The 2011 Broad Prize for Urban Education Summary of Data Collection and Analysis Procedures

This document discusses the data collection and analysis procedures used to generate the summary tables created by MPR Associates for the 2011 Broad Prize and available at www.broadprize.org. First, it describes the criteria and data sources used to identify the eligible districts. Second, it reviews each of the achievement measures and the data on which they are based. Third, it explains the methodologies used to analyze the achievement data to measure performance and improvement.

Please note that several changes have been made to the quantitative analyses for The 2011 Broad Prize. Descriptions of additions or significant methodological changes are shown in bold italics throughout this document.

The Broad Prize for Urban Education selection process is based on the procedure developed in the inaugural year of the Prize. The winner is selected through a two-step process. First, a Review Board selects finalists from an initial pool of eligible districts, based on a review of data collected from federal, state, and other sources. Second, a Selection Jury uses these data, augmented by qualitative information gathered through site visits to the finalists and additional quantitative analyses, to choose the winner.

In an effort to ensure that the school districts in consideration for The Broad Prize are widely considered to be "urban" and "large," which is the intention behind the award, The Broad Foundation tightened the eligibility rules for The 2011 Broad Prize. To be eligible for The Broad Prize, school districts had to meet the revised criteria related to size, poverty, and urbanicity described below under "Eligible Districts." (Please note that 2010 finalists who would otherwise have been ineligible in 2011 were "grandfathered in" for the 2011 Prize only.)

As a result of the narrowed definition of eligibility, 75 school districts (instead of the previous 100) will now be in annual contention for The Broad Prize. For the 2011 Prize, these districts were located in 31 states and the District of Columbia. See Table 1.1 for a list of districts and a description of their demographic characteristics as of 2009.

Given the smaller set of eligible districts, and The Broad Foundation's interest in highlighting the most improved and highest-performing large urban school districts, the Prize has also been modified so that only four school districts (instead of five) will be selected as finalists by the Review Board.

To support the selection of the four finalists, MPR Associates, Inc. collected extensive data on the pool of 75 eligible districts plus the three previous winners, analyzed the data to develop multiple measures of performance and improvement, and prepared

detailed profiles of each district as well as summary tables containing data on all eligible districts.

1) Eligible Districts

For The 2011 Broad Prize, in order to ensure that only the largest, most urban districts were included, eligibility criteria were revised and the number of eligible districts was reduced from 100 to 75. The pool of eligible districts was identified based on data from the U.S. Department of Education's Common Core of Data. To be eligible for The Broad Prize, school districts must meet certain criteria set by The Broad Foundation related to size, poverty, and urbanicity. Winners from the previous three years are ineligible (currently Gwinnett County Public Schools in Georgia, and Aldine and Brownsville Independent School Districts in Texas). ***The revised criteria for eligibility in 2011 included the following:***

- ***K–12 districts serving at least 37,500 students that have at least 40 percent of students eligible for free or reduced-price school lunch (FRSL), at least 40 percent of students from minority groups, and an urban designation (Locale Code 11, 12, or 21 in the Common Core of Data[1]) are identified. In states where more than 10 districts qualify under these criteria, only the 10 largest qualifying districts are eligible (69 districts met these criteria in 2011).***
- ***In states with no districts meeting the above criteria, the next largest districts in the nation with at least 40 percent FRSL, at least 40 percent minority, and an urban designation are identified, in order to bring the total number of eligible districts to 75. Only one district per state can qualify under these criteria (3 districts were included in 2011 based on these criteria).***[2]
- ***Three finalists for the 2010 Broad Prize who did not meet the above criteria were "grandfathered in" for the current Prize year only.***[3]

For The 2011 Broad Prize, data on school district demographics were obtained from the National Center for Education Statistics' Common Core of Data (CCD) for 2009 (the most recent year for which data were available) and used to determine the list of 75 eligible districts.

The 75 eligible school districts are located in 31 states and the District of Columbia. See Table 1.1 for a list of districts and a description of their demographic characteristics as of 2009.[4]

[1] CCD locale code 11 represents a large city; code 12 represents a mid-size city; and code 21 represents a large suburb. Sable, J. (2008). *Documentation to the NCES Common Core of Data Local Education Agency Universe Survey: School Year 2006–07 Version 1a* (NCES 2009-301). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC

[2] These include Indianapolis, Indiana; Des Moines, Iowa; and Norfolk, Virginia.

[3] These include Montgomery County Public Schools, Maryland; and Socorro and Ysleta Independent School Districts in Texas.

2) Measures of Student Achievement

Detailed data on various measures of student achievement were obtained for each district, using federal, state, and other sources. Wherever possible, data were collected by grade level, race/ethnicity (African-American, Asian, Hispanic, and White), and low-/non-low-income status. To provide context for these data, state data on student achievement on nationally standardized tests were also obtained. Measures of student achievement examined include:

- State scores on National Assessment of Educational Progress (NAEP) and Northwest Evaluation Association (NWEA) tests,
- Reading and mathematics proficiency rates and advanced proficiency rates as determined by state tests used for federal accountability[5],
- High school graduation rates, and
- College readiness measures.

These data were analyzed in order to report indicators of both recent performance and improvement over time. Performance indicators reflect the most recent year of available data for any given measure. Improvement indicators generally reflect the four most recent years of available data and were determined by calculating the "average change" for each measure. Average change was calculated as the slope of the best fit line among the four most recent years of available data points, generally determined by regressing the relevant outcome measure on year.[6] If only one data point was available, or if data were missing for both of the two most recent years, average change was not calculated.

Each of the analyses provides information on both performance and improvement. In theory, districts with high initial performance levels might be expected to have lower rates of improvement. For example, if the residuals analysis reveals that a district performed consistently above expectations during all four years, but did not improve, that district could still be thought of as consistently high-performing.

State Scores on NAEP and NWEA Tests

On a regular basis, the NAEP, administered by the National Center for Education Statistics (NCES), publishes achievement scores for the nation for students in the 4th, 8th, and 12th grades, and achievement scores for participating states for students in the 4th and 8th grades.[7] These scores are based on tests administered to samples of

[4] States without eligible districts this year are Alaska, Arkansas, Connecticut, Delaware, Hawaii, Idaho, Maine, Mississippi, Missouri, Montana, New Hampshire, North Dakota, Oregon, Rhode Island, South Dakota, Vermont, Washington, West Virginia, and Wyoming. Hawaii is ineligible because it has a statewide school system.
[5] The "advanced" level was defined as the combination of all performance levels above "proficient" on a state's test.
[6] When only two data points were available, the slope was equal to $(X_2 - X_1)/(Year_2 - Year_1)$.
[7] Beginning in 2009, NAEP scores in reading and mathematics for grade 12 were reported for 11 states that volunteered to participate as part of a pilot state study.

students at the national and state levels, respectively. State NAEP scores are presented here as a means of "calibrating" among states. The test is constructed to allow comparisons over time as well as across the states.

State NAEP data for the 4th- and 8th-grade reading and mathematics assessments in 2005, 2007, and 2009 are used here to report information on students scoring at the proficient level or above (Table 2.1). The percentage of students scoring at the proficient level or above on the 2009 administration of NAEP is presented, as well as the simple change between 2005 and 2009 and between 2007 and 2009, including an indication of whether the change was statistically significant. These data are meant to show whether a district is located in a state that performs above or below the national average and whether the state's improvement is above or below the national average. These data provide some context where test content and standards vary across states and state proficiency rates are not directly comparable.

NAEP Trial Urban District Assessment (TUDA) data for 2005, 2007, and 2009 are also presented. The NAEP TUDA was launched in 2002 to pilot a district-level NAEP assessment that would allow for more detailed analyses of urban education. Eighteen large urban school districts, all currently eligible for The Broad Prize, participated in at least one of the NAEP TUDA administrations given through 2009. As with the State NAEP, the percentage of students scoring at the proficient level or above for the 2009 administration of NAEP TUDA is presented (Table 2.2). The change between 2005 and 2009 and between 2007 and 2009 are shown, where possible, including an indication of whether the change is statistically significant. For those districts participating in the TUDA, these data provide a direct comparison of student achievement on a national achievement test.

Tables 2.3a and 2.3b show results from an ongoing NWEA study that map state proficiency standards onto NWEA's Measures of Academic Progress (MAP)[8] scales for reading and mathematics, respectively. The scale score equivalents shown represent the NWEA MAP scale scores that were determined to be equivalent to the states' proficiency standards for elementary (grades 3–5), middle (grades 6–8), and high school (grades 9–11) in reading and math. They provide an indication of the relative rigor of the state proficiency standards. The states' score equivalents were ranked, with "1" indicating the highest score equivalent and most rigorous state proficiency standards.

