors in reproducing the percentages meeting standards in the schools based on the mapping to the size of errors expected by random measurement and sampling error if the two assessments were perfectly parallel. A value of 1.0 for this relative error is expected, and a value greater than 1.5 suggests that the mapping is questionable.[19] The numeric values of the NAEP scale score equivalents for the primary standards displayed in figures 3 and 4, as well as other standards, appear in tables B-1 and B-3 in Appendix B.

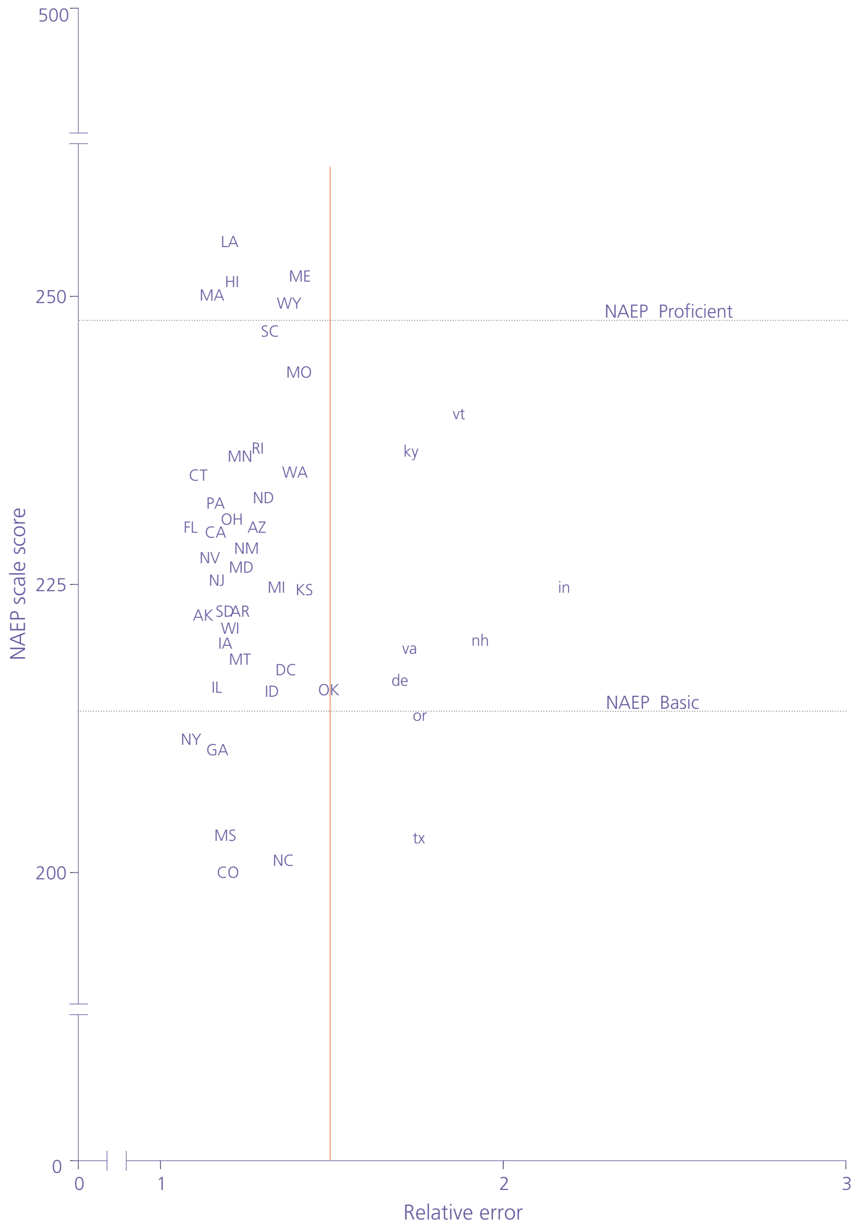
Eight of the 46 grade 4 mathematics standards have relative errors greater than 1.5, as indicated by their position to the right of the vertical line in the figure, and they are displayed in lower case letters in figure 3, indicating that the variation in results for individual schools was large enough to call into question the use of these equivalents. In six of these eight states (Delaware, Indiana, Kentucky, New Hampshire, Oregon, and Virginia), the state assessment results are for grade 3 or 5, so the comparison with NAEP fourth grade results is indirect. The grade discrepancy appears to be a more severe problem for mapping mathematics than for reading, possibly because the elementary school mathematics curriculum has more grade-specific learning objectives than reading. The problem with the mapping for Texas relates to a restriction of range: at 88 percent passing, it was the most extreme of the states' primary standards, leaving relatively little room for reliable measurement of achievement differences between schools. The other state for which the grade 4 mapping is questionable is Vermont.

17. This criterion is different from the usual standard error of equipercentile equating, which is related to the coarseness of the scales, not their correlation. With the relative error criterion we assessed the extent to which the error of the estimate is larger than it would be if NAEP and the state assessment were testing exactly the same underlying trait; in other words, by evaluating the accuracy with which each school's reported percentage of students meeting a state standard can be reproduced by applying the linkage to NAEP performance in that school. The method of estimation discussed in appendix A ensures that, on average, these percentages match, but there is no assurance that they match for each school. To the extent that NAEP and the state assessment are parallel, the percentages should agree for each school, but if NAEP and the state assessment are not correlated, then the mapping will not be able to reproduce the individual school results.

