URBAN
INSTITUTE · ELEVATE · THE · DEBATE

America's Gradebook: How Does Your State Stack Up?

Technical Appendix

Matthew Chingos and Kristin Blagg

This tool draws on restricted-use, student-level data on main National Assessment of Educational Progress (NAEP) reading and mathematics tests. We have access to all available datasets (through 2015) from the main NAEP, which has been administered to nationally representative samples of fourth- and eighth-grade students since 1990 (mathematics) and 1992 (reading). The most recent administration of main NAEP was in 2017, but that student-level data are not yet available to researchers.

NAEP Variables

We use data on students in the state reporting sample (RPTSAMP= 1). To minimize the time that students sit for the tests, no student takes an entire test. For the analysis, we use a statistical estimate of what each student's score on the test would have been had he or she taken the test in its entirety. For the 2013 and 2015 main NAEP, using the procedures described in the documentation provided with the restricted-use data, this estimate is based on 20 plausible test-score values and 62 replicate weights (in previous years, including 2003, there were 5 plausible values instead of 20).

For demographic adjustments of the data over time, we use a set of student-level control variables that were all available in the data from 1996 (mathematics) and 1998 (reading) onward. The variables used for the tool adjustment and their codes are as follows:

- SDRACE: race and ethnicity variable used by the National Center for Education Statistics (NCES) to report trends (white, black, Hispanic, Asian/Pacific Islander, American Indian, unclassified/two or more races)

  » This is a school-reported variable, though supplemented at times with student-reported data, and is coded as SDRACEM in the 2015 assessment and as DRACE in the 1990, 1992, 1996, 1998, and 2000 assessments.

  » The SDRACE "unclassified" category was relabeled "two or more races" in 2011. We have labeled this category as "multirace_unclass" in the Excel dataset to acknowledge this change.

- SLUNCH: eligibility for federal free and reduced-price lunch program (not eligible, eligible for reduced-price lunch, eligible for free lunch, or other or missing)

  » The SLUNCH variable is only available in the data from 1996 (mathematics) and 1998 (reading) onward. We have included data from earlier years in the Excel data file, but the adjusted scores exclude those based on models that would control for the SLUNCH variable.

- BMONTH and BYEAR: Age on February 1 of testing year, using date of birth estimated as 15th day of birth month in birth year, with ages more than two years from the mean weighted national age recoded to the mean

  » This variable is coded as MOB and YOB in the 2015 assessment

- LEP: student classified as an English language learner (yes or no)

- IEP: student classified as having a disability (yes or no)

- B018201: language other than English spoken at home (never, once in a while, about half of the time, or all or most of the time)

  » This variable is coded as B003201 in the 1998 and 2000 assessments and as B003201A in the 1990, 1992, 1994, and 1996 assessments (never, sometimes, or always); we recoded B018201 responses of "once in a while" or "about half of the time" as "sometimes", and "all or most of the time" as "always" to make the variable consistent over time.

NAEP Adjustment Methodology

All analyses were run separately by grade and subject and are weighted to be representative of the state mean (using weight variable ORIGWT).

We performed adjustments by estimating regression coefficients using the student-level data from 2003, roughly halfway between the available periods for mathematics (1996–2015) and reading (1998–2015). Specifically, we regress the test score of each student in 2003 on the set of control variables described above.[1] Control variables are included in the regression using dummy variables identifying each of the groups of students for each construct, with the exception of the one arbitrarily chosen group that is the omitted category (except for age, which is included as a continuous variable).

Using this regression, we estimate a residual for each student across all assessment years, which is the difference between their actual score and a predicted score based on the relationship between the predictors and test scores in 2003. We calculate the adjusted state score as the sum of the mean residual and the mean score in the given test year (e.g., 1996, 1998, and so on), re-normed to the mean of the given test year. Essentially, we perform an adjustment for each testing year, but we base the adjustment on the relationship between NAEP scores and the control variables in 2003.

In the data file available for download on the tool's page, we also provide adjusted scores that are comparable across time (NAEP_overtime tab). The methodology is the same as for the scores used in the tool, except that we do not re-norm the adjusted scores to the unadjusted mean of each test year. These adjusted scores can be used to calculate adjusted score changes, but they should not be used to compare adjusted and unadjusted scores in a given year (as we do in the tool using the re-normed adjusted scores).

Since we do not have student-level data for the 2017 NAEP administration, we have performed a pseudo-adjustment for the data tool. We do this by applying the 2015 adjustment—the difference between unadjusted and adjusted scores in 2015—to the 2017 state average scores as released by the NCES. The assumption is that the underlying demographics of a given state likely have not changed substantially in the two years since the last test was given. A validation analysis using the 2011 and 2013 data showed that an adjustment like this produces estimated adjusted scores that are very highly correlated ($r = 0.97$) with the actual adjusted scores.

NAEP Data Notes

Before 2003, state participation in NAEP 4th- and 8th-grade tests was voluntary. As a result, we cannot report data for all states before 2003.

Student survey data, which includes the B018201 (frequency of English spoken at home) variable, are missing for nearly all students in Alaska, so adjusted scores for Alaska are calculated using all available controls excluding B018201 status. For example, when the tool shows the adjusted scores with all controls, the Alaska adjusted score is represented by adjusting for all controls except the missing B018201. This is unlikely to be a significant limitation given that scores for the other 49 states based on this limited set of control variables are highly correlated with the metrics based on the full set of controls (coefficients range from 0.95 to 0.97 depending on subject or grade). The B018201 variable is also missing for Utah students in the 2015 assessment, so we make the same adjustment for the Utah 2015 assessments as we do for Alaska.

Except for the 2017 adjustment, all scores, including unadjusted scores, are derived from the student-level data. The scale scores for 1990, 1992, 1994, 1996, 1998, and 2000 are from test administrations where accommodations were not permitted.

In 1990 and 1992, state-level score estimates were based solely on the assessment of public school students. In the NAEP Data Explorer,[2] the NCES reports state scores based only on public school students for 1994, 1996, and 1998. Because private school students were assessed in these years, we have opted to use an estimated state score generated from the reporting sample, which includes students at both private and public schools. For most states, this change causes a 1- to 2-point difference from the official NCES score, but for some states, the difference can be larger (5 to 7 points). This method ensures comparability with later years, where both public and private school students were assessed and their scores were reported at the state level.

Students who are Native Hawaiian or other Pacific Islander are classified under "Asian/Pacific Islander" in the race variable used in the tool and data file to maintain a consistent measurement over time (Native Hawaiian and Pacific Islander students were not identified separately from the broader Asian category until 2003). Because of the high proportion of students who are Native Hawaiian or other Pacific Islander in the state of Hawaii (approximately 46 percent in 2015), adjusted scores that control for race may be misleading for this state. Table 1 shows the adjusted scores for Hawaii based on the full set of controls in the tool alongside adjusted scores that define race in a way that separates Native Hawaiian and other Pacific Islander students. These scores may be biased toward the unadjusted

score because of the large proportion (roughly 20 percent) of US Native Hawaiian students who live in Hawaii (i.e., our adjustment methodology is thus comparing Native Hawaiian students in Hawaii to themselves to a significant degree).

Full Controls Score Adjustments for Hawaii, with and without Native Hawaiian and Pacific Islander Control

	Mathematics Grade 4		Mathematics Grade 8		Reading Grade 4		Reading Grade 8	
Year	Full set of controls	Full set of controls, N.H./P.I. control	Full set of controls	Full set of controls, N.H./P.I. control	Full set of controls	Full Set of controls, N.H./P.I. control	Full set of controls	Full set of controls, N.H./P.I. control
2003	216.8	219.5	255.9	258.3	200.2	203.1	245.7	248.7
2005	219.7	222.4	255.4	257.9	201.4	204.4	242.8	245.6
2007	224.0	226.6	257.6	260.3	204.4	207.2	245.0	248.0
2009	225.2	227.9	262.6	264.7	202.1	204.8	248.0	250.8
2011	227.7	229.9	265.7	268.2	204.5	207.1	249.7	252.6
2013	232.1	234.5	270.7	273.2	205.1	207.7	254.1	257.0

Note: N.H./P.I. = Native Hawaiian/Pacific Islander.

Notes

1. This methodology differs slightly from the one used in a 2015 report (Chingos), which estimated the adjustment regression using 2013 data and excluded each state's data from the calculation of its adjusted scores. The methodology used in this report has the significant advantage of producing adjusted scores that can be compared over time, and it produces results that are highly correlated (r = 0.98 - 0.99) with those from the earlier report.

2. "NAEP Data Explorer," National Center for Education Statistics, accessed May 12, 2016, http://nces.ed.gov/nationsreportcard/naepdata/.

Reference

Chingos, Matthew. 2015. *Breaking the Curve: Promises and Pitfalls of Using NAEP Data to Assess the State Role in Student Achievement*. Washington, DC: Urban Institute.

Acknowledgments

For more information on this project, see http://apps.urban.org/features/naep/.

ABOUT THE URBAN INSTITUTE

URBAN INSTITUTE

2100 M Street NW
Washington, DC 20037

www.urban.org

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.