What We Can (and Can't) Learn from Publicly Available Data About K12 Math Outcomes

Ben Backes, Michael DeArmond, Elise Dizon-Ross, & Dan Goldhaber
CALDER Brief No. 36
May 2023

Math achievement plays a pivotal role in students' success in school and beyond.

Advanced math course-taking in high school can significantly increase the odds of graduating (James, 2013), success in college (Long et al., 2012), and lifetime earnings potential (Rose & Betts, 2004). But disparities in math performance and access to accelerated coursework are widespread, particularly among students of color (Asim et al., 2019; Baker et al., 2023). And gaps in math achievement have only widened during the COVID-19 pandemic (Goldhaber et al., 2023; Strunk et al., 2023).

Given math's importance, states are adopting a range of K12 policies to improve math outcomes.

Some policies focus on access and opportunity, like automatically enrolling qualified students in advanced classes (e.g., SB 2124 in Texas, RCW 28A.320.195 in Washington state, and H.R. 986 in North Carolina). Others target achievement, like mandating individualized support plans for struggling students (e.g., H.B. 7039 in Florida and part of S.B. 294 in Arkansas). And still others invest in teacher training and guidance to strengthen math instruction (e.g., H.B. 23-1231 in Colorado and California's 2023 Mathematics Framework).

As these and other policies unfold, policymakers need to monitor intended outcomes and ultimately gauge impact.

Policymakers need answers to basic questions: Are more students taking and passing algebra and advanced math? Are performance gaps changing between student groups or schools? What are the characteristics of our math teachers? Are they changing? How are math teachers with different characteristics distributed across schools and students? Policymakers also need answers to more complex questions: Did a given policy have an effect? If so, why, under what conditions, and for whom? Were there any unintended consequences?

With these kinds of questions in mind, this CALDER brief singles out the value of longitudinal, individual-level state administrative data by reviewing questions that can—and can't—be answered by the more limited data made publicly available by states. To illustrate, we focus on three states that are attempting to improve math outcomes and that enroll large numbers of students from historically marginalized student groups: California, Florida, and Texas.

What Can (and Can't) Public Data Tell Us?

Like other states, California, Florida, and Texas post school-level performance data on their state education agency websites. With a little effort, anyone can use this data to answer the question: What share of students in each state are "meeting standards" or "proficient" for different math outcomes?

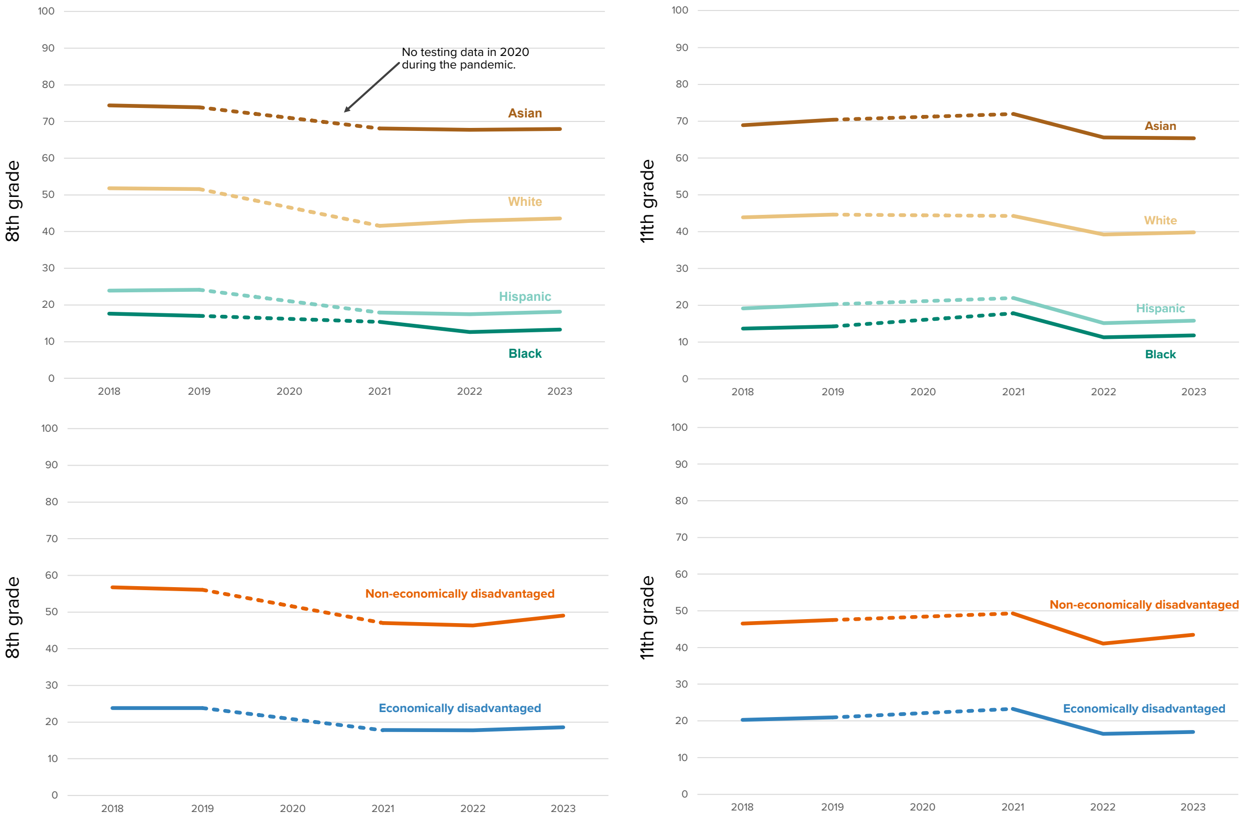
In Florida and Texas, for example, you can download data showing the share of students meeting standards in various math courses (Algebra I in Texas, Algebra I and Geometry in Florida). In California, public data cover the share of students meeting grade-level standards in math in 8th and 11th grade. Texas and California report their outcomes for different student groups, including groups based on economic advantage/disadvantage and groups based on race/ethnicity. Florida's public data does not break out results for different student groups.