18. No percentages meeting mathematics achievement standards were available for this report for Alabama, Nebraska, Tennessee, Utah, and West Virginia.

19. The computation on which this distinction is made is described in appendix A.

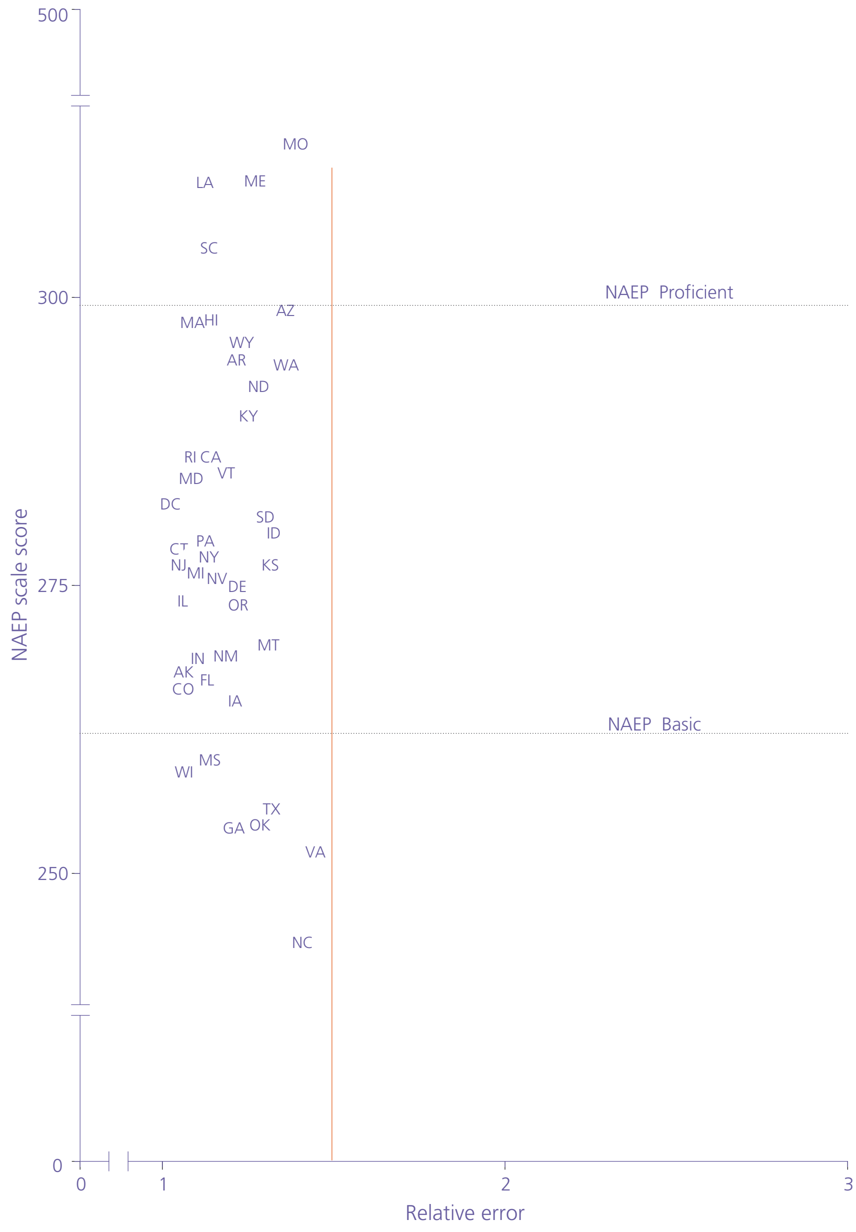
Figure 3. NAEP scale equivalents of primary state mathematics achievement standards, grade 4 or adjacent grade, by relative error criterion: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error. Standards for the eight states displayed in lowercase letters to the right of the vertical line have relative errors greater than 1.5; the variation in results for individual schools in these states is large enough to call into question the use of these equivalents.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 4. NAEP scale equivalents of primary state mathematics achievement standards, grade 8 or adjacent grade, by relative error criterion: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Primary state standards for grade 8 mathematics are displayed in figure 4. For three states for which grade 4 data were available, we did not have data available for grade 8 comparisons.[20] The mappings for the remaining 43 states all appear to be acceptable.

Because this is an initial application of the relative error criterion for evaluating the validity of mapping state mathematics achievement standards to the NAEP scale, we have included the states for which our mappings are questionable in the comparison analyses. However, findings of differences between NAEP and state assessment results for trends and gaps should not be surprising given the quality of the mapping.

The thresholds for these primary state mathematics standards range from below the NAEP basic threshold (e.g., North Carolina and Georgia) to above the NAEP proficient threshold (e.g., Louisiana and Maine); this variation can have profound effects on the percentages of students states find to be meeting their standards. Focusing on the primary mathematics achievement standards, we can ask:

- How variable are the standards from one state to another?
- How is variability of standards related to the percentages of students meeting them?
- How is variation among standards related to the performance of students on NAEP?

