3 Comparing Schools' Percentages Meeting State Standards

A fundamental question is whether state assessments in different states would identify the same schools as having high and low mathematics achievement. No state assessments are administered exactly the same way, with the same content, across state lines, so analysts cannot answer this question directly. However, NAEP provides a link, however imperfect, to address this question. If the pattern of NAEP results matches the pattern of state assessment results across schools in each of two different states, then either of those two states' assessments would likely identify the same schools as having students who are good in mathematics, compared to other schools in their respective states.

The correlation coefficient is the standard measure of the tendency of two measurements to give the same results, varying from +1 when two measurements give functionally identical results, to 0 when they are completely unrelated to each other, to –1 when they represent opposites. A high correlation (near +1) between two mathematics achievement tests means that schools (or students) whose performance is relatively high on one test also demonstrate relatively high performance on the other test. It does not mean that the two tests are necessarily similar in content and format, only that their results are similar. And it is the results of the tests that are of concern for accountability purposes.

To compute a correlation coefficient, one needs the results of both tests for the same schools (or students). State assessment statistics are available at the school level, and NAEP data can be aggregated from student records to create school-level statistics for the same schools.[24] Therefore, the correlations presented in this report are of school-level statistics, and a high correlation indicates that two assessments are identifying the same schools as high scoring (and the same schools as low scoring).[25]

24. NAEP does not report individual school-level statistics because the design of NAEP precludes measurement of school-level statistics with sufficient accuracy (reliability) to justify public release. However, for analytical purposes, aggregating these school-level statistics to state-level summaries provides reliable state-level results (e.g., correlations between NAEP and state assessment results).

25. The value of a correlation coefficient at the student level and one at the school level need not be the same, even though they are based on the same basic data. The student level correlation will tend to be somewhat lower because it does not give as much weight to systematic variation in distributions of higher and lower achieving students to different schools. Because policy analysts are interested in systematic variation between schools, the school-level correlation is the appropriate statistic.

State assessments have traditionally reported scores in a wide variety of units, including percentile ranks, scale scores, and grade equivalent scores, among others, but since 1990 there has been a convergence on reporting in terms of the percentages of students meeting standards (which is translated into the percentages of students earning a score above some *cutpoint*). While this does not present an insurmountable problem for computation of correlation coefficients, it does raise three issues that need to be addressed.

Most important of these is the match between the two standards being correlated. The correlation between the percentage of students achieving a very easy standard (e.g., one which 90 percent of students pass), with the percentage of students achieving a very hard standard (e.g., one which only 10 percent of students pass) will necessarily be lower than the correlation between two standards of matching difficulty. For this reason, the correlations presented in this report are between (a) the school-level percentages meeting a state's standards as measured by its own measurement, and (b) the corresponding percentages meeting a standard of the same difficulty as measured by NAEP. In the preceding chapter, NAEP cutpoints of difficulty equivalent to state standards were identified (in figures 3 and 4), and they are used in this analysis.

The second issue concerns the position of the standards in the achievement distribution even when they are matched in difficulty. Extremely easy or difficult standards necessarily have lower intercorrelations than standards near the median of the population. It is impossible to dictate where a state's standards fall in its achievement distribution, but it is possible to estimate how much the extremity of the standards might affect correlations.

The third issue concerns the fact that percentages meeting standards necessarily hide information about variations in achievement within the subgroup of students who meet the standard (and within the subgroup of students who fail to meet the standard). One might expect this to set limits on the correlation coefficients. However, empirical comparison of correlations of percentages meeting standards near the center of the distribution with correlations of median percentile ranks or mean scale scores has indicated that there is only modest loss of correlational information in using percentages meeting standards near the center of the distribution (MacLaughlin, 2005 and Shkolnik and Blankenship, 2006).