What can these data tell us? In **Figure 1,** we aggregate the public, school-level data in each state for the outcomes and groups that are available on state websites for five years (2018-2023). In all three states, the results reflect the across-the-board drop in performance seen nationwide during the COVID-19 pandemic and its aftermath. In California and Texas, the plots also highlight persistent performance gaps based on social and economic advantage. (Again, because Florida does not post public data for different student groups, the state's plots have one line for all students).

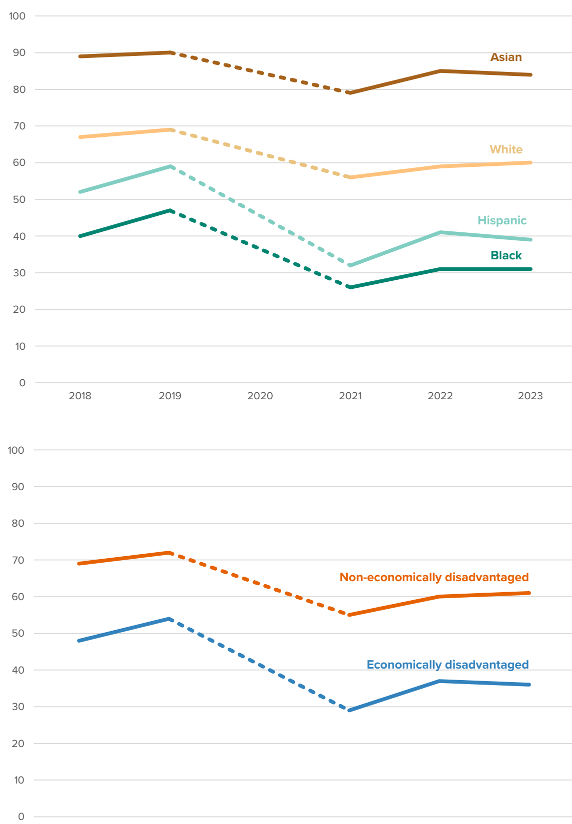
The public data in Figure 1 can tell us the share of students who are proficient in a given math outcome. But it has important limitations. First, because states use different assessments and proficiency standards, we can't compare results across states. And even

FIGURE 1. Publicly Available State Data on Middle/Secondary Math Performance in CA, FL, and TX 2018-2023