In a broader arena, most states have set multiple standards, or achievement levels, and it may be of value to examine the variation in their placement of all levels in relation to the NAEP scale and to their student populations.

- Is there a pattern in the placement of standards relative to expected student performance?

These questions are addressed in the following pages.

How variable are the performance standards from one state to another?

In order to interpret information about the percentage of students meeting one state's standard and compare it to the percentages of students in other states meeting those other states' standards, it is essential to know how the standards relate to each other. Although many of the standards are clustered between the NAEP basic and proficient thresholds, there is great variability. The primary standards range from approximately the 15th to the 80th percentile of the NAEP mathematics achievement distribution. Thus it should not be surprising to find reports that in some states 70 percent of students are meeting the primary standard while 30 percent of students in other states are meeting their states' primary standards, but the students in the latter states score higher on NAEP. Such a result does not necessarily indicate

20. Grade 8 state mathematics assessment data were not available for Minnesota, New Hampshire, and Ohio.

that schools are teaching differently or that students are learning mathematics differently in the different states; it may only indicate variability in the outcomes of the standard setting procedures in the different states.

The variability of the NAEP scale equivalents of the states' primary standards is summarized in table 3. The standard deviations of 13.7 and 16.7 NAEP points among states' primary standards can be translated into the likelihood of finding contradictory assessment results in different states. To see this concretely, imagine a set of students who take one state's mathematics assessment and then another state's mathematics assessment. How different would the percentage of these students meeting the two states' standards be? In some pairs of states, with standards set at the same level of difficulty, we would expect only random variation, but in extreme cases, such as among fourth graders in Louisiana and North Carolina, the difference might be 50 percent (i.e., of a nationally representative sample of students, 50 percent more would appear to show *consistent mastery* in North Carolina than would appear to demonstrate *mastery* in Louisiana). On average, for any random pair of states, this discrepancy would be about 15 percentage points. That is, among sets of students in two randomly selected states who are actually achieving in mathematics at the same level, about 15 percent would be classified differently as to whether they were meeting the state's primary mathematics standard in the two states.

Table 3. Mean and standard deviation of primary mathematics standard cutpoints across states, by grade: 2003

Level	Number of states	Average cutpoint	Standard error	Standard deviation	Standard error of standard deviation
Grade 4	46	227.0	0.16	13.7	0.17
Grade 8	43	229.3	0.18	16.7	0.18

NOTE: Primary standard is the state's standard for *proficient* performance.

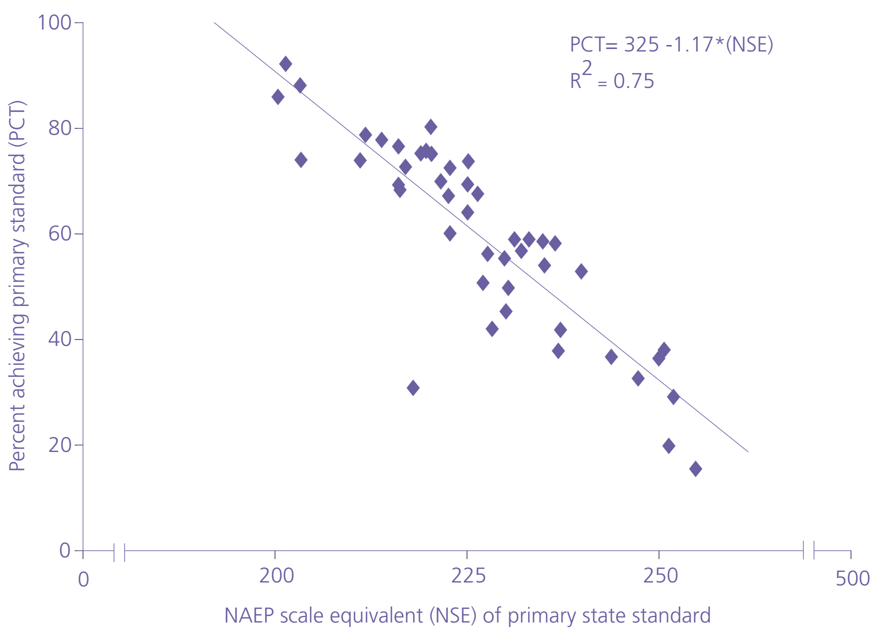
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

How is variability of performance standards related to the percentages of students meeting them?

Is it possible that states are setting standards in relation to their particular student populations, with higher standards set in states where mathematics achievement is higher? Perhaps one could imagine that public opinion might lead each state education agency to set a standard to bring all students up to the level currently achieved by the median student in its state. Then variation in standards would just be a mirror of variation in average achievement among the states. If that is not the case, then we should expect to see a negative relationship between the placement of the standard on the NAEP scale and the percentage of students meeting the standard.