The correlations between the percentage of schools' students meeting the NAEP and the state assessment primary standards are shown in table 4. The selection of the standard and the short name of the standard included in the table are based on interpretation of information on the state's web site. The grade indicated is generally the same as NAEP (4 or 8), but in a few states, scores were available only for grade 3, 5, 7, or 9 (or for E or M, which represent aggregates across elementary or middle grades). The correlations for primary standards range from .44 to .89, with a median of .76, for grade 4 mathematics, and from .62 to .97, with a median of .81, for grade 8 mathematics. The distributions of correlations are shown in figures 13 and 14.

Table 4. Correlations between NAEP and state assessment school-level percentages meeting primary state mathematics standards, grades 4 and 8, by state: 2003

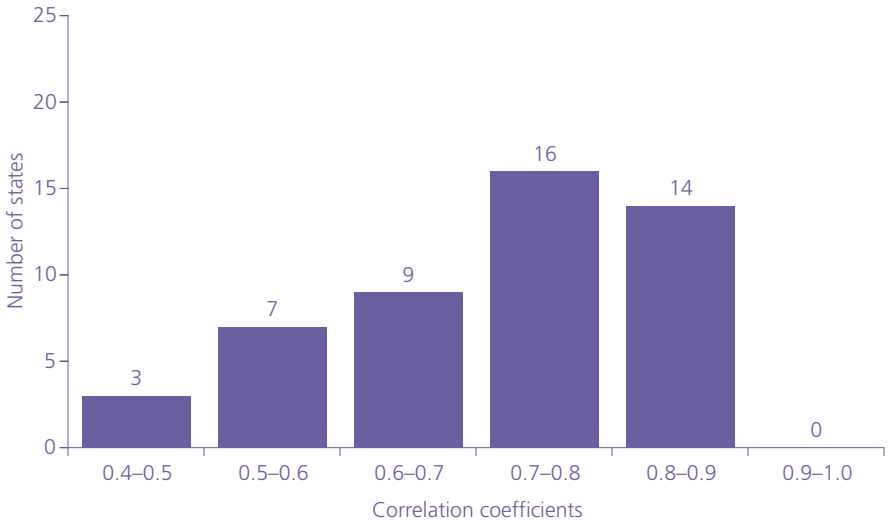
State/ jurisdiction	Name of standard	Grades for state assessment		Grade 4 correlation	Grade 8 correlation
Alabama	Percentile Rank	4	8	0.80	0.84
Alaska	Proficient	4	8	0.78	0.86
Arizona	Meeting	5	8	0.77	0.69
Arkansas	Proficient	4	8	0.81	0.77
California	Proficient	4	7	0.84	0.88
Colorado	Partially Proficient	5	8	0.79	0.87
Connecticut	Goal	4	8	0.89	0.89
Delaware	Meeting	5	8	0.58	0.79
District of Columbia	Proficient	4	8	0.69	0.97
Florida	(3)PartialSuccess	4	8	0.89	0.86
Georgia	Meeting	4	8	0.83	0.80
Hawaii	Meeting	5	8	0.78	0.83
Idaho	Proficient	4	8	0.67	0.70
Illinois	Meeting	5	8	0.84	0.92
Indiana	Pass	3	8	0.44	0.83
Iowa	Proficient	4	8	0.77	0.77
Kansas	Proficient	4	7	0.66	0.72
Kentucky	Proficient	5	8	0.53	0.72
Louisiana	Mastery	4	8	0.79	0.82
Maine	Meeting	4	8	0.56	0.69
Maryland	Proficient	5	8	0.83	0.88
Massachusetts	Proficient	4	8	0.82	0.87
Michigan	Meeting	4	8	0.74	0.87
Minnesota	(3)Proficient	5	—	0.77	—
Mississippi	Proficient	4	8	0.79	0.82
Missouri	Proficient	4	8	0.69	0.62
Montana	Proficient	4	8	0.72	0.72
Nebraska	Meeting	—	—	—	—
Nevada	Meeting:3	4	7	0.81	0.82
New Hampshire	Basic	3	—	0.46	—
New Jersey	Proficient	4	8	0.84	0.90
New Mexico	Top half	4	8	0.77	0.81
New York	Meeting	4	8	0.86	0.85
North Carolina	Consistent Mastery	4	8	0.63	0.71
North Dakota	Meeting	4	8	0.64	0.75
Ohio	Proficient	4	—	0.81	—
Oklahoma	Satisfactory	5	8	0.58	0.71
Oregon	Meeting	5	8	0.51	0.77
Pennsylvania	Proficient	5	8	0.83	0.87
Rhode Island	Proficient	4	8	0.78	0.90
South Carolina	Proficient	4	8	0.74	0.80
South Dakota	Proficient	4	8	0.77	0.71
Tennessee	Percentile Rank	4	8	0.76	0.81
Texas	Passing	4	8	0.52	0.71
Utah	Percentile Rank	5	8	0.68	0.72
Vermont	Achieved	4	8	0.47	0.74
Virginia	Proficient	5	8	0.54	0.63
Washington	Met	4	7	0.69	0.69
West Virginia	Top half	—	—	—	—
Wisconsin	Proficient	4	8	0.81	0.90
Wyoming	Proficient	4	8	0.64	0.74