Reading and Mathematics Proficiency as Determined by State Tests

Key indicators of student performance include scores on state-mandated achievement tests used for federal accountability and trends in these scores over time. Test score data in reading and mathematics were collected from each state containing an eligible

[8] MAP is a computerized assessment used for diagnostic and accountability purposes by schools and school systems in many states. Thirty-seven states (25 of which have at least one eligible district) had enough students participate in both MAP and their state assessment to allow for estimating state cut score equivalents on the MAP scales for at least some grades.

district for the period 2007 through 2010.[9] These data were used to calculate the percentage of students in each district scoring at or above the proficient level on state tests in reading and mathematics in each grade 3 through 12, where available. Weighted by the number of test-takers at each grade level, these data on student achievement were aggregated across elementary grades (3–5), middle grades (6–8), and high school grades (9–12). These state assessment data were analyzed (using methods described below) to calculate actual versus expected performance (Tables 3.1–3.4); decile rankings of proficiency rates at both the proficient or above level (Tables 4.1–4.4) ***and the advanced level (Tables 4.5–4.8)***; and gaps between low- and non-low-income students, White and African-American students, and White and Hispanic students (Tables 5.1–5.6). In Sections 3, 4, and 5, within-state decile rankings were applied to both performance and improvement measures as a further indicator of a district's relative standing in its state. Test data were suppressed if they were deemed unreliable or if the subgroup being reported at a given level represented fewer than 5 percent of the test-takers at that level.

It is important to note that standardized tests differ across states in a number of respects. These differences include test standards (some are more rigorous than others as indicated in Tables 2.3a and 2.3b), proficiency requirements, cut points, grades tested, testing requirements for English language learners and students with disabilities, and many other details. Thus, meeting proficiency requirements is not an indicator that can be readily compared across states.

Moreover, many state tests changed during the period from 2007 to 2010, including changes in grades tested, test standards, proficiency cut points, and population inclusion policies. Table 1.2 summarizes state test comparability and indicates which years and grades were included in our analyses in Sections 3 through 5. When state test standards or policies changed during the four-year period analyzed for this Prize year, priority was given to using those data that would maximize the number of years available for trend analyses. Therefore, some grade-level data may have been excluded in certain years in order to increase the total number of years available for trend analysis for a particular subject and education level.

Because states use different tests and different standards of proficiency, individual states may be subject to "floor effects" or "ceiling effects." If proficiency levels are generally very high in a state (near 90 percent, for example), then high-performing districts may not be able to demonstrate their relative advantage because their proficiency level cannot increase above 100 percent. Similarly, if state proficiency levels are very low, then the relative advantage of higher performers may again be understated because lower performing districts cannot fall below 0 percent. Table 1.3 shows state proficiency rates over the period from 2007 to 2010. This table can help identify possible floor and ceiling effects associated with the state tests. Also, examining performance and improvement at the advanced level (Tables 4.5 through 4.8), especially when state proficiency rates are generally above 80 percent, may provide

[9] The data were provided directly by state agencies or downloaded from their websites.

some further insight into a district's performance and improvement in states with possible ceiling effects.

While differences in testing standards from state to state prevent the direct comparison of proficiency rates across states, one can compare the relative performance of school districts within their own states to the relative performance of other districts within their own states. One approach for conducting such comparisons is within-state ranking. Several of the state-test analyses in this report include within-state decile rankings of performance and improvement measures. These within-state decile rankings can be used to determine how an eligible district's relative performance within its state compares with the relative performance of eligible districts in other states.

The following provides an example of how decile rankings are generally determined for any given measure. Suppose that a district in State A had a 2010 elementary reading proficiency rate on the state's reading assessment of 49% for all students. In order to understand the relative standing of that performance level in the state, the proficiency rates of all districts in the state were ranked, highest to lowest, and divided into deciles (ten groupings). In a state with 300 school districts, there would be 30 districts in each decile; in a state with 30 school districts, there would be 3 districts in each decile. In this example, the district proficiency rate of 49% fell in the 7th decile in the state (that is, in the top 70 percent—or bottom 40 percent—of districts in the state). See the table below, which illustrates where the proficiency rate of 49% fell in the distribution of proficiency rates for State A (colored orange).

STATE A PROFICIENCY RATES (2010)

Decile Ranking	1	2	3	4	5	6	7	8	9	10
State A: Elementary Reading Proficiency Rates (2010)	95–88	88–82	81–75	74–70	69–67	66–54	**54–48**	47–39	38–31	30–18

Because testing standards differ from state to state, a proficiency rate of 49% may have a very different standing in another state. Suppose that a district in State B also had a 2010 elementary reading proficiency rate of 49% for all students. As the table below illustrates, a proficiency rate of 49% fell in the 3rd decile in State B (colored orange).

STATE B PROFICIENCY RATES (2010)

Decile Ranking	1	2	3	4	5	6	7	8	9	10
State B: Elementary Reading Proficiency Rates (2010)	65–59	58–52	**51–48**	47–44	43–41	40–37	36–32	31–26	26–18	30–18

Based on the examples above, even though both districts had the same absolute proficiency rate, the eligible district in State B was performing better relative to other districts in its state than the eligible district in State A was performing within its state context.

Decile rankings were applied to several different types of measures, as explained in the Data Analysis Methods section, below.

High School Graduation Rates

Another key measure of student performance is the high school graduation rate. There are several ways to estimate graduation rates, and three are presented here (Tables 6.3–6.5). The average of these three graduation rate measures is also presented for each district (Table 6.1).

In order to highlight districts' relative standing among all eligible districts, both the districts' average of their three graduation rates for 2008 and the average of their three improvement rates for 2005 to 2008 were ranked among the 75 eligible districts and the rankings are presented in Table 6.2.

The data needed to calculate graduation rates, which were obtained from the CCD, included total and subgroup enrollment figures and diploma counts for each district for the high school classes of 2005 through 2008 (the most recent year for which diploma counts were available at the time of the analysis). Data were suppressed if they were deemed unreliable or if the subgroup being reported represented less than 5 percent of the district's total enrollment.

College Readiness Measures

District measures of the college readiness of students include SAT and ACT scores and participation rates. These two tests are designed to assess readiness for college-level work. Scale scores for each SAT subject (reading, writing, and math) range from 200 to 800. Scale scores for the composite ACT test (covering English, mathematics, reading, and science) range from 1 to 36. With district permission, the College Board and ACT provided mean test scores for each district for 2007 through 2010 (Tables 7.1–7.3 and 7.5), along with the number of seniors who had taken the test (regardless of when they took the test during high school).[10] Participation rates were calculated for each test using CCD enrollment data for 12th-graders in the relevant year as the denominator (Tables 7.4 and 7.6).

Data provided by College Board and ACT are not student-level and therefore do not allow for the calculation of a combined SAT and ACT participation rate. In order to provide some indication of whether the SAT or ACT is the predominantly used test in a district, results reported in Tables 7.1 through 7.6 are de-emphasized (grayed out) when the 2010 participation rate for "all students" on the relevant test was less than 20 percent.

Another measure of college readiness is the extent to which students take and pass Advanced Placement (AP) examinations. These examinations provide a standardized

[10] Where students had taken the test more than once, the most recent score was reported.

measure of student performance in college-level courses taken while in high school. AP grades are reported on a 5-point scale:

5 = Extremely well qualified
4 = Well qualified
3 = Qualified (equivalent to passing)
2 = Possibly qualified
1 = No recommendation

Again with permission from each district, the College Board provided data for 2007 through 2010 on the number of AP examinations taken by juniors and seniors in the district and the number of passing scores (3 or above) . Exam passing rates using these data are presented (Table 7.7). The College Board also provided the number of juniors and seniors who took any AP test; overall student participation rates using CCD enrollment data for 11th- and 12th-graders in the relevant year as the denominator are also presented (Table 7.8).

Data were suppressed if they were deemed unreliable. Test scores were suppressed if they were based on the performance of fewer than 15 students, as required by the College Board. Participation rates were suppressed if a subgroup represented less than 5 percent of enrollment in the relevant grades. In addition, subgroup results were suppressed if data on the number of test takers whose race/ethnicity was identified represented less than 90 percent of the total number of test takers for a given test and year.[11]

3) Data Analysis Methods

The summary tables include data collected on district characteristics (Table 1.1), background on state tests (Tables 1.2 and 1.3), and state test rigor (Tables 2.1–2.3b). For the remaining tables, data on student achievement described above were analyzed to develop measures of the following:

- Overall actual performance and improvement on state tests versus expected performance when differences in the distribution of results on the various state tests and the proportion of low-income students in the districts are taken into consideration.
- Overall and subgroup performance and improvement on state tests, at both the proficient or above ***and the advanced levels***, compared with other districts in the state.
- Achievement gaps between Whites and other racial/ethnic groups—African-Americans and Hispanics—and between low-income and non-low-income students, and the progress that is being made in closing these gaps.

[11] Race/ethnicity is self-reported on SAT, ACT, and AP exams, and the amount of missing race/ethnicity data varies by district and year.

- Overall and subgroup high school graduation rates.
- Overall and subgroup performance on and participation in college readiness exams.

Performance and Improvement on State Tests

Restricting Analyses to Local School Districts

Test files obtained from states often contain results for all local education agencies in the state. State test analysis files created by MPR for The Broad Prize were restricted to local school districts, defined as Common Core of Data (CCD) agency types 1 and 2 (local school districts***). Beginning in 2011, charter school agencies, defined as CCD agency type 7, were also included in analysis files.***[12] As a result, supervisory union administrative centers, regional education services agencies, state- and federally operated institutions serving special-needs populations, and other agencies not fitting the mentioned categories were removed from The Broad Prize data files and analyses.

Calculating Improvement Measures

For each of the state test analyses described below, improvement or "average change" was calculated as the slope of the best fit line among the available data points for 2007 through 2010. The slope was generally determined by regressing the available indicator of interest (i.e., standardized residuals, percentages of students at the proficient or above level, ***percentages of students at the advanced level,*** and achievement gaps) on year. If only one data point was available, or if the data points were missing for both 2009 and 2010, average change was not calculated.

Standardized residuals were used in improvement or "average change" calculations regardless of any changes in state tests from 2007 to 2010, as long as a test change was implemented statewide. This practice was followed because of the relative nature of the measure. Standardized residuals indicate a district's performance relative to that of other districts in the state, regardless of the particular test administered in a given year. In the rare case where a test change is implemented differently across districts in the state, residuals in the affected years would be suppressed and not included in improvement calculations.

Trends for the more direct comparisons of student achievement, however, can be affected by year-to-year changes in state tests. Therefore proficiency rates, ***advanced proficiency rates***, or achievement gap measures in a given year may have been excluded from trend analyses if there were changes in state testing standards or policies.

[12] Prior to 2008, the CCD combined charter school agencies with other types of agencies that may have served special populations. CCD has since created a separate category for "other education agencies," allowing the inclusion of charter agencies in district-to-district comparisons. Wherever applicable, charter agencies were included in these analyses.

Regression Analysis and Calculation of Standardized Residuals

An ordinary least squares regression analysis was conducted to determine the extent to which each district performed better or worse than other districts in its state given the district's percentage of low-income students. Specifically, the dependent variable in the regression analysis was the percentage of test-takers in each of the three grade-level groupings (elementary, middle, and high school) in the district who were proficient or above on the state test. The independent variable was the percentage of test-takers in each grade-level group in the district who were low income.

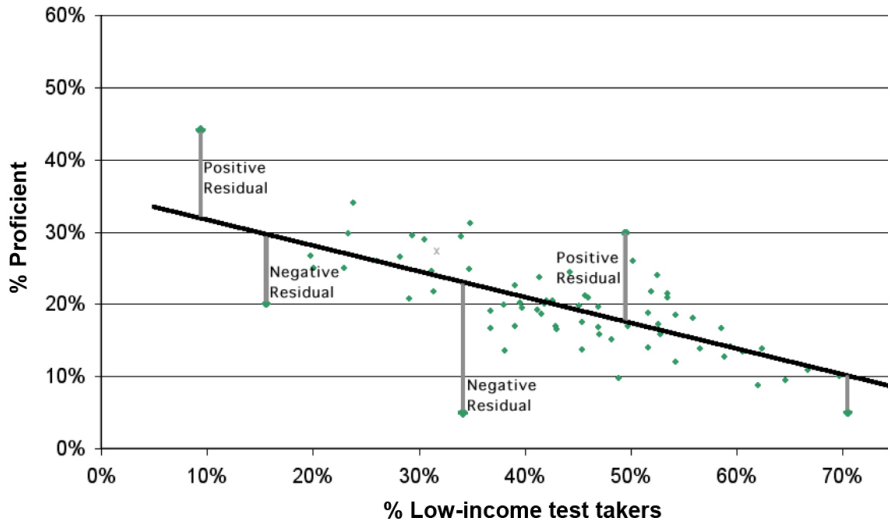
Regressions were weighted by district size, as measured by enrollment, giving greater weight in the regressions to larger districts, and avoiding possible undue influence of very small districts on the regression results. It should be noted, however, that this practice is likely to moderate the size of the residuals for very large districts, which have a greater influence on the slope of the regression line than smaller districts. Such moderating effects will be more likely to occur in states where a single district is substantially larger than all other districts in the state.

Running the regressions requires data for all districts in the state, as well as data on the percentage of test-takers in each district who were low income. In cases where data were available for the state and eligible district, but not for other districts in the state, regressions could not be run.[13]

Each district's expected or predicted proficiency level was calculated based on the regression. The difference between the district's actual percentage of students who tested at the proficient or above level and the predicted or expected value is the residual. A positive residual indicates that the district is performing better than expected on the state test given the percentage of low-income students taking the test, while a negative residual indicates lower-than-expected performance. Figure 1 (below) illustrates this approach.

[13]Regressions could not be run for the District of Columbia (DC), because data for charter agencies not affiliated with DC Public Schools were not available at the time of analysis.

Figure 1



Separate regressions for each year of data were calculated.

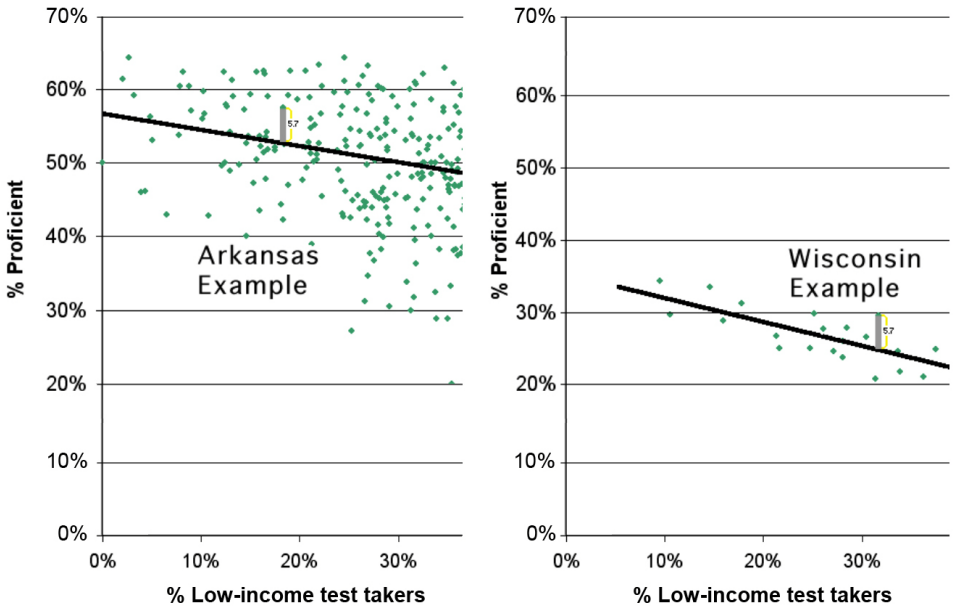
Some states changed tests over the period under review, and tests differ from state to state. Consequently, the interpretation of residuals varies. In order to have a measure with greater comparability, The Broad Prize methodology uses "standardized residuals." A district's standardized residual is typically calculated by dividing its residual by the standard deviation of all residuals from the state regression.[14]

It should be emphasized that residuals are *relative* performance measures. A district's performance was assessed relative to that of other districts in the state, not in absolute terms.

This point is illustrated in Figure 2. As an example, a district in Arkansas may have a residual in elementary reading of 5.7 (meaning that they had 5.7 percent more students reach the benchmark than their "expected level" given their district's poverty level). At the same time, a district in Wisconsin may also have a residual of 5.7 in elementary reading. The assessment of how well each district is performing, however, may not be the same even though both have the same residual. If the majority of districts in Arkansas are within 6 percentage points of the expected level, while the majority of districts in Wisconsin are within 2 percentage points of the expected level, the Wisconsin district is performing much better compared with its peers than the Arkansas district is compared with its peers. Standardizing the residuals helps account for such differences in variability.

[14] When weighting the regressions by district size for The 2011 Broad Prize, standardized residuals were calculated by formula.

Figure 2



Caution must be exercised in comparing these standardized residuals across states. For example, a district performing above average in a state that ranks below the national average on NAEP may be performing no better (in absolute terms) than a district performing below average in a state that ranks above the NAEP national average. To see which states perform below or above the national average, it may be helpful to consult the state-by-state NAEP data in Table 2.1.

Residuals were suppressed if the overall regression (as represented by the F statistic) was not significant, if the regression produced implausible (out-of-range) predicted values for the eligible district, or if the underlying data were determined to be unreliable.

Separate standardized residuals were calculated for each subject (reading and mathematics), level (elementary, middle, and high school), and year (2007, 2008, 2009, and 2010), for a total of 24 possible regressions for each state. In order to compare the magnitude of standardized residuals *within* a state, the 2010 residuals are presented (Table 3.1). (See cautionary note about moderating effects for very large districts, above.)

Improvement or "average change" was calculated as the slope of the best fit line among the available data points for 2007 through 2010 (Table 3.3). The slope was generally determined by regressing the available standardized residuals on year. If only one data point was available, or if residuals were missing for both 2009 and 2010, average change was not calculated.

As mentioned above, standardized residuals were used in improvement or "average change" calculations regardless of any changes in state tests from 2007 to 2010, as long as the test change was implemented statewide. Just as the performance measure is based on relative performance, the improvement measure is based on improvement in relative performance. Thus, a district whose scores improved, but more slowly than those of other districts in the state, could find itself moving upward on the graph from year to year more slowly than the upward movement of the regression line. Such a district would show negative (relative) improvement in the analysis.

For purposes of comparing the magnitude of standardized residuals *across* states, within-state decile ranks based on standardized residuals for all districts in a state regression were computed and the decile ranks for 2010 are presented (Table 3.2). In addition, within-state decile ranks of the average change in residuals between 2007 and 2010 were also computed (Table 3.4). Decile ranks of 1–3, which represent the upper 30 percent of residuals in a state, are highlighted for both average 2010 decile ranks and for average change.[15] Decile ranks were calculated separately by level (elementary, middle, and high school) and subject.

Unadjusted Proficiency Rate Decile Rankings

While the preceding regression analyses take poverty into account when comparing district performance within a state, the analyses discussed in this section directly compare student performance among districts in a state for each reported subgroup *without* taking into account district poverty levels. District results are then compared across states.

Section 4 tables present direct comparisons of the performance of four subgroups (all students, African-American, Hispanic, and low-income students) in a district with their state peers. These comparisons focus on the proficient or above level (Tables 4.1–4.4) ***and the advanced level (Tables 4.5–4.8)*** on state tests.[16] Each district subgroup's 2010 proficiency rates were ranked within their state using deciles, and the average decile rank across the three education levels (elementary, middle and high school) is shown for both reading and math. Counts of decile ranks of 1–3, which represent performance in the upper 30 percent of districts in the state, for each subgroup listed above are also shown in the tables.

Similarly, each district subgroup's improvement or "average change" rate for both the proficient or above level ***and the advanced level*** was ranked within their state using deciles, and the average decile rank across the three education levels (elementary, middle and high school) is shown for both reading and math. Counts of decile ranks of 1–3, which represent improvement in the upper 30 percent of districts in the state, for each subgroup listed above are also shown in the tables. "Average change" was calculated as the slope of the best fit line among the available data points for 2007

[15] Decile ranks range from 1 for the largest positive residuals to 10 for the largest negative residuals in a state.

through 2010. The slope was generally determined by regressing the available proficiency rates on year. If only one data point was available, or if data were missing for both 2009 and 2010, average change was not calculated. Improvement at the advanced level may be a particularly useful measure in states with high average proficiency rates and possible ceiling effects.

Test data were suppressed if they were deemed unreliable or if the subgroup being reported at a given level represented fewer than 5 percent of the test-takers at that level.

Achievement Gaps

Three achievement gaps were calculated:

- African-American vs. White gap: This compares the performance of African-American students with White students.
- Hispanic vs. White gap: This compares the performance of Hispanic students with White students.
- Low-income vs. non-low-income gap: This compares the performance of low-income students with non-low-income students.

Achievement gaps are represented by negative numbers in the district data reports, and the closing of gaps is represented by positive numbers. For example, if a district's African-American students perform 30 percentage points below the district's White students, this gap is represented by -30. If the gap decreases to -10 in subsequent years, then the magnitude of the gap closure is the later year's gap minus the earlier year's gap (-10 minus -30 equals +20).

Gaps were measured in three different ways:

Internal District Gap

This measure was used to compare the performance of different subgroups within the district. Some caution must be exercised in comparing internal gaps across districts because these comparisons may be distorted by the following factors:

- The relative absence of one of the subgroups in some districts (e.g., few White or non-low-income students). To address this issue, internal gaps were not calculated in districts where either of the groups being compared represented fewer than 5 percent of the district's test-takers.
- Differences between districts in the composition of analogous groups (e.g., high-income Whites in one district and moderate-income Whites in another).
- Higher than average performance or improvement by the advantaged group in some districts and lower than average performance or improvement by the

advantaged group in others (which could cause districts with lower performing advantaged students to appear to be doing a better job of "closing the gap").

- Ceiling or floor effects (discussed above), which can distort the comparison of gaps across states.

An internal district gap was considered to be closing if the district's disadvantaged group proficiency was increasing and the district's advantaged group proficiency was either steady or increasing. The gap was closing because the district's disadvantaged group proficiency was increasing at a faster rate than the district's advantaged group proficiency.

Internal District Versus Internal State Gap

This measure was used to calculate the *district's* internal gap minus the *state's* internal gap. The "state" internal gaps against which district internal gaps were compared generally excluded the district's results. That is, unless otherwise indicated, district proficiency rates were removed from state averages to produce "rest of state" values for comparison purposes. This approach was particularly important in cases where a very large eligible district enrolled a significant proportion of the population in a state and would otherwise have been compared largely with itself. In states with multiple eligible districts, the "state" values will vary, because each district was compared separately with all other districts in the state except itself.

Positive numbers indicate that the district outperformed the state on the measure. For example, if the district's Hispanic students are performing 10 percentage points below the district's White students (a gap of -10), but the state's Hispanic students are performing 15 percentage points below the state's White students (a gap of -15), then the district's gap is 5 percentage points smaller than the state's gap. This is shown in tables in the district data reports as an internal district vs. internal state gap of +5 percentage points (-10 minus -15 equals +5).

By similar reasoning, a positive change in this measure over time for Hispanic students would indicate that the district's Hispanics are improving faster relative to the district's Whites than the state's Hispanics are improving relative to the state's Whites.

An internal district vs. internal state gap was considered to be closing if the district's internal district gap was closing (as defined above), and was closing at a faster rate than the state internal gap was closing.

External Gap: District Disadvantaged Versus State Advantaged

This measure was used to compare the performance of the *district's disadvantaged* group with that of the *state's advantaged* group. Again, the "state's" advantaged proficiency rate against which the district's disadvantaged group was compared generally excluded the district's results. That is, unless otherwise indicated, district proficiency rates were removed from state averages to produce "rest of state" values for

comparison purposes. This approach was particularly important in cases where a very large eligible district enrolled a significant proportion of the population in a state and would otherwise have been compared largely with itself. In states with multiple eligible districts, the "state" proficiency rates and gaps will vary, because each district was compared separately with all other districts in the state except itself.

If 30 percent of District A's Hispanic students, 40 percent of District B's Hispanic students, and 50 percent of the state's White students are proficient on the state test, District A's external gap for Hispanics is 30 percent minus 50 percent (or -20 percentage points), and District B's external gap is 40 percent minus 50 percent (or -10 percentage points). A negative number indicates the district's disadvantaged group proficiency was lower than the states' advantaged group proficiency. External gap statistics are generally negative numbers, but improvement in external gaps (improvement in the performance of the district's disadvantaged students relative to the state's advantaged students) will show up as positive numbers. External gap statistics are particularly helpful in evaluating districts whose populations of advantaged students are too small to allow for the calculation of an internal district gap.

An external gap was considered to be closing if the district's disadvantaged group proficiency was increasing at a faster rate than the state's advantaged group proficiency was increasing.

Table 5.1 shows a summary of gap closures for each of the three subgroups and overall. Tables 5.2–5.4 show detail on each gap type (internal district, internal district vs. internal state, external) for the three subgroups of interest (low income, African American, and Hispanic, respectively). Table 5.5 shows the number of gaps for each district that were among the smallest 30 percent of gaps in the state. To identify districts with the smallest gaps, decile ranks based on gaps for all districts in a state were computed and gaps that had a decile rank of 1, 2, or 3 were counted.[17] Decile ranks could only be calculated for internal district gaps. Table 5.6 shows the number of gaps for each district that were among the top 30 percent of fastest closing gaps in the state. To identify districts with the fastest closing gaps, decile ranks based on the rate of gap closures were computed and gaps that were closing (as defined above) and that had a decile rank of 1, 2, or 3 were counted.

Important Note Regarding Achievement Gap Data

The same cautions must be exercised in comparing gaps across states as were discussed earlier for regression residuals and proficiency deciles. Also, additional caution must be used because the three gap types are not "standardized" and are even more vulnerable to ceiling and floor effects than are standardized measures. Table 1.3 shows state proficiency rates over the period from 2007 to 2010, and can help identify possible floor and ceiling effects associated with the state tests.

[17] Decile ranks range from 1 for the smallest gaps to 10 for the largest gaps in a state.

High School Graduation Rates

There are a number of ways of estimating high school graduation rates. The binder includes calculations using three different methods:

1. The Averaged Freshman Graduation Rate (AFGR)
2. Urban Institute Graduation Rate (Cumulative Promotion Index or CPI)
3. Manhattan Institute Graduation Rate (Greene's Graduation Indicator or GGI)

Graduation rates were calculated using district enrollment data (for different grades and years) and district diploma counts, according to the equations below. Enrollments and diploma counts were obtained from the National Center for Education Statistics' Common Core of Data (CCD). For each of the three measures, graduation rates were suppressed if needed enrollments (obtained from school-level files) summed to less than 95 percent or greater than 100 percent of total district enrollment. Graduation rates were also suppressed as implausible if they exceeded 100 percent. Finally, graduation rates were suppressed for racial/ethnic groups that made up less than 5 percent of the district's enrollment. Additional suppression rules specific to the CPI and GGI rates are described below.

Each methodology is briefly explained here for your convenience.

Averaged Freshman Graduation Rate (AFGR)

This method divides the number of students graduating in Year *y* by an average of the 8th-grade enrollment in Year *y* – 4, 9th-grade enrollment in Year *y* – 3, and 10th-grade enrollment in Year *y* – 2:

$$\text{Graduation Rate} = \frac{G_y}{\left(S_{8,y-4} + S_{9,y-3} + S_{10,y-2}\right)/3}$$

Where: G = Number of graduates receiving a regular diploma
y = School year
Denominator = Smoothed estimator for first-time 9th-grade enrollment

Urban Institute Graduation Rate (Cumulative Promotion Index or CPI)

Also known as Swanson's Cumulative Promotion Index (SCPI), this method assumes that graduation is a process composed of three grade-to-grade promotion transitions (9 to 10, 10 to 11, and 11 to 12) in addition to the graduation event (grade 12 to diploma). Each of the transitions is calculated as a probability, dividing the enrollment of the following year by the enrollment of the current year for the grade in question. These separate probabilities are then multiplied to produce the probability that a student in that school system will graduate within four years of entering 9th grade.

$$\text{Graduation Rate} = \frac{S_{10,y+1}}{S_{9,y}} * \frac{S_{11,y+1}}{S_{10,y}} * \frac{S_{12,y+1}}{S_{11,y}} * \frac{G_y}{S_{12,y}}$$

Where: S_{grade} = Number of students in a specified grade
y = School year
G = Number of graduates receiving a regular diploma

As recommended by Swanson[18] and the National Center for Education Statistics (NCES),[19] the following quality controls were followed. When data were missing for any grade 9–12 for either of the two consecutive years required to calculate the rate, the graduation rate was reported as missing. In addition, promotion rates greater than 100 percent but not exceeding 115 percent were "trimmed" to 100 percent before calculating the graduation rate. Promotion rates exceeding 115 percent were suppressed, as was the resulting graduation rate.

Manhattan Institute Graduation Rate (Greene's Graduation Indicator or GGI)

The number of students who receive a diploma is divided by the product of a measure of high school population change over time and an estimate of the number of first-time 9th-graders. The population change quantity adjusts for enrollment variability due to student mobility among districts and states rather than dropping out.

Graduation Rate =

$$\frac{G_y}{\left(1+\frac{\left(S_{9,y}+S_{10,y}+S_{11,y}+S_{12,y}\right)-\left(S_{9,y-3}+S_{10,y-3}+S_{11,y-3}+S_{12,y-3}\right)}{S_{9,y-3}+S_{10,y-3}+S_{11,y-3}+S_{12,y-3}}\right)*\left(\frac{S_{8,y-4}+S_{9,y-3}+S_{10,y-2}}{3}\right)}$$

Where: G = Number of graduates receiving a regular diploma
y = School year
S_{grade} = Number of students in a specified grade

As recommended by Greene and Winter[20] and the National Center for Education Statistics (NCES),[21] the following quality controls were followed. Rates for districts with fewer than 200 9th graders in a given year were suppressed, as were rates for districts with population changes greater than 30 percent, and districts with fewer than 2,000 9th graders *and* a population change of more than 20 percent.

The cited NCES study reported that when calculating a statewide graduation rate, the Averaged Freshman Graduation Rate came closest to approximating a longitudinal graduation rate. The different methodologies sometimes lead to very different results,

[18] Swanson, C. (2003). *Who Graduates? Who Doesn't? A Statistical Portrait of Public High School Graduation, Class of 2001*. New York: Urban Institute; and updates.
[19] Seastrom, M., Chapman, C., Stillwell, R., McGrath, D., Peltola, P., Dinkes, R., and Xu, Z. (2006). *User's Guide to Computing High School Graduation Rates, Volume 1: Review of Current and Proposed Graduation Indicators* (NCES 2006-604). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office, and by the same authors, *Volume 2: Technical Evaluation of Proxy Graduation Indicators* (NCES 2006-605).
[20] Greene, J. and Winters, M. (2005). *Public High School Graduation and College-Readiness Rates: 1991–2002*. Education Working Paper No. 8. Manhattan Institute. February.
[21] Ibid.

because each uses different types of data from different years. All three have strengths and weaknesses but are considered acceptable methodologies. It should be remembered that all three measures are estimates of the true longitudinal graduation rate. The smaller the district, state, or student group being analyzed, the less precisely the three graduation rates estimate the true longitudinal graduation rate.

Tables 6.3–6.5 show district results on the three graduation rate measures for the classes of 2005 through 2008,[22] and Table 6.1 shows the average of the three graduation rates. Improvement or "average change" was calculated as the slope of the best fit line among available data points from 2005 to 2008, generally determined by regressing the relevant outcome measure on year.[23] If only one data point was available, or if data were missing for both 2007 and 2008, average change was not calculated. In addition, both the districts' average of the three rates for 2008 and the average of the three improvement rates for 2005 to 2008 were ranked among the 75 eligible districts and the rankings are presented (Table 6.2).

College Readiness

Three district-level indicators of the college readiness of their students were taken into account:

- SAT
- ACT
- AP tests

Two main types of measures were calculated for each of the three college readiness exams: 1) test score performance, and 2) participation rates. ***SAT mean scores are presented for each of the three subjects tested: reading, writing and mathematics.*** In contrast, mean composite scores (including English, mathematics, reading, and science) are reported for ACT. For the AP exam, the percentage of all AP tests taken that received scores of 3 or higher was calculated in order to provide an indicator equivalent to a passing rate. These performance measures are shown in Tables 7.1–7.3, 7.5, and 7.7.

With regard to participation rates, data provided by the College Board and ACT on the number of students taking the different tests were used, as well as data on district enrollments obtained from the Common Core of Data (CCD). SAT and ACT participation rates were based on the number of 12th-graders who took the tests, divided by 12th-grade enrollment for the district in that year (Tables 7.4 and 7.6). The overall AP participation rate was based on the number of 11th- and 12th-graders who took the test, divided by the sum of 11th- and 12th-grade enrollments for the district in that year (Table 7.8). It should be noted that 2010 enrollment data were not yet available at the time of the analysis, so 2009 enrollments were used as an estimate of 2010 enrollments in the calculation of participation rates for 2010.

[22] At the time of the analysis, diploma counts data were not available for 2009 and 2010.
[23] When only two data points were available, the slope was equal to $(X_2 - X_1)/(Year_2 - Year_1)$.

The SAT and ACT participation rates cannot be added together to obtain an overall participation rate in college entrance examinations. Students sometimes take both tests, so adding participation rates together may result in double counting. In addition, low participation rates on one or the other test may be due to regional test preferences. In order to provide some indication of whether the SAT or ACT is the predominantly used test in a district, results reported in Tables 7.1 through 7.6 are de-emphasized (grayed out) when the 2010 participation rate for "all students" on the relevant test was less than 20 percent.

In the Section 7 tables, data are presented for each district for the following student groups: All students, African-Americans, Hispanics, and White students. Improvement or "average change" was calculated as the slope of the best fit line among available data points from 2007 to 2010, generally determined by regressing the relevant outcome measure on year.[24] If only one data point was available, or if data were missing for both 2009 and 2010, average change was not calculated.

Data were suppressed if they were deemed unreliable. Test scores were suppressed if they were based on the performance of fewer than 15 students, as required by the College Board. Participation rates were suppressed if a subgroup represented less than 5 percent of enrollment in the relevant grades. In addition, subgroup results were suppressed if data on the number of test takers whose race/ethnicity was identified represented less than 90 percent of the total number of test takers for a given test and year.[25]

4) Summary Tables

Summary tables are categorized into seven sections: 1) Eligible Districts and State Tests; 2) State Test Rigor; 3) Standardized Residuals; 4) Proficiency and Advanced Proficiency Decile Ranks; 5) Achievement Gaps; 6) High School Graduation Rates; and 7) College Readiness.

Trend data as well as performance and improvement measures are presented where appropriate. A national average or eligible district average is included at the top of the tables where an average was available or could be calculated.

Section 1: Eligible Districts and State Tests

For The 2011 Broad Prize, data on school district demographics were obtained from the National Center for Education Statistics' Common Core of Data (CCD) for 2009 (the most recent year for which data were available) and used to determine a pool of eligible districts meeting size, poverty, and urbanicity criteria.[26]

[24] When only two data points were available, the slope was equal to $(X_2 - X_1)/(Year_2 - Year_1)$.
[25] Race/ethnicity is self-reported in SAT, ACT, and AP, and the amount of missing race/ethnicity data varies by district and year.
[26] Winners from the previous three years are ineligible (currently Gwinnett County Public Schools in Georgia, and Aldine and Brownsville Independent School Districts in Texas).

Table 1.1: Description of districts eligible for The 2011 Broad Prize

This table lists the eligible districts and describes their demographic characteristics as of 2009. Some data that were missing from, or unreliable in, the CCD were obtained from state or local education agencies.

Table 1.2A & 1.2B: Test Changes and Grades Included in Analysis, by State

This table summarizes state test comparability over the period from 2007 to 2010 and indicates which years and grades were included in analyses in Sections 3, 4, and 5. Noncomparable test years were not included in trend analyses in Sections 4 and 5. Information for reading is summarized in Table 1.2A, and information for mathematics is summarized in Table 1.2B.

Table 1.3: State Proficiency Rates

This table shows state-level proficiency rates over the period from 2007 to 2010, as well as average change in these rates, and can help identify possible floor and ceiling effects associated with the state tests. Only comparable test years were included in the average change calculations in the table. State proficiency rates Information for reading are summarized in Table 1.3A, and rates for mathematics are summarized in Table 1.3B.

Section 2: State Test Rigor

The National Assessment of Educational Progress (NAEP), also known as "the Nation's Report Card," is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and the arts. About every 2 years, NAEP publishes achievement scores for the nation for students in the 4th, 8th, and 12th grades, and achievement scores for participating states for students in the 4th and 8th grades. The national NAEP is structured to be nationally representative.

The State NAEP is structured to be representative of participating states and is voluntary. As of 2009, 43 states and DC opted to participate. State NAEP scores for the 4th- and 8th-grade reading and mathematics assessments are presented here as a means of "calibrating" between states. For example, if a state performs poorly relative to other states on NAEP, but the districts in that state perform very well on the state test, this might suggest that the state test is not very rigorous. NAEP Trial Urban District Assessment (TUDA) data for 4th- and 8th-grade reading for 11 large urban school districts and for mathematics for 18 large urban districts are also presented. All 18 were eligible for the 2011 Broad Prize. The State and TUDA tests are constructed to allow comparisons over time as well as across states and districts.

NAEP defines three achievement levels, which are established by the National Assessment Governing Board (NAGB):

- Basic: Denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
- Proficient: Represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
- Advanced: Signifies superior performance.

NAEP reports four performance categories: Below Basic, Basic, Proficient, and Advanced. In this section, the percentage of students scoring proficient or above is presented.

As mentioned above under Measures of Student Achievement, results from an ongoing Northwest Evaluation Association (NWEA) study mapping state proficiency standards onto NWEA's Measures of Academic Progress (MAP)[27] scales for reading and mathematics are also presented. Table 2.3 summarizes findings from their initial study. Since 2006, NWEA has conducted follow up evaluations of state tests in several states not included in the initial 2006 study or where test changes occurred since their initial analysis, and these updates are included in the summary and rankings.

Table 2.1: State NAEP—Percent of students scoring at or above proficient

This table shows the percentage of students scoring proficient or above on the 2009 NAEP assessments for grades 4 and 8 in reading and mathematics, as well as the change in the proficiency rate from 2005 to 2009 and from 2007 to 2009. State scores that are statistically above the national average in 2009 are shaded orange, as are statistically significant change values.

Table 2.2: NAEP Trial Urban Districts Assessment (TUDA)—Percent of students scoring at or above proficient

Similar to Table 2.1, this table shows the percentage of students scoring proficient or above on the 2009 NAEP assessments for grades 4 and 8 in reading and mathematics and the change in the proficiency rate from 2005 to 2009 and from 2007 to 2009 for the 18 districts that had participated in the TUDA as of that year. State scores that are statistically above the national average in 2009 are shaded orange, as are statistically significant change values.

Tables 2.3a and Table 2.3b: Average NWEA scale scores equivalent to state proficiency standards

These tables presents results from an ongoing Northwest Evaluation Association (NWEA) study mapping state proficiency standards onto NWEA's Measures of Academic

[27] MAP is a computerized assessment used for diagnostic and accountability purposes by schools and school systems in many states.

Progress (MAP) scales for reading and mathematics, respectively.[28] The scale score equivalents represent the NWEA MAP scale scores that were determined to be equivalent to the states' proficiency standards for elementary (grades 3–5), middle (6–8), and high school (10) in reading and math. They provide an indication of the relative stringency of the state proficiency standards. Each state's score equivalents are ranked, with "1" indicating the highest score equivalent and most stringent state proficiency standards. Thirty-seven states had enough students participate in both MAP and their state's reading assessments and 36 states had enough participate in both MAP and their state's mathematics assessment to allow for estimating state cut score equivalents on the MAP scales in those subjects for at least some grades. In several states, the most recent test included in The Broad Prize analysis was not yet evaluated by NWEA, as indicated in the table.

Section 3: Standardized Residuals

Tables in Section 3 present standardized residuals (as described above under Data Analysis Methods) for all eligible districts, where available. Residuals for each district are presented by school level (elementary, middle, high school) for reading and math for all students in the district.

2010 residuals are presented in Table 3.1 for purposes of comparing the magnitude of standardized residuals within a state. In addition, decile ranks based on standardized residuals for all districts in a state regression were computed and the average decile ranks for 2008 and 2009 are presented (Table 3.2) to allow for comparisons across states. Decile ranks of 1–3, which represent the top 30 percent of residuals in the state, are highlighted.[29] These average decile ranks were calculated separately by level (elementary, middle, and high school) and subject, and then averaged to produce a single performance measure for each eligible district.

Average change in residuals from 2007 through 2010 is presented in Table 3.3 for purposes of comparing the magnitude of districts' improvement in residuals within a state. In addition, decile ranks of the average change in residuals in a state were computed and the decile ranks of these average change values are presented (Table 3.4) to allow for comparisons across states. Decile ranks of 1–3, which represent the top 30 percent of improvement in residuals values in the state, are highlighted.[30] Improvement or "average change" was calculated as the slope of the best fit line among the available data points for 2007 through 2010. The slope was generally determined by regressing the available standardized residuals on year. If only one data point was available, or if residuals were missing for both 2009 and 2010, average change was not calculated.

[28] While the initial study was conducted in 2006, with some data collected in fall 2005, NWEA has conducted ongoing updates of both tests in states not included in the original mapping or in states included in the original study where testing standards may have changed. Dates of updates are indicated in Table 2.3.
[29] Decile ranks range from 1 for the largest positive residuals to 10 for the largest negative residuals in a state.
[30] Decile ranks range from 1 for the largest positive residuals to 10 for the largest negative residuals in a state.

Table 3.1: Actual vs. Expected Performance—PERFORMANCE residuals for ALL STUDENTS: 2010

This table is presented for purposes of comparing the magnitude of standardized residuals within a state. The table shows the standardized residuals for all students in reading and math, separately by school level (elementary, middle, and high school), followed by the number of positive and available residuals for the district, and the percentage of available residuals that were positive. A district with 5 "positive residuals" out of 6 "available residuals" is performing above expectations in 5 out of the 6 (or 83 percent of the) possible comparisons. Residuals for "All Students" indicate the number of standard deviations that a district performed above or below its expected performance, based on the district's percentage of test-takers eligible for free or reduced-price lunch (FRSL) at each school level and in each subject. Positive residuals are shaded in orange and indicate that the district performs above expectations compared with districts in the same state with similar percentages of low-income students for a given school level and subject. Percentages of positive residuals that were 67 percent or higher are also colored orange.

Table 3.2: Actual vs. Expected Performance—DECILE RANK of residuals for ALL STUDENTS: 2010

Decile ranks for 2010 are presented in this table, based on standardized residuals for all districts in the state regressions for the relevant years. The table shows six decile ranks for reading and math at the elementary, middle, and high school levels, followed by the average of these six decile ranks. The next two columns show the count of the first six decile ranks that had values of 1 to 3 and the count of available decile ranks.[31] The final column shows the percentage of available ranks that had values of 1 to 3. Decile ranks with values between 1 and 3, which represent the top 30 percent of performance residuals in the state, and percentages of available ranks with values of 1 to 3 that are 67 percent or greater are colored orange.

Table 3.3: Actual vs. Expected Performance—AVERAGE CHANGE in residuals for ALL STUDENTS: 2007–2010

This table shows the average change in residuals for All Students in reading and math at the elementary, middle, and high school levels, followed by the number of positive and available average change values and the percentage of available values that were positive. Positive residuals are shaded in orange and indicate that the district improved above expectations compared with districts in the same state with similar percentages of low-income students for a given school level and subject. Percentages of positive residuals that were 67 percent or higher are also colored orange.

[31] Decile ranks range from 1 for the largest positive residuals to 10 for the largest negative residuals in a state.

Table 3.4: Actual vs. Expected Performance—DECILE RANK of the AVERAGE CHANGE in residuals for ALL STUDENTS: 2007–2010

Decile ranks of improvement or "average change" in residuals from 2007 to 2010 are presented in this table, based on standardized residuals for all districts in the state regressions for the relevant years. The table shows six decile ranks for reading and math at the elementary, middle, and high school levels, followed by the average of these six ranks. The next two columns show the count of the first six decile ranks that had values of 1 to 3 and the count of available decile ranks. The final column shows the percentage of available ranks that had values of 1 to 3. Decile ranks with values between 1 and 3, which represent the top 30 percent of improvement residuals in the state, and percentages of available ranks with values of 1 to 3 that are 67 percent or greater are colored orange.