This question is addressed in figures 5 and 6, which graph the relations between the difficulty of meeting each standard, as measured by its NAEP scale equivalent, and the percentage of students meeting the standard. The higher the standard is placed, the smaller the percentage of students in the state meeting the standard. In fact, the negative relation is so strong that for every point of increased NAEP difficulty (which corresponds roughly to one percent of the population, except in the extremes), about one percent (1.17 percent in grade 4 and 1.04 percent in grade 8) fewer students meet the standard. There is clearly much greater variability between states in the placement of mathematics standards than in the mathematics achievement of students: the standard deviations of state mean NAEP scale scores for the states included in this analysis are 7.0 points at grade 4 and 8.9 points at grade 8, compared to the standard deviations of their standards placement of 13.7 points and 16.7 points (table 3).

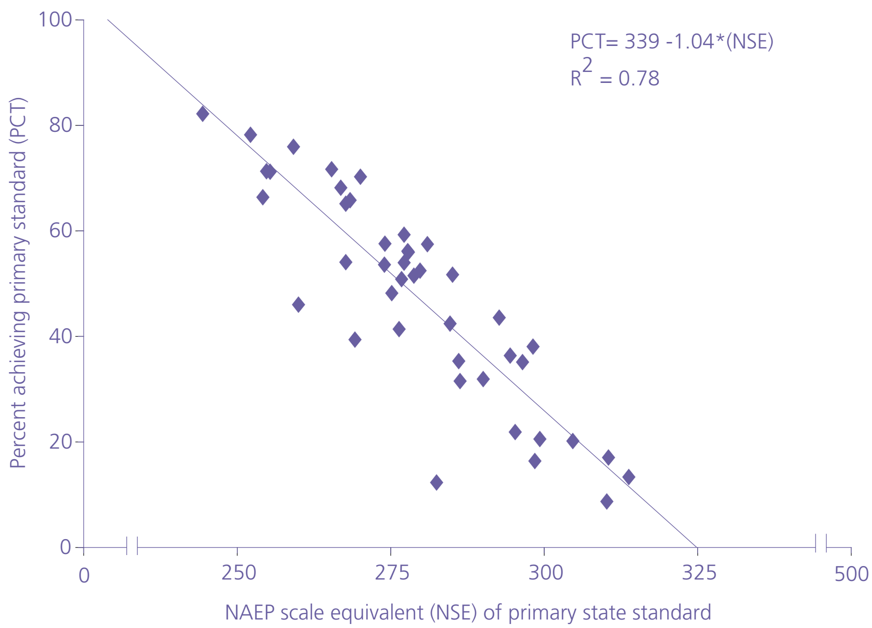
Figure 5. Relationship between the NAEP equivalents of grade 4 primary state mathematics standards and the percentages of students meeting those standards: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. Each diamond in the scatter plot represents the primary standard for one state. The relationship between the NAEP scale equivalent of grade 4 primary state mathematics standards (NSE) and the percentages of students meeting those standards in a state (PCT) is estimated over the range of data values by the equation PCT = 325 - 1.17(NSE). In other words, a one point increase in the NAEP difficulty implies 1.17 percent fewer students meeting the standard. For example, the 220 point on the NAEP scale equivalent represents approximately 67.6 percent of students achieving primary standard (67.6 = 325 - 1.17(220)) and at 221 on the same scale indicates 66.4 percent (325 - 1.17(221) = 66.4).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 6. Relationship between the NAEP equivalents of grade 8 primary state mathematics standards and the percentages of students meeting those standards: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. Each diamond in the scatter plot represents the primary standard for one state. The relationship between the NAEP scale equivalent of grade 8 primary state mathematics standards (NSE) and the percentages of students meeting those standards (PCT) is estimated over the range of data values by the equation

PCT = 339 - 1.04(NSE). In other words, a one point increase in the NAEP difficulty implies 1.04 percent fewer students meeting the standard. For example, the 270 point on the NAEP scale equivalent represents approximately 58.2 percent of students achieving primary standard (58.2 = 339 - 1.04(270)) and at 271 on the same scale indicates 57.2 percent (339 - 1.04(271) = 57.2).

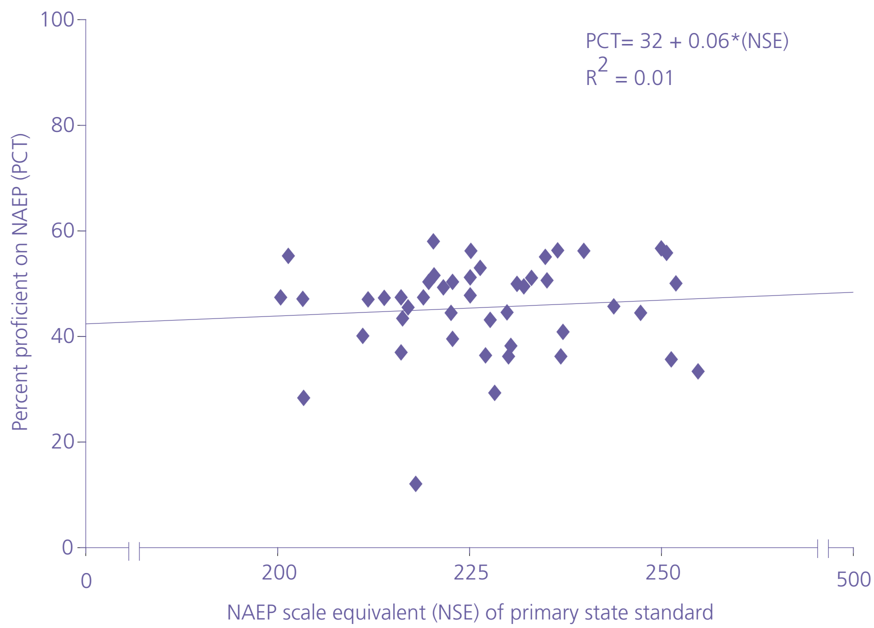
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

How is variation among performance standards related to the performance of students on NAEP?