— Not available.

NOTE: Primary standard is the state's standard for *proficient* performance. In Alabama, Tennessee, and Utah, correlations are based on school-level median percentile ranks. In West Virginia, E and M represent aggregates across elementary and middle grades, respectively.

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.

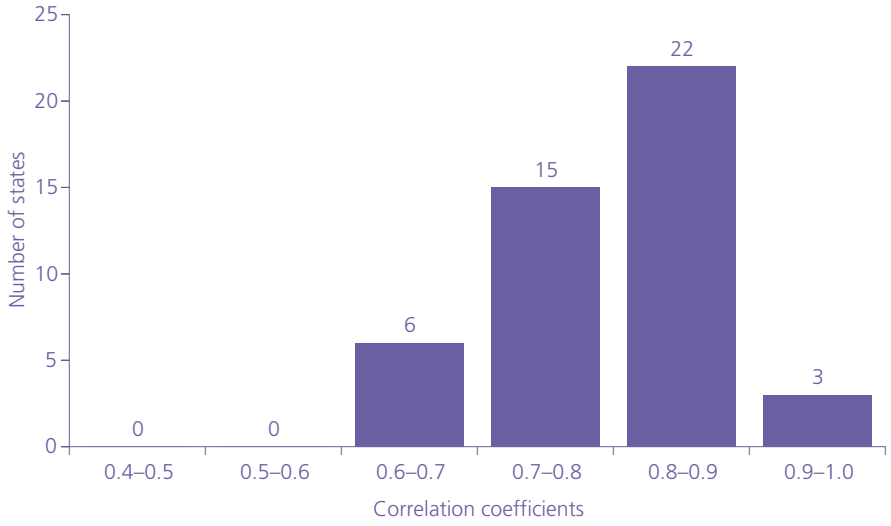
Figure 13. Frequency of correlations between school-level NAEP and state assessment percentages meeting the primary grade 4 state mathematics standard: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations are of median percentile ranks for Alabama, Tennessee, and Utah.

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 14. Frequency of correlations between school-level NAEP and state assessment percentages meeting the primary grade 8 state mathematics standard: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations are of median percentile ranks for Alabama, Tennessee, and Utah.

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.

As an overall criterion, one would like to have correlations greater than .70 for analyses that depend on a linkage between the results of two assessments. In states that do not meet this criterion, i.e., the two assessments have less than 50 percent of common variance, divergences in comparisons of trends and gaps in later sections of this report may reflect the impact of whatever factors cause the correlations to be lower. Because this is a research report, we do not exclude data from states with lower correlations.

There are many sources of variation in correlation coefficients; results presented here can only set a context for in-depth analysis of the differences which analysts may wish to pursue. The tendency to say that "they must be measuring different things" should be resisted, however. Even if the tests were sampling and measuring different parts of the mathematics construct, they still might be highly correlated; that is, they might still identify the same schools as high achieving and low achieving.

