

Measuring racial educational disparities over time amongst top achievers

<https://doi.org/10.1038/s41586-024-08536-0>

Udit Karna¹, Min Sok Lee², John A. List^{2,3,4}✉, Andrew Simon⁵ & Haruka Uchida²✉

Received: 25 March 2024

Accepted: 17 December 2024

Published online: 12 February 2025

 Check for updates

Educational disparities remain a key contributor to increasing social and wealth inequalities. To address this, researchers and policymakers have focused on average differences between racial groups or differences among students who are falling behind¹. This focus potentially leads to educational triage, diverting resources away from high-achieving students, including those from racial minorities^{2,3}. Here we focus on the ‘racial excellence gap’—the difference in the likelihood that students from racial minorities (Black and Hispanic) reach the highest levels of academic achievement compared with their non-minority (white and Asian) peers. There is a shortage of evidence that systematically measures the magnitude of the excellence gap and how it evolves^{4,5}. Using longitudinal, statewide, administrative data, we document eight facts regarding the excellence gap from third grade (typically ages 8–9) to high school (typically ages 14–18), link the stability of excellence gaps and student backgrounds, and assess the efficacy of public policies. We show that excellence gaps in maths and reading are evident by the third grade and grow slightly over time, especially for female students. About one third of the gap is explained by a student’s socioeconomic status, and about one tenth is explained by the school environment. Top-achieving racial minority students are also less likely to persist in excellence as they progress through school. Moreover, state accountability policies that direct additional resources to reduce non-race-based inequality⁶ had minimal effects on the racial excellence gaps. Documenting these patterns is an important step towards eliminating excellence gaps and removing the ‘racial glass ceiling’.

Educational disparities have been a key contributor to increasing social and wealth inequities over the past few generations. This has motivated modern societies to focus on closing opportunity gaps in education. In the USA, for example, attempts to close achievement gaps among children from different racial groups have typically come in the form of advancing ‘one-size-fits-all’ accountability policies, such as No Child Left Behind (NCLB)^{1,7–13}. The emphasis on targeting differences in specific average scores in high-stakes tests leads schools to participate in educational triage: ignoring the lowest and highest-performing students and rationing resources towards students who are most likely to improve aggregate scores^{2,14,15}.

Such triaging potentially creates a nefarious externality across demographic groups when applied across school districts, serving to increase gaps between races in the right tail of the distribution. These racial excellence gaps—the difference in the percentage of students between racial groups that reach the top levels of overall performance—can have broad deleterious social effects. For the purposes of this study, we compare students in under-represented racial groups (Black, Hispanic and other students; collectively called ‘minority’ students here) in the highest level of education and science,

technology, engineering and mathematics (STEM) fields to ‘non-minority’ (white or Asian) students. For instance, there will be a deficiency of minority representation in scarce, high-level jobs, high-ranking policy officials and other key decision-making roles. A lack of such representation can have deep social multiplier effects, wherein the next generation views human capital investment as having a low return. In this spirit, policies that target and eliminate excellence gaps can help to achieve both educational equity and set a path to shatter the ‘racial glass ceiling’.

At this point, however, there is a shortage of evidence concerning the excellence gap, and we do not understand its causes and consequences at a deep enough level to advance useful policy advice⁴. Here, to fill this gap, we provide systematic, longitudinal evidence of the excellence gap by following students from third grade (typically ages 8–9) to the end of high school (typically ages 14–18). We use comprehensive administrative data from the state of Michigan, USA to study trends in racial differences regarding those who reach the highest levels of academic achievement over time—that is, as they age—to establish several features of excellence gaps.

We first document the extent of the excellence gap and its drivers. We then examine variability of excellence gaps across schools.

¹Department of Economics, Columbia University, New York, NY, USA. ²Department of Economics, University of Chicago, Chicago, IL, USA. ³Research School of Economics, Australian National University, Canberra, Australian Capital Territory, Australia. ⁴NBER, Cambridge, MA, USA. ⁵Frank Batten School of Leadership and Public Policy, University of Virginia, Charlottesville, VA, USA.