Does setting high standards lead to higher achievement? Finding out whether it does must await the accumulation of trend information over time, but the relation between the difficulty level of a state's primary mathematics standard and the performance of that state's students on the NAEP mathematics assessment is relevant. This question is addressed in figures 7 and 8, which display the percentage of each state's students meeting the NAEP proficient standard as a function of the placement of their own primary mathematics standard.

These graphs present a stark contrast to the relations shown in figures 5 and 6. In 2003, there was virtually no relation between the level at which a state sets its primary mathematics standard and the mathematics achievement of its students on NAEP. In most states, between 30 and 60 percent of students meet the NAEP proficient standard, and that percentage is no higher among states that set high primary standards than among states that set low primary standards.

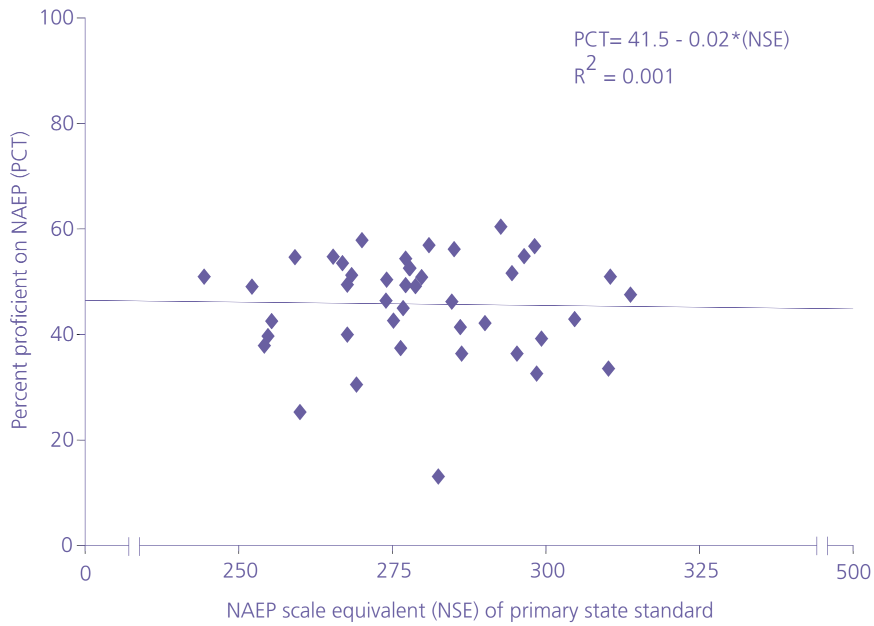
Figure 7. Relationship between the NAEP equivalents of grade 4 primary state mathematics standards and the percentages of students meeting the NAEP mathematics proficiency standard: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. The relationship between the NAEP scale equivalent of grade 4 primary state mathematics standards (NSE) and the percentages of students meeting NAEP mathematics proficiency standard (PCT) is estimated over the range of data values by the equation PCT = 32 + 0.06(NSE). There is virtually no relation between the level at which a state sets its primary mathematics standard and the mathematics achievement of its students on NAEP.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 8. Relationship between the NAEP equivalents of grade 8 primary state mathematics standards and the percentages of students meeting the NAEP mathematics proficiency standard: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. The relationship between the NAEP scale equivalent of grade 8 primary state mathematics standards (NSE) and the percentages of students meeting NAEP mathematics proficiency standard (PCT) is estimated over the range of data values by the equation PCT = 41.5 - 0.02(NSE). There is virtually no relation between the level at which a state sets its primary mathematics standard and the mathematics achievement of its students on NAEP.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Is there a pattern in the placement of a state's performance standards relative to the range of student performance in the state?

As we saw in figures 5 and 6, the placement of the standards can have consequences for the ability to demonstrate school-level gains. It is therefore useful to see where states are setting their standards, single and multiple alike. The scatter plots in figures 9 and 10 extend the charts of primary standards shown in figures 3 and 4 to show the entire range of 130 grade 4 and 120 grade 8 state mathematics standards. In these scatter plots, standards higher than the primary standard are shown as plus/minus signs, primary standards as open/filled diamonds, and lower standards as open/filled circles. The 34 grade 4 standards and 5 grade 8 standards that have sufficiently high relative errors to question the validity of the mapping are indicated by dashes and unfilled diamonds and circles, and are to the right of the vertical line in each figure. Grade 8 mathematics standards were more easily equated than grade 4 mathematics standards.