The following (non-exhaustive) list of reasons for lower correlations should be considered before selecting any particular interpretation of low correlations.

- Reliability of the measure (the school-level test score)
  - Student sample size in schools (small school suppression may increase correlation)
  - Reliability of the student-level measure
  - Measures from different grades
- Conditions of testing
  - Different dates of testing (including testing in different grades)
  - Different motivation to perform
- Requirements for enabling skills
  - Different response formats (different demands for writing skills)
- Similarity of location of the measure relative to the student population
  - Extreme standards will not be as strongly correlated as those near the median
- Similarity of testing accommodations provided for students with special needs
  - Accommodations given on one test but not the other reduce correlations
- Match of the student populations included in the statistics
  - Representativeness of NAEP samples of students and of schools
  - Extent of student exclusion or non-participation
- Differences in the definition of the target skill (mathematics)
  - States have varying emphasis on computation, problem-solving, and conceptual understanding.

To understand the potential impact of these factors, consider the effects of three factors on the correlations of NAEP and state assessment percentages meeting standards: (1) extremity of the standard, (2) size of the school sample of students on which the percentage is based, and (3) grade level of testing (same grade or adjacent grade). As a meta-analysis of the correlation coefficients, we carried out a linear

regression accounting for variation in correlations for 130 standards in 46 states in grade 4 and 120 standards in 43 states in grade 8.[26] Results are shown in table 5.

Table 5. Standardized regression coefficients of selected factors accounting for variation in NAEP-state mathematics assessment correlations: 2003

Factor	Grade 4	Grade 8
Extreme standards	-0.51 *	-0.54 *
Small school samples	-0.44 *	-0.20 *
Grade difference	-0.27 *	-0.11
Sample size	130	120
R^2	0.47	0.32

* Coefficient statistically significantly different from zero (p<.05)

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.

The values of R^2 were .47 and .32 for grades 4 and 8. That is, these three factors accounted for one-half of the variance of correlations between NAEP and state standards in grade 4 and one-third of the variance in grade 8 correlations. At grade 4, all three predictors were significant, but for grade 8, the effect of the grade difference factor was not significantly different from zero. This may be due to the fact that in only four states did we have to use grade 7 state assessment scores.

Applying the results of the linear regression, one can estimate what each correlation might have been if it were based on a standard set at the student population median, in the same grade as NAEP, and with no school samples of fewer than 30 students. The results are displayed in table 6 and summarized in figures 15 and 16. Nearly all of the adjusted correlations are greater than .70 for grade 8, with a median of .86, and four-fifths of them are at least .70 for grade 4, with a median of .84.

At grade 8, correlations in two states remained less than .70, after adjusting for effects of grade differences, small schools, and extreme standards, although their correlations rounded to .68 (Virginia) and .70 (Idaho). At grade 4, adjusted correlations in nine states were less than .70: Delaware, Indiana, Kentucky, New Hampshire, Oregon, Virginia, Texas, Vermont, and Washington. In the first six of these states, the correlations were based on state assessment scores for grade 3 or 5. It is possible that the adjustment did not capture all of the effects of that factor. The factors affecting the correlation coefficients in the other three states are not known at this time.

26. All state standards, not merely the primary one for each state, were included. The specific predictors were: (1) $(d/50)^4$, where d was the difference between the average percentage meeting the standard and 50 percent; (2) the maximum of 0 and the amount by which the average NAEP school's student sample size was less than 30 (34 was the largest average school sample size in grade 4); and (3) a dichotomy, 1 if the tested grade was not 4 or 8, 0 if it was 4 or 8.