✉e-mail: jlist@uchicago.edu; uchida@uchicago.edu

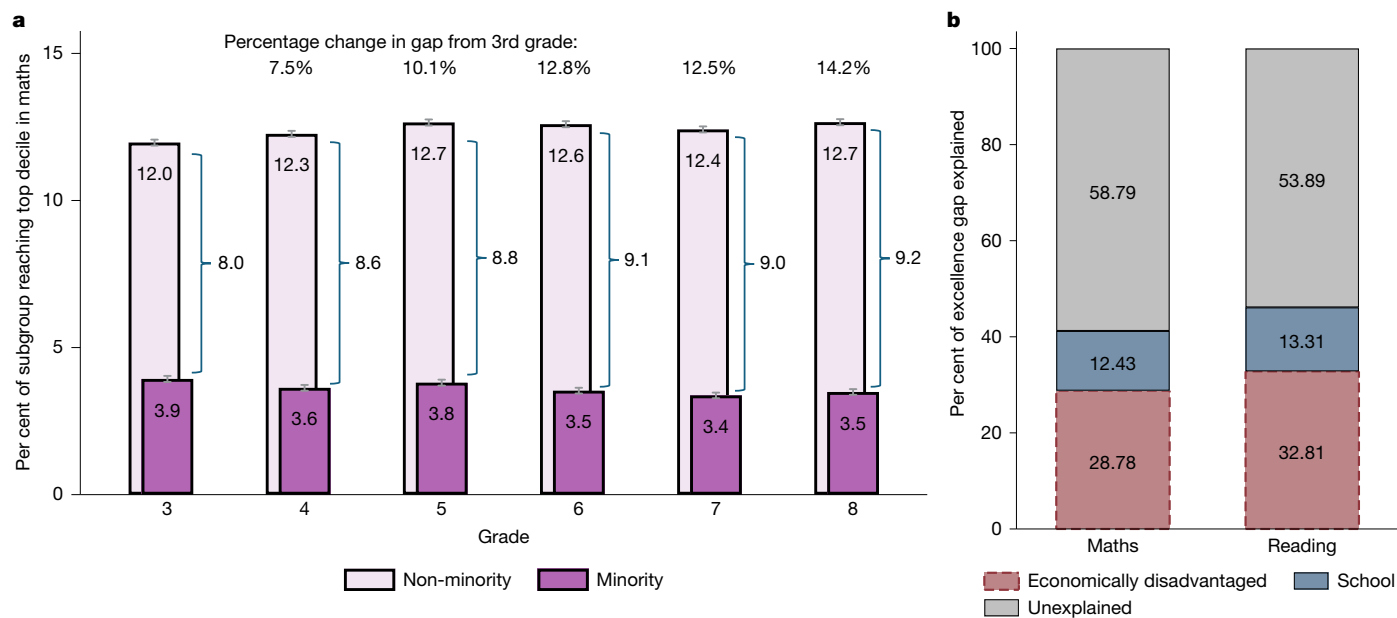


Fig. 1 | Racial excellence gaps from third to eighth grade. a, The likelihood that a minority (Black or Hispanic) or non-minority (white or Asian) student reaches excellence. Data are mean $\pm$ 95% confidence interval. **b**, Gelbach decomposition of the minority and non-minority difference in reaching excellence in maths, pooling students from third to eighth grade. Excellence is defined as reaching the top decile on the maths or reading statewide standardized exam, Michigan Educational Assessment Program (MEAP), for all years until 2014, and Michigan Student Test of Educational Progress (M-STEP)

for all years after 2014. ‘Economically disadvantaged’ is an indicator for whether a student is classified as economically disadvantaged by Michigan, which is defined by whether the student is determined to be eligible for free or reduced-price meals, is in a household that receives food or cash assistance, is homeless, a migrant, or in foster care. ‘School’ refers to the inclusion of school fixed effects. The analysis uses our main sample, as defined in Methods, which consists of 534,122 students in total: 136,137 classified as racial minority students and 397,985 are classified as non-minority students.

We additionally consider whether and how public policies, such as accountability programmes, affect the dynamics of the excellence gap. Finally, we use the richness and longitudinal nature of our data to document various facts around mobility in and out of the highest levels of achievement and explore potential drivers of mobility.

Our main contribution is to provide a comprehensive documentation of the excellence gap over time using longitudinal, statewide data that follows students from third grade to the end of high school. This extends past work, which has been more limited in scope and based on available survey data (for example, in refs. 4,5,16–18). We additionally consider the potential drivers of the excellence gap and show that racial differences in socioeconomic status help to explain 29% of the excellence gap for maths and 33% for reading. This acknowledges the important role of poverty and family resources^{19,20}. We also find that school factors have a smaller but notable role and explain 12% and 13% of the maths and reading excellence gaps, respectively, complementing past work that has explored the role of schools in achievement gaps²¹. About half of the excellence gap is explained by factors that are not covered in our data.

We further contribute to the growing body of evidence²² on whether and which policies can meaningfully affect the excellence gap. We do so in two ways. First, we document the importance of specific drivers, which can be helpful for understanding which policies may work, especially since we find that school-specific factors only explain about 10% of the excellence gaps. Second, we estimate the causal effects of past Michigan accountability policies on the excellence gap and educational triage. This analysis specifically adds to the mixed evidence on the effects of K-12 school accountability interventions^{22–24}. Our results indicate that the Michigan accountability policies had little impact on the excellence gap, building on descriptive quantitative and qualitative work (for example, in refs. 5,25–32). More broadly, our results highlight the need to identify policies that jointly pursue the goals of equity and excellence.

Setting

We use statewide administrative data on all public school students in the state of Michigan. First, we present summary statistics and the relevant US averages (Extended Data Table 1) to understand the representativeness of our sample and explore how observables that are correlated with preferences, beliefs and constraints vary across our setting and target settings of interest³³. We find that Michigan is largