URBAN INSTITUTE · ELEVATE · THE · DEBATE

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

Appendix

Matthew Chingos, Kristin Blagg, and Grace Luetmer

October 2019

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 2017) 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 the main NAEP was in 2019, but those 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 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 she taken the entire test. For the 2013, 2015, and 2017 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 student-level control variables that have all been available in the data since 1996 (mathematics) and 1998 (reading). The variables used for the tool adjustment and their codes are as follows:

- **SDRACE:** Race or ethnicity variable used by the National Center for Education Statistics (NCES) to report trends (white, black, Hispanic, Asian or Pacific Islander, American Indian, and unclassified or 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 and 2017 assessments 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.
- **BMONTH and BYEAR:** Age on February 1 of testing year, using date of birth estimated as the 15th day of the birth month in the 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 and 2017 assessments.
- **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 "once in a while" or "about half of the time" as "sometimes," and we recorded "all or most of the time" as "always" to make the variable consistent over time.
  - » This variable was unavailable in the 2017 NAEP.
- **SLUNCH:** Eligibility for the 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 available in the data only 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.

- **Imputed SLUNCH variable:** A measure of free lunch eligibility in schools that use a special provision to provide school lunch. This measure is imputed by the Urban Institute from other data about the student, the student's school, and the school district and is available for the 2005 administration forward. Further information about this variable's development is in the section Free and Reduced Price Lunch Adjustment.

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. Specifically, we regress the test score of each student in 2003 on the control variables described above.[1] Control variables are included in the regression using dummy variables identifying each group of students for each construct, except for 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 her 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 web 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 2019 NAEP administration, we have performed a pseudo-adjustment for the data tool. We do this by applying the 2017 adjustment—the difference between unadjusted and adjusted scores in 2017—to the 2019 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 previous test was given. A validation analysis using the 2015 and

2017 data showed that an adjustment like this produces estimated adjusted scores that are highly correlated (r = 0.97–0.98) with the actual adjusted scores.

NAEP SLUNCH Imputation

Through special provisions available from the United States Department of Agriculture, which administers the free and reduced-price lunch program, high-poverty schools and districts have reduced or eliminated the collection of paper forms to determine eligibility for free lunch. But the share of students identified as low income, through the free lunch proxy, may be less consistently measured in these provision schools than in schools where the paper forms are used annually. This is of particular concern in schools that have adopted the Community Eligibility Provision, which allows schools with high shares of students whose families use social safety net programs, such as the Supplemental Nutrition Assistance Program and Temporary Assistance for Needy Families, to provide free lunch to all students.

When reporting the SLUNCH variable for NAEP, states have two options for determining the free lunch status of students in provision schools. All students at the school can be reported as receiving free lunch, which likely overestimates student poverty within a school. Alternatively, the school can report only students who were identified as participating in social safety net programs (a process known as direct certification), which likely underestimates the share of students who meet the program's income requirements (185 percent of the federal poverty level). To remedy this issue for our demographic adjustments, we have devised a strategy to impute free and reduced-price lunch for students from provision schools.

For the 2005 through 2017 test administrations, we use an ordered probit regression model for each year, grade, and test subject to predict the probability that each student is eligible for free, reduced-price, or paid school lunch. Although the Community Eligibility Provision was not in place in 2005, other school lunch provisions (provisions 1, 2, and 3) allowed some flexibility in collecting forms that may have affected coding of students. In provision schools, we substitute this imputed probability for the SLUNCH variable. If a school is not a provision school, we retain the information reported on the student's free lunch status.

For this adjustment, we use as much additional contextual data as we can to determine the likelihood of each student falling into the free, reduced-price, or paid lunch categories. Because the data available vary by administration, we use a slightly different set of socioeconomic variables in different test years. Further, a few states do not administer the student survey. For these states, we first impute

free lunch status for all other students in provision schools that administered the survey and then impute the free lunch status for the remaining students in nonsurvey provision schools without those additional survey questions.

In all years, we use the following school- and district-level variables in our imputation:

- whether the school is a part of the large city sample, used for comparison with Trial Urban District Assessment school districts
- the school's detailed locale designation (1 of 12 categories, such as midsize suburb, small suburb, fringe town, or remote rural); the 2005 data have a simpler local variable, with 8 categories
- the share of students eligible for targeted Title I services in the school
- the share of students that was reported as eligible for free and reduced-price lunch schoolwide
- the share of school-age children in poverty within the geographic school district, as measured by the Small Area Income and Poverty Estimates
- a state-level fixed effect control

We also use the following student-level variables for each year in our imputation (* indicates a student survey question not used for nonsurvey states:

- **2005–11**: Gender, limited English proficiency (LEP) status, special education (SPED) status, age, race or ethnicity, Hispanic origin, number of books in the home*, whether there are magazines in the home*, whether there is a computer in the home*, frequency of English spoken at home*, and parent education (eighth grade only)*.
  - » Nonsurvey state: Alaska
- **2013**: Gender, LEP status, SPED status, age, race or ethnicity, Hispanic origin, number of books in the home*, whether the student has internet at the home*, whether there is a computer in the home*, frequency of English spoken at home*, the child's parent status (e.g., single mother, single father, or foster parent)*, and parent education (eighth grade only)*.
  - » Nonsurvey state: Alaska
- **2015**: Gender, LEP status, SPED status, age, race or ethnicity, Hispanic origin, number of books in the home*, whether the student has internet at the home*, whether there is a computer in

the home*, frequency of English spoken at home*, the child's parent status (e.g., single mother, single father, or foster parent)*, and parent education (eighth grade only)*.

  - » Nonsurvey states: Alaska and Utah

- **2017**: Gender, LEP status, SPED status, age, race or ethnicity, Hispanic origin, number of books in the home*, whether the student has internet at the home*, whether there is a computer in the home*, the child's parent status (e.g., single mother, single father, or foster parent)*, and parent education (eighth grade only)*.

  - » Nonsurvey states: Alaska, Colorado, Louisiana, Montana, South Dakota, Utah, and Wyoming

NAEP Data Notes

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

Student survey data, which include the B018201 (frequency of English spoken at home) variable, are missing for and has not been available in previous years for students in Utah (in 2015) and Alaska (all years). In 2017, students were not asked on the survey about frequency of English spoken at home. As a result, adjusted scores for these states in these years are calculated using all available included controls excluding B018201 status. For example, when the tool shows the adjusted scores with all controls in 2015, the Alaska and Utah adjusted scores are represented by adjusting for all controls except the missing B018201. This is unlikely to be a significant limitation given that scores for the other 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).

Except for the 2019 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 the NAEP Data Explorer,[2] the NCES reports state scores based only on public school students for 1996 and 1998. We opted to use an estimated state score generated from the reporting sample for those years, which includes students at both private and public schools. For most states, this change causes a one-to-two-point difference from the official NCES score, but for some states, the difference can be five to seven points. From 2000 onward, state scores are reported only for public school

students. This approach is in contrast to our *Breaking the Curve* report (Chingos 2015), which includes private school students.

Students who are Native Hawaiian or other Pacific Islander are classified under Asian or 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 Hawaii, adjusted scores that control for race may be misleading for this state.

Notes

[1] This methodology differs slightly from the one used in Chingos (2015), 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 October 29, 2019, http://nces.ed.gov/nationsreportcard/naepdata/.

References

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

This research was originally funded by an anonymous donor. The 2018 update was made possible with support from the Overdeck Family Foundation. The 2019 update was supported by a grant from the Bill & Melinda Gates Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

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

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

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.