Table 6. Adjusted correlations between NAEP and state assessment school-level percentages meeting primary state mathematics standards, grades 4 and 8, by state: 2003

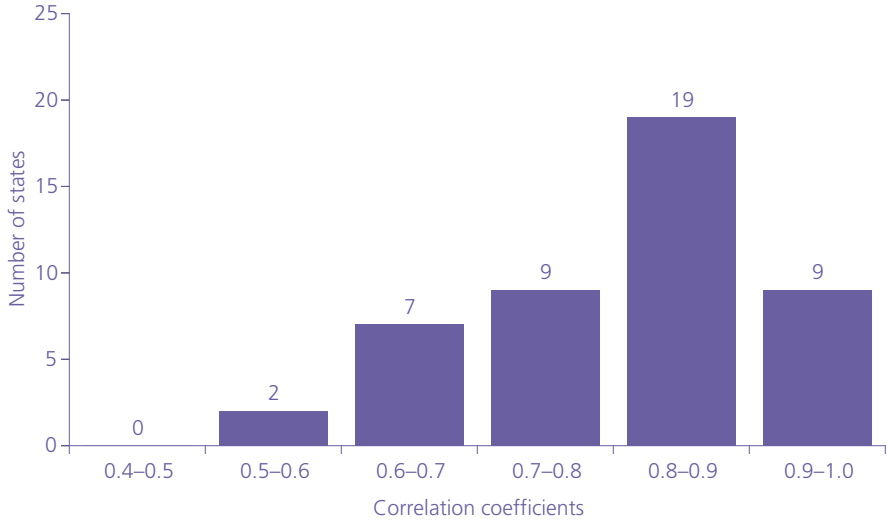
State/ jurisdiction	Name of standard	Grades for state assessment		Grade 4 correlation	Grade 8 correlation
Alabama	Percentile Rank	4	8	—	—
Alaska	Proficient	4	8	0.98	0.90
Arizona	Meeting	5	8	0.87	0.79
Arkansas	Proficient	4	8	0.84	0.86
California	Proficient	4	7	0.84	0.94
Colorado	Partially Proficient	5	8	1.00	0.93
Connecticut	Goal	4	8	0.89	0.91
Delaware	Meeting	5	8	0.69	0.79
District of Columbia	Proficient	4	8	0.79	1.00
Florida	(3)Partial Success	4	8	0.89	0.89
Georgia	Meeting	4	8	0.84	0.81
Hawaii	Meeting	5	8	0.93	0.92
Idaho	Proficient	4	8	0.73	0.70
Illinois	Meeting	5	8	0.93	0.96
Indiana	Pass	3	8	0.55	0.85
Iowa	Proficient	4	8	0.88	0.82
Kansas	Proficient	4	7	0.80	0.83
Kentucky	Proficient	5	8	0.64	0.77
Louisiana	Mastery	4	8	0.91	1.00
Maine	Meeting	4	8	0.75	0.77
Maryland	Proficient	5	8	0.92	0.91
Massachusetts	Proficient	4	8	0.86	0.88
Michigan	Meeting	4	8	0.76	0.92
Minnesota	(3)Proficient	5	—	0.86	—
Mississippi	Proficient	4	8	0.80	0.86
Missouri	Proficient	4	8	0.71	0.76
Montana	Proficient	4	8	0.97	0.82
Nebraska	Meeting	—	—	—	—
Nevada	Meeting:3	4	7	0.81	0.86
New Hampshire	Proficient	3	—	0.65	—
New Jersey	Proficient	4	8	0.84	0.93
New Mexico	Top half	4	8	0.85	0.81
New York	Meeting	4	8	0.89	0.90
North Carolina	Consistent Mastery	4	8	0.80	0.76
North Dakota	Meeting	4	8	0.90	0.86
Ohio	Proficient	4	—	0.81	—
Oklahoma	Satisfactory	5	8	0.78	0.79
Oregon	Meeting	5	8	0.68	0.81
Pennsylvania	Proficient	5	8	0.93	0.89
Rhode Island	Proficient	4	8	0.79	0.90
South Carolina	Proficient	4	8	0.74	0.87
South Dakota	Proficient	4	8	0.98	0.79
Tennessee	Percentile Rank	4	8	—	—
Texas	Passing	4	8	0.60	0.71
Utah	Percentile Rank	5	8	—	—
Vermont	Achieved	4	8	0.69	0.78
Virginia	Proficient	5	8	0.66	0.68
Washington	Met	4	7	0.69	0.77
West Virginia	Top half	—	—	—	—
Wisconsin	Proficient	4	8	0.90	0.96
Wyoming	Proficient	4	8	0.87	0.75

— Not available.