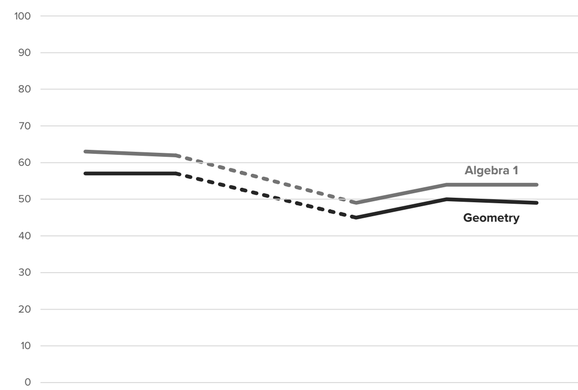
California: Share of students meeting standards in 8th and 11th grade math by race/ethnicity and economic advantage, 2018-2023



Texas: Share of students meeting standards in Algebra I by race/ethnicity and economic advantage, 2018-2023



Florida: Share of students meeting standards in Algebra I and Geometry, 2018-2023



Sources: California Department of Education, Florida Department of Education, and Texas Education Agency websites.

within states, proficiency standards can change over time, making single-state performance trends sometimes hard to interpret. Second, and less obviously, reducing performance to "proficiency" is problematic because it masks how the full distribution of scores influences our interpretations of performance. As Ho (2016) explains, our interpretations of "proficiency" are heavily influenced by where the threshold for proficiency is set and by its relationship to the underlying distribution of scores. To take just one example: it's much easier for a school to improve proficiency rates from a baseline of 50% proficient than from 10% or 90%. These and other issues lead Ho (2016) to conclude that assessing schools by their percent of students at proficiency leads to "limiting and often inaccurate" interpretations of their performance (p. 351).

Centralized public data sources can help address some—but not all—of these problems. For example, the National Assessment of Educational Progress (NAEP), known as the Nation's Report Card, facilitates cross-state comparisons because it uses a single proficiency standard [however, NAEP's standard differs from grade-level proficiency, leading to possible misinterpretations (Loveless, 2016)]. Despite this and other advantages, the NAEP has limitations: it isn't administered annually; it uses a sample of students, rather than the population of test takers in the states; and its results can be distorted by demographic shifts if not adjusted (Chingos, 2015).

A newer public data source leverages both the NAEP and state-level proficiency data to facilitate performance comparisons nationwide. By taking state proficiency rates and using the NAEP to place state proficiency thresholds on the same scale, the Educational Opportunity Project at Stanford publishes estimates of average test scores, learning rates, and trends in average test scores across states. The approach addresses concerns about comparability and the influence of underlying distributions.

Although the NAEP and Stanford data do a better job monitoring performance across states, they can't answer more complex questions about what is driving achievement or questions about causal policy impacts. For that, we need more detailed data.

What Can More Detailed Data Tell Us?

Recognizing the value of individual-level longitudinal data for addressing key policy questions, the National Center for Education Statistics launched the Education Department's Statewide Longitudinal Data System (SLDS) grant program in 2006. The SLDS program provided resources for states to develop K12 data systems and connect them with data from other agencies (e.g., early learning, postsecondary, and workforce data). As of 2023, IES has awarded more than $900 million in total grants through the program (Institute of Education Sciences, 2023).

Unlike the public, school-level data, SLDS data includes individual-level data about students and teachers. Because the data includes unique identification numbers for individuals, SLDS systems can link data across years. This level of detail raises important privacy concerns, so states only share this type of data under strict rules and in ways that protect the identity of the students and teachers who appear in the data.

Longitudinal, individual-level data in SLDS systems have significant advantages for answering questions about system performance and policy effectiveness. Researchers can use these data to follow students' progress over time and to understand the factors that influence their achievement and progress (Figlio et al., 2017).

For example, this granular data would allow researchers to model relationships between math scores and factors, such as prior achievement, course access, student mobility, and exposure to effective teachers. The data could be used to evaluate policy effects on important outcomes, like achievement gaps, course access/taking, and the compositions of the teacher workforce. If states link K12 data to post-secondary data systems, researchers and decision-makers could potentially track policy impacts on longer-term outcomes beyond K12, like college attendance and career outcomes (Bleiberg, 2017).