Section 4: Proficiency and Advanced Proficiency Decile Ranks

Similar to Section 3, this section provides information on districts' relative performance compared with other districts within their state, but does not make any adjustment for district poverty. Proficiency deciles were determined for four subgroups (all students, African-American, Hispanic, and low-income students) by locating the reading and mathematics proficiency rates and average change in proficiency rates for each eligible district in the distribution of proficiency rates for all districts in the state.[32] Decile ranks of the percentage of students performing at the proficient or above level in 2010 were computed for each of the four subgroups. Decile ranks were also computed for the average change in the percentage of students at the proficient or above level between 2007 and 2010 for each of the four subgroups. Tables 4.1 through 4.4 summarize both the average decile ranks for each group at each education level (elementary, middle, and high school) and the number and percentage of times the districts' decile ranks were in the top 30 percent (decile ranks 1–3) of districts in the state***. In addition, for The 2011 Broad Prize, decile ranks of the percentage of students performing at the advanced level in 2010 were computed for each of the four subgroups, as well as for the average change in the percentage of students at the advanced level between 2007 and 2010***[33]***. Tables 4.5 through 4.8 summarize results for the advanced proficiency level.***

Table 4.1: Proficiency Deciles—Average DECILE RANK and count of best decile ranks across educational levels for reading and mathematics proficiency rates, for ALL STUDENTS

The table shows a summary of decile ranks at the proficient or above level for both 2010 proficiency rates, presented on the left side of the table, and average change in proficiency rates for 2007–2010, presented on the right, for all students. In each set of

[32] Only local school districts (CCD district types 1 or 2) and charter agencies (CCD district type 7) were included in this analysis.
[33] The "advanced" level was defined as the combination of all performance levels above "proficient" on a state's test.

columns, under "Avg. decile rank," the district's average decile rank across the three education levels (elementary, middle and high school) is presented first for reading and then mathematics. Under "Count of best ranks (1–3)," the first column shows the count of each district's decile ranks that were among the top 30 percent in the state (i.e., had a decile rank of 1 to 3).[34] The second column under "Avail." represents the number of 2010 proficiency rates or average change rates that received a decile rank. The third column under "Pct." represents the percent of available decile ranks that had values of 1 to 3. Average decile ranks of 3 or better (between 1 and 3) and percent values greater than 67 percent are colored orange.

Tables 4.2–4.4: Proficiency Deciles—Average DECILE RANK and count of best decile ranks across educational levels for reading and mathematics proficiency rates, among subgroups

These tables show detail on proficiency deciles at the proficient or above level for each of the three subgroups of interest (low-income, African-American, and Hispanic students). Table 4.2 summarizes proficiency decile ranks for low-income students. Table 4.3 summarizes proficiency decile ranks for African-American students. Table 4.4 summarizes proficiency decile ranks for Hispanic students. For each subgroup, the table shows a summary of decile ranks for both 2010 proficiency rates, presented on the left side of the table, and average change in proficiency rates for 2007–2010, presented on the right. In each set of columns, under "Avg. decile rank," the district's average decile rank across the three education levels (elementary, middle and high school) is presented first for reading and then mathematics. Under "Count of best ranks (1–3)," the first column shows the count of each district's decile ranks that were among the top 30 percent in the state (i.e., had a decile rank of 1 to 3). The second column under "Avail." represents the number of 2010 proficiency rates or average change rates that received a decile rank. The third column under "Pct." represents the percent of available decile ranks that had values of 1 to 3. Average decile ranks of 3 or better (between 1 and 3) and percent values greater than 67 percent are colored orange.

Table 4.5: Advanced Proficiency Deciles—Average DECILE RANK and count of best decile ranks across educational levels for reading and mathematics advanced proficiency rates, for ALL STUDENTS

The table shows a summary of decile ranks at the advanced level for both 2010 proficiency rates, presented on the left side of the table, and average change in proficiency rates for 2007–2010, presented on the right, for all students. In each set of columns, under "Avg. decile rank," the district's average decile rank across the three education levels (elementary, middle and high school) is presented first for reading and then mathematics. Under "Count of best ranks (1–3)," the first column shows the

[34] To identify districts with the best ranks, decile ranks based on proficiency rates and average change for all districts in a state were computed. Decile ranks ranged from 1 for the highest proficiency rates or largest average change to 10 for the lowest proficiency rates or smallest average change (or largest negative change) in a state. Proficiency rates or average change values that had a decile rank of 1 through 3 were considered to be the "best" in the state.

count of each district's decile ranks that were among the top 30 percent in the state (i.e., had a decile rank of 1 to 3). [35] ***The second column under "Avail." represents the number of 2010 advanced proficiency rates or average change rates that received a decile rank. The third column under "Pct." represents the percent of available decile ranks with values of 1 to 3. Average decile ranks of 3 or better (between 1 and 3) and percent values greater than 67 percent are colored orange.***

Tables 4.6–4.8: Advanced Proficiency Deciles—Average DECILE RANK and count of best decile ranks across educational levels for reading and mathematics advanced proficiency rates, among subgroups

These tables show detail on proficiency deciles at the advanced level for each of the three subgroups of interest (low-income, African-American, and Hispanic students). Table 4.6 summarizes advanced proficiency decile ranks for low-income students. Table 4.7 summarizes advanced proficiency decile ranks for African-American students. Table 4.8 summarizes advanced proficiency decile ranks for Hispanic students. For each subgroup, the table shows a summary of decile ranks for both 2010 advanced proficiency rates, presented on the left side of the table, and average change in advanced proficiency rates for 2007–2010, presented on the right. In each set of columns, under "Avg. decile rank," the district's average decile rank across the three education levels (elementary, middle and high school) is presented first for reading and then mathematics. Under "Count of best ranks (1–3)," the first column shows the count of each district's decile ranks that were among the top 30 percent in the state (i.e., had a decile rank of 1 to 3). The second column under "Avail." represents the number of 2010 proficiency rates or average change rates that received a decile rank. The third column under "Pct." represents the percent of available decile ranks that had values of 1 to 3. Average decile ranks of 3 or better (between 1 and 3) and percent values greater than 67 percent are colored orange.

Section 5: Achievement Gaps

Tables in Section 5 provide information on each district's achievement gaps. As noted above in Data Analysis Methods, three gap types were measured:

- Internal District Gap: This measure calculates the gap in performance between the district's disadvantaged and the district's advantaged students.
- Internal District vs. Internal State Gap: This measure calculates the district's internal gap minus the state's internal gap.
- External Gap: This measure calculates the gap in performance between the district's disadvantaged students and the state's advantaged students.

[35] To identify districts with the best ranks, decile ranks based on advanced proficiency rates and average change for all districts in a state were computed. Decile ranks ranged from 1 for the highest advanced proficiency rates or largest average change to 10 for the lowest advanced proficiency rates or smallest average change (or largest negative change) in a state. Advanced proficiency rates or average change values that had a decile rank of 1 through 3 were considered to be the "best" in the state.

For each of the above gap types, three subgroup comparisons were made:

- Income Gaps: These compared the performance of 1) low-income students with non-low-income students.
- Racial/Ethnic Gaps: These compared the performance of 2) African American students with White students and 3) Hispanic students with White students.

Table 5.1: Achievement Gaps—Total number and percent of achievement gaps closing, by subgroup: 2007–2010

This table shows a summary of gap closures for all three gap types for the low-income, African-American, and Hispanic subgroups compared with their non-low-income or White peers. The total possible number of achievement gap closures is 54, with 18 per subgroup. For each student subgroup and for the final set of Total columns, the *first column* under "Clos." indicates the number of gaps that were closing. The *second column* under "Avail." indicates the number of gaps for which data were available. The *third column* under "Pct." indicates the percentage of available gaps that were closing. Percent values greater than 50 percent are colored orange.

Tables 5.2 through 5.4: Achievement Gaps—Number and percent of gaps closing between subgroups: 2007–2010

These tables show detail on each of the three gap types for the three subgroups of interest (low-income, African-American, and Hispanic, respectively). Table 5.2 summarizes achievement gap closures between low-income and non-low-income students. Table 5.3 summarizes achievement gap closures between African-American and White students. Table 5.4 summarizes achievement gap closures between Hispanic and White students. For each gap type in the tables, the *first column* under "Clos." indicates the number of gaps that were closing. The *second column* under "Avail." indicates for the number of gaps for which data were available. The *third column* under "Pct." indicates the percentage of available gaps that were closing. Percent values greater than 50 percent are colored orange.

Table 5.5: Achievement Gaps—Total number of gaps that are among the smallest in the state, by subgroup: 2010

This table shows a summary of 2010 gaps for the three subgroups (low-income, African-American, and Hispanic students) that were among the smallest 30 percent of gaps in the state (with decile ranks of 1 to 3).[36] For each student subgroup and for the final set of Total columns, the *first column* under "Small" represents the number of 2010 gaps with decile ranks of 1 to 3. The *second column* under "Avail." represents the number of available gaps with decile ranks. The *third column* under "Pct." represents the percent of

[36] To identify districts with the smallest gaps, decile ranks based on gaps for all local school districts and charter agencies (where relevant) in a state (CCD district types 1, 2 or 7) were computed. Decile ranks could only be calculated for internal district gaps and ranged from 1 for the smallest gaps to 10 for the largest gaps in a state. Gaps that had a decile rank of 1 to 3 were considered to be "small."

available district gaps that had decile ranks of 1 to 3. Percent values greater than or equal to 50 percent are colored orange.

Table 5.6: Achievement Gaps—Total number of gaps that are among the fastest closing in the state, by subgroup: 2007–2010

This table shows a summary of gap closure rates for the three subgroups (low-income, African-American, and Hispanic students) that were among the fastest closing 30 percent of gaps in the state (with decile ranks of 1 to 3).[37] For each student subgroup and for the final set of Total columns, the first column under "Fast" represents the number of gap closure rates with decile ranks of 1 to 3. The second column under "Avail." represents the number of available gap closures with decile ranks. The third column under "Pct." represents the percent of available district gap closures that had decile ranks of 1 to 3. Percent values greater than or equal to 50 percent are colored orange.

Section 6: High School Graduation Rates

The tables in this section present graduation rates for the classes of 2005 through 2008 for all students, African-American students, Hispanic students, and White students.[38] Table 6.1 shows the average of the three graduation rate estimates calculated for The Broad Prize. Table 6.2 shows the rankings among the 75 eligible districts of their 2008 graduation rates and the average change in their graduation rates as reported on Table 6.1. Tables 6.3 to 6.5 show the graduation rates calculated by each method: Averaged Freshman Graduation Rate (AFGR), Urban Institute method, and Manhattan Institute method, respectively. Graduation rates for the minority subgroups may be particularly volatile due to small sample sizes across years. Therefore, the recommendation is to focus primarily on the graduation rates for all students and the average of the three rates.

In Tables 6.1 and 6.3–6.5, for each student group, the first column contains the 2008 graduation rate and the second column contains the average change in the graduation rate. "Average change" was calculated as the slope of the best fit line among available data points from 2005 to 2008, generally determined by regressing the graduation rate on year.[39] If only one data point was available, or if data were missing for both 2007 and 2008, average change was not calculated. For each student group in Table 6.2, the first column contains the ranking of the district's 2008 graduation rate and the second

[37] To identify districts with the fastest closing gaps, decile ranks based on gap closure rates for all local school districts and charter agencies (where relevant) in a state were computed. Decile ranks could only be calculated for internal district gaps and ranged from 1 for the fastest closing gaps to 10 for the slowest closing or widening gaps in a state. Gap closure rates that had a decile rank of 1 to 3 were considered to be "fast."

[38] At the time of this analysis, diploma counts were not available for 2009 and 2010. The most recent data available on district diploma counts were for the graduating class of 2008, so graduation rates were calculated for the classes of 2005 through 2008.

[39] When only two data points were available, the slope was equal to $(X_2 - X_1)/(Year_2 - Year_1)$.

column contains the ranking of the district's average change in graduation rate, among all eligible districts with available data.

Table 6.1: Graduation Rates—AVERAGES of performance (2008) and of average change (2005–2008) for the three graduation rate methods, by subgroup

This table presents an average of the three graduation rate estimates reported in Tables 6.3 through 6.5, giving equal weight to each available measure. 2008 graduation rates that are 3 or more percentage points above the eligible district average are colored orange. In addition, positive average change values are colored orange.

Table 6.2: Graduation Rate— DISTRICT RANKING of performance (2008) and average change (2005–2008) measures for the average of the three graduation rate methods, by subgroup

This table presents each district's ranking, among eligible districts with available data, of the average of the three graduation rate estimates as presented in Table 6.1. Rankings of both the 2008 graduation rate and the average change in graduation rates between 2005 and 2008 are presented. Ranks from 1 to 25 are colored orange.

Tables 6.3 through 6.5: Graduation rate using different methods

These tables show graduation rates computed using one of the three methods described above for all students and for African-American, Hispanic and white students. Under each student group, the first column shows the graduation rate for the class of 2008 and the second column shows improvement in the graduation rate from 2005 to 2008.

Section 7: College Readiness

The tables in Section 7 present college readiness measures based on three different assessments: SAT, ACT, and Advanced Placement (AP) examinations. Data are presented for each district for the following student groups: All students, African American students, Hispanic students, and White students.

The SAT and ACT are designed to assess readiness for college-level work. ***This year, scale scores for the SAT are reported by individual subject (i.e., separately for reading, writing, and mathematics) in Tables 7.1, 7.2 and 7.3, respectively. Scale scores for each subject range from 200 to 800.*** Scale scores for the composite ACT exam (covering English, mathematics, reading, and science) shown in Table 7.5 range from 1 to 36.

In order to provide some indication of whether the SAT or ACT is the predominantly used test(s) in a district, results reported on Tables 7.1 through 7.6 are de-emphasized (grayed out) when the 2010 participation rate on the relevant test was less than 20 percent for all students.

Table 7.1: SAT Reasoning Test MEAN SCORES performance (2010) and average change (2007–2010) in reading, by subgroup

This table shows the average reading score for seniors in 2010 as well as the average change in these scores from 2007 to 2010 (calculated as described above). When the 2010 SAT participation rate for all students was less than 20 percent, the district's results were de-emphasized (grayed out). For districts with at least 20 percent of students participating in 2010, scores that are 20 or more points above the eligible district average and positive change values are colored orange.

Table 7.2: SAT Reasoning Test MEAN SCORES performance (2010) and average change (2007–2010) in writing, by subgroup

This table shows the average writing score for seniors in 2010 as well as the average change in these scores from 2007 to 2010 (calculated as described .above). When the 2010 SAT participation rate for all students was less than 20 percent, the district's results were de-emphasized (grayed out). For districts with at least 20 percent of students participating in 2010, scores that are 20 or more points above the eligible district average and positive change values are colored orange.

Table 7.3: SAT Reasoning Test MEAN SCORES performance (2010) and average change (2007–2010) in mathematics, by subgroup

This table shows the average mathematics score for seniors in 2010 as well as the average change in these scores from 2007 to 2010 (calculated as described above). When the 2010 SAT participation rate for all students was less than 20 percent, the district's results were de-emphasized (grayed out). For districts with at least 20 percent of students participating in 2010, scores that are 20 or more points above the eligible district average and positive change values are colored orange.

Table 7.4: College Readiness—SAT Reasoning Test participation rates (reading, mathematics, and writing combined) performance (2010) and average change (2007–2010), by subgroup

This table shows the SAT participation rates for seniors in 2010, as well as the average change in these rates from 2007 to 2010 (calculated as described above). When the 2010 SAT participation rate for all students was less than 20 percent, the district's results were de-emphasized (grayed out). For districts with at least 20 percent of students participating in 2010, rates that are 3 or more percentage points above the eligible district average and positive change values are colored orange.

Table 7.5: College Readiness—ACT mean composite scores (English, reading, mathematics, and science combined) performance (2010) and average change (2007–2010), by subgroup

This table shows the average composite ACT score for seniors in 2010, as well as the average change in these scores from 2007 to 2010 (calculated as described above). When the 2010 ACT participation rate for all students was less than 20 percent, the district's results were deemphasized (grayed out). For districts with at least 20 percent of students participating in 2010, scale scores that are 1.0 or more points above the eligible district average and positive change values are colored orange.

Table 7.6: College Readiness—ACT participation rates (English, reading, mathematics, and science combined) performance (2010) and average change (2007–2010), by subgroup

This table shows the ACT participation rates for seniors in 2010, as well as the average change in these rates from 2007 to 2010 (calculated as described above). When the 2010 ACT participation rate for all students was less than 20 percent, the district's results were deemphasized (grayed out). For districts with at least 20 percent of students participating in 2010, rates that are 3 or more percentage points above the eligible district average and positive change values are colored orange.

Table 7.7: College Readiness—Advanced Placement (AP) percent of tests taken with scores of 3 or above (in all subjects) performance (2010) and average change (2007–2010), by subgroup

This table shows the percentage of AP exams taken by juniors and seniors in 2010 that received passing scores of 3 or above, as well as the average change in these passing rates from 2007 to 2010 (calculated as described above). Rates that are 3 or more percentage points above the eligible district average and positive change values are colored orange.

Table 7.8: College Readiness—Advanced Placement (AP) participation rates (in all subjects) performance (2010) and average change (2007–2010), by subgroup

This table shows the participation rate for juniors and seniors who took AP exams in 2010, as well as the average change in these rates from 2007 to 2010 (calculated as described above). Rates that are 3 or more percentage points above the eligible district average and positive change values are colored orange.