NOTE: Primary standard is the state's standard for *proficient* performance. For Alabama, Tennessee, and Utah, adjusted correlations could not be estimated since percentages meeting standards were not available.

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.

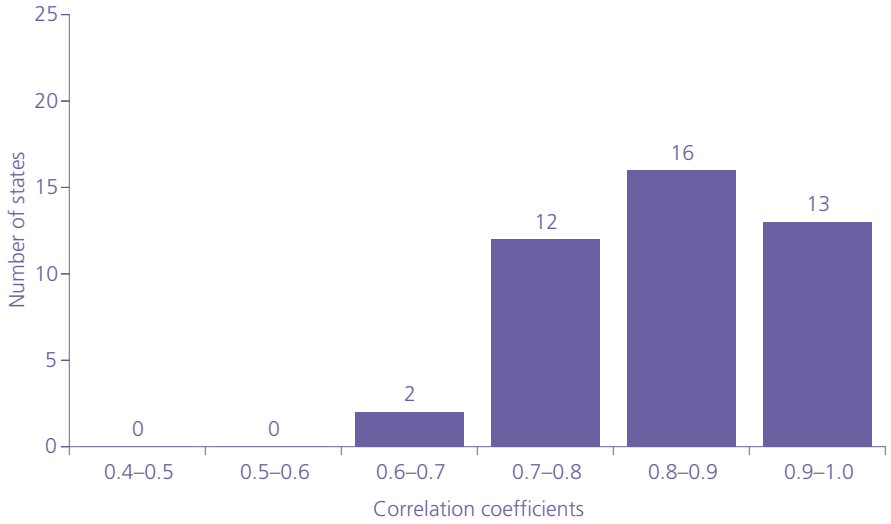
Figure 15. Frequency of adjusted correlations between school-level NAEP and state assessment percentages meeting the primary grade 4 state mathematics standard: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations of median percentile ranks for Alabama, Tennessee, and Utah are not included.

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 16. Frequency of adjusted correlations between school-level NAEP and state assessment percentages meeting the primary grade 8 state mathematics standard: 2003



NOTE: Primary standard is the state's standard for *proficient* performance. No correlations lie on the boundaries of the categories. Correlations of median percentile ranks for Alabama, Tennessee, and Utah are not included.

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.

The high adjusted correlations between NAEP and state assessment measures of the percentages of schools' students meeting mathematics achievement standards indicate that in most states, NAEP and state assessments in 2003 were in general agreement about which schools have high and low mathematics achievement. Nevertheless, the findings of relatively low correlations in a few states need to be considered in interpreting results of gap and trend comparisons of trends and gaps as reported by NAEP and state assessments. Gaps and trends may be similar, in spite of low correlations, but when gaps or trends differ significantly, the reasons for the low correlations require further study.

SUMMARY

An essential criterion for the comparison of NAEP and state assessment results in a state is that the two assessments agree on which schools are high achieving and which are not. The critical statistic for testing this criterion is the correlation between schools' percentages achieving the standard, as measured by NAEP and the state assessment.

In 2003, correlations between NAEP and state assessment measures of mathematics achievement were greater than .70 in 30 of 49 states for grade 4 and 41 of 46 states for grade 8. An analysis of the correlations focused on two methodological factors that tend to depress some of these correlations: (1) small enrollments in schools limit the reliability of percentages of students meeting a standard; and (2) standards set either very high or very low tend to be less correlated than standards set near the middle of a state's achievement distribution. Estimates of what the correlations would be if they were all based on scores on non-extreme standards in the same grade in schools with more than 30 students per grade resulted in correlations greater than .70 in 37 of 46 states/jurisdiction for grade 4 and 42 of 43 states/jurisdiction for grade 8.