A recent study in Texas illustrates the kinds of questions these data can answer. Researchers used detailed SLDS data to see whether a comprehensive educator evaluation and compensation reform in the Dallas Independent School District improved student achievement (Hanushek et al., 2023). Besides allowing the researchers to estimate value-added measures of teacher effectiveness, the state's SLDS allowed them to track teacher movement between schools and districts. This meant they could investigate not only whether the reform improved student achievement but also whether it did so, in part, by shifting the workforce's composition (as opposed to increased effort or some other driver). They concluded not only that the reforms raised math and reading achievement, but they attributed 15% of the math gains to shifts in the composition of the district's teachers. Although national assessments haven't found positive effects from evaluation reform (Bleiberg et al., 2021), studies from districts with strong implementations, like Dallas and Washington D.C. (Dee and Wyckoff, 2015, James and Wycoff, 2020), suggest they may hold more promise.

Of course, SLDS data aren't perfect. Administrative data can have errors (Goldhaber et al., 2019), and it does not include data on all the student outcomes policymakers care about (Conaway et al., 2015). Fortunately, it's possible to combine SLDS data with other sources of information to address issues like how instructional materials, instructional time, educator perspectives, or other school-level practices and policies influence student outcomes. For instance, researchers analyzed state test scores along with information on math textbook adoption and usage—including teacher surveys—across 6,000 schools in six states (California, Louisiana, Maryland, New Jersey, New Mexico, and Washington) (Blazer et al., 2020). Contrary to some prior evidence, the analysis didn't find significant evidence of an association between different math textbooks and variations in student achievement growth.

The bottom line: As states embark on a new wave of initiatives to improve math outcomes, policymakers will face important questions about math access and achievement, and the factors that shape them. Answering those questions will require leveraging—and continuing to invest in—the detailed longitudinal data collections housed in SLDS systems.

References

Asim, M., Kurlaender, M., & Reed, S. (2019). 12th grade course-taking and the distribution of opportunity for college readiness in mathematics. Policy Analysis for California Education.

Baker, M., & Wade, G. (2023). Opportunities denied: High achieving Black and Latino students lack access to advanced math. Education Trust.

Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S., Goldhaber, D., Harris, D., Hitch, R., Holden, K., & Kurlaender, M. (2020). Curriculum reform in the Common Core era: Evaluating elementary math textbooks across six U.S. states. Journal of Policy Analysis and Management, 39, 966–1019.

Bleiberg, J. (2017). Policymakers shouldn't underestimate the value of high-quality education data [Commentary]. Brookings Institution.

Chingos, M. M. (2015). Breaking the curve: Promises and pitfalls of using NAEP data to assess the state role in student achievement [Research Report]. Urban Institute.

Conaway, C., Keesler, V., & Schwartz, N. (2015). What research do state education agencies really need? The promise and limitations of state longitudinal data systems. Educational Evaluation and Policy Analysis, 37(1_suppl), 16S-28S.

Figlio, D., Karbownik, K., & Salvanes, K. (2017). The promise of administrative data in education research. Education Finance and Policy, 12(2), 129–136.

Goldhaber, D., Holden, K. L., & Grout, C. (2019). Errors in administrative education data: A cautionary tale. Educational Researcher, 48(3), 179–182.

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2023). The educational consequences of remote and hybrid instruction during the pandemic. American Economic Review: Insights, 5(3), 377–392.

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. Educational Researcher, 37(6), 351–360.

Institute of Education Sciences (2023). History of SLDS grant program: Expanding states' capacity for data-driven decisionmaking. U.S. Department of Education. https://nces.ed.gov/programs/slds/pdf/History_of_the_SLDS_Grant_Program_Oct2023.pdf

James, J. (2013). The surprising impact of high school math on job market outcomes (Economic Commentary 2013–14). Federal Reserve Bank of Cleveland.

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. American Educational Research Journal, 49(2), 285–322.

Loveless, T. (2016). The NAEP proficiency myth [Commentary]. Brookings Institution.

National Academies of Science, Engineering, and Medicine (2022). The future of education research at IES: Advancing equity-oriented science. The National Academies Press.

Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. The Review of Economics and Statistics 86(2), 497-513.

Strunk, K. O., Hopkins, B., Kilbride, T., Imberman, S. A., & Yu, D. (2023). The path of student learning delay during the COVID-19 pandemic: Evidence from Michigan. EPIC, Michigan State University.