But how is this variability related to the student populations in the states? This question is addressed in an exploratory manner in figures 11 and 12, which display the frequencies of standards met by differing percentages of the population.[21] Thus, for example, the relatively easiest 14 standards for grade 4 were achieved by more than 90 percent of the students in their respective states, and the highest 11 standards were achieved by fewer than 10 percent of the students (figure 11).[22] At grade 8, five standards were achieved by more than 90 percent of the students in their respective states, while 14 were achieved by fewer than 10 percent (figure 12).

Standards for grade 4 (or grades 3 or 5, where there is no grade 4 state mathematics assessment) are set at every level, from very easy (more than 90 percent passing) to very difficult (fewer than 10 percent passing). The same is true for standards for grade 8 (or grade 7, where there is no grade 8 state mathematics assessment), although a greater percentage of the grade 8 standards are set to be difficult for eighth graders to pass: 38 of 120 are set where they are passed by fewer than 20 percent of eighth graders, compared to 28 of 130 grade 4 standards.[23] NAEP basic, proficient, and advanced mathematics achievement levels, by comparison, are met by about 70 percent (77 percent at grade 4 and 68 percent at grade 8), 30 percent, and 5 percent, respectively, of students nationally.

We conclude this section on state standards by pointing out the assumptions made in these comparisons. The major assumption is that the state assessment results are correlated with NAEP results—although the tests may look different, it is the correlation of their results that is important. If NAEP and the state assessment identify the same pattern of high and low achievement across schools in the state, then it is meaningful to identify NAEP scale equivalents of state assessment standards. The question of correlation is discussed in the next section.

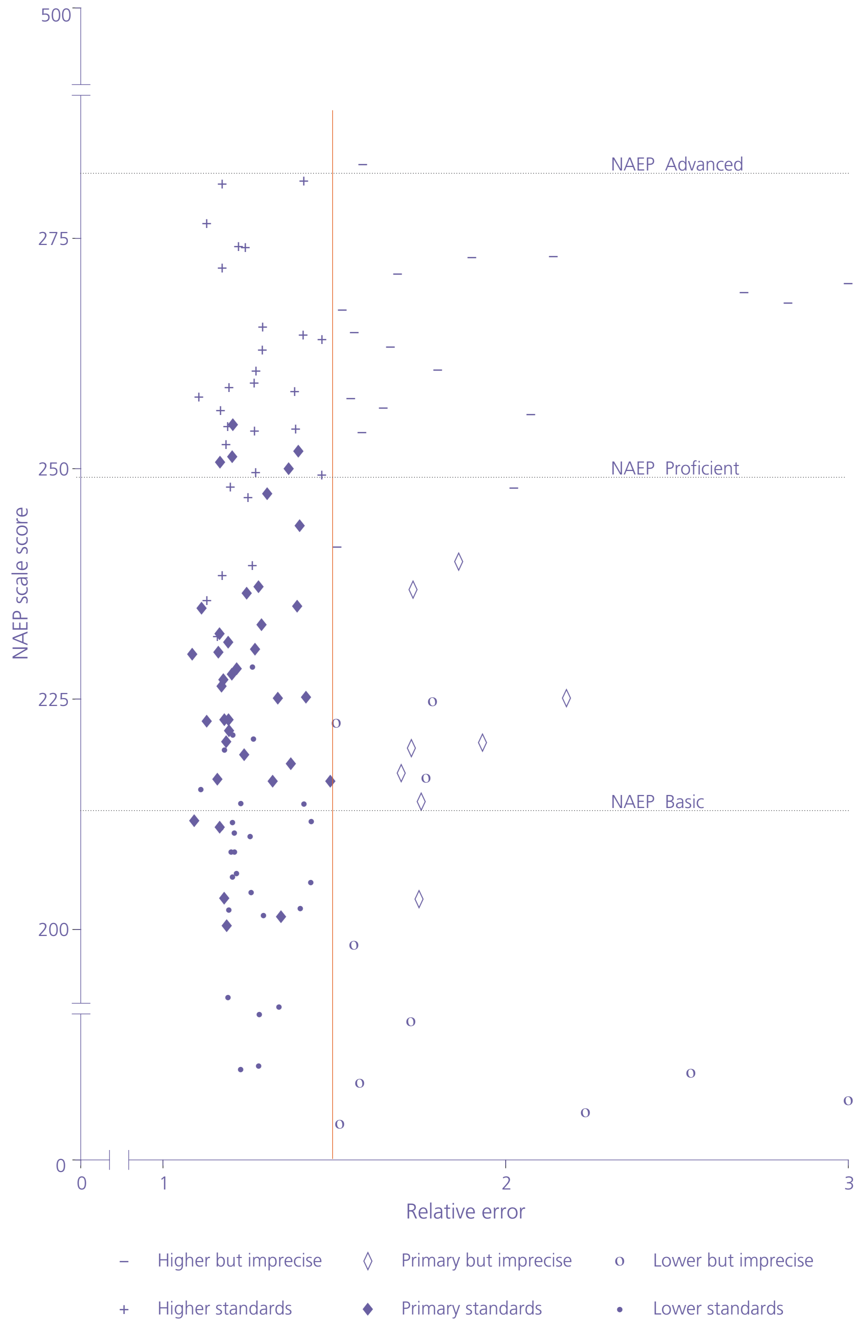
The other important assumption is that the assessments are measuring the same population, in the same way. If substantial numbers of students participate in one of the assessments but not the other, this can have a biasing effect on the standard comparison. While we cannot account for state assessment non-participation in this comparison, we do account for NAEP non-participation by use of weighting and imputation of achievement of excluded students (see appendix A for a discussion of the imputation).

21. The grade 4 and grade 8 standards include a few that are for adjacent grades, as indicated in table 4, below.

22. If most students in a state can pass a performance standard, the standard must be considered relatively easy, even if fewer students in another state might be able to pass it.

23. On a simple test of proportions, the probability of this pattern is less than .05, but inferring statistical significance from this would be inappropriate because there was no *a priori* hypothesis of finding this pattern.

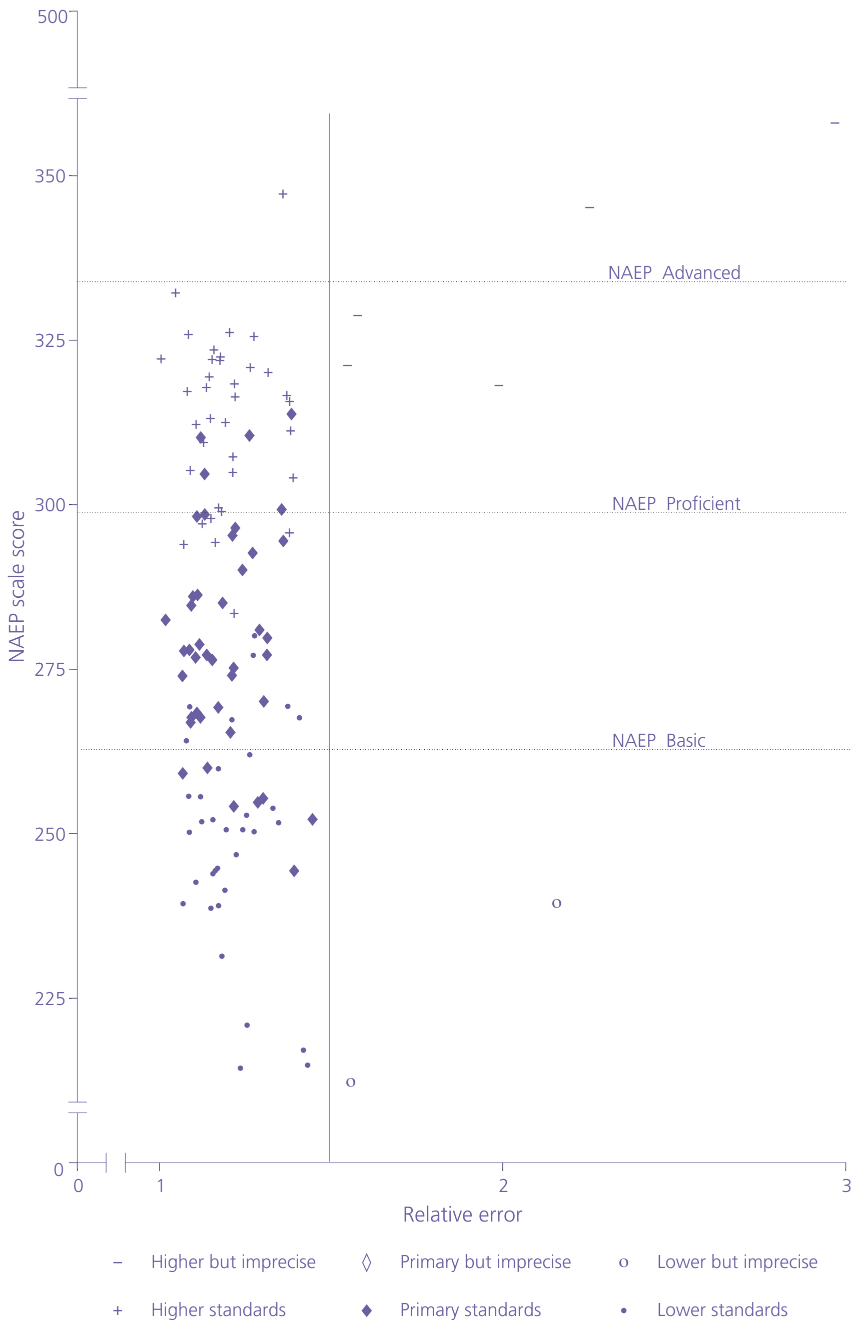
Figure 9. NAEP equivalents of state grade 4 primary mathematics achievement standards, including standards higher and lower than the primary standards, by relative error criterion: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

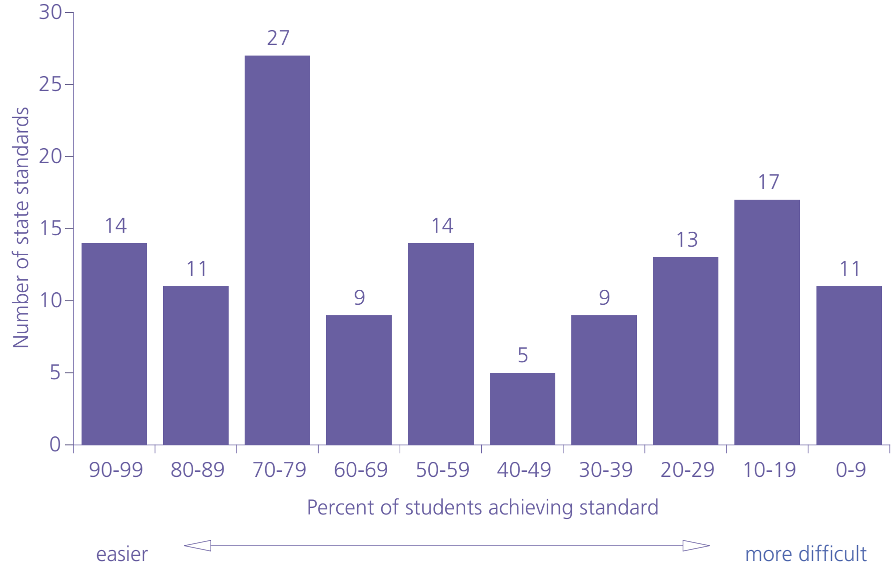
Figure 10. NAEP equivalents of state grade 8 primary mathematics achievement standards, including standards higher and lower than the primary standards, by relative error criterion: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. Relative error is a ratio measure of reproducibility of school-level percentages meeting standards, described in appendix A. The vertical line indicates a criterion for maximum relative error.

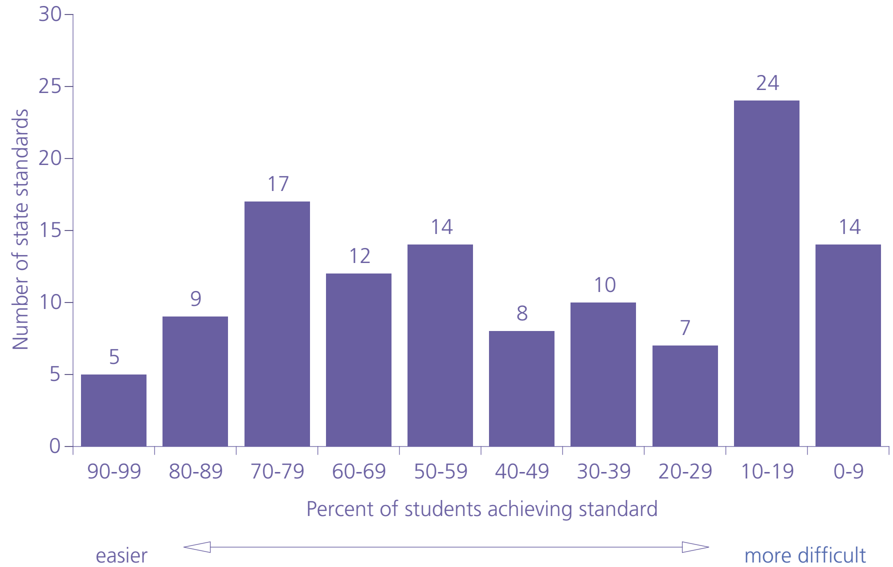
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment: Full population estimates. The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 11. Number of state mathematics standards by percentages of grade 4 students meeting them: 2003



SOURCE: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Figure 12. Number of state mathematics standards by percentages of grade 8 students meeting them: 2003



SOURCE: The National Longitudinal School-Level State Assessment Score Database (NLSLSASD) 2004.

Finally, there is the issue of accommodations, or non-standard test administrations, provided for some students with disabilities and English language learners. It is not known at present how these accommodations (e.g., extended time and one-on-one testing) affect the distribution of assessment scores.

Summary

By matching percentages of students reported to be meeting state standards in schools participating in NAEP with the distribution of performance of students in those schools on NAEP, cutpoints on the NAEP scale can be identified that are equivalent to the state standards. The accuracy of the determination of the NAEP equivalent of the standard depends on the correlations between NAEP and state assessment results. Most states have multiple standards, and these can be categorized into *primary standards*, which are generally the standards used for reporting adequate yearly progress, and standards that are above or below the primary standards. In most states, the primary standards are referred to as *proficient* or *meets the standard*.

In the majority of the states examined, the standards were sufficiently correlated to warrant reporting the NAEP equivalents of standards. At grade 4, 8 of the 46 primary standards were judged to be questionable; none were at grade 8. The mapping of state standards to the NAEP scale is an essential step in comparing achievement trends and gaps as measured by NAEP and state assessments.

The primary standards vary significantly in difficulty, as reflected in their NAEP equivalents. On average, for any two randomly selected states, about 15 percent of the students who meet the primary standard in one state would not meet the standard in the other state; between some states, the disparity is much larger.

As might be expected, the higher the primary standard is set, the fewer the students who meet it. Between states, there is more variability in the setting of standards than in actual student achievement. On the other hand, students in states with high primary standards score just about the same on NAEP as students in states with low primary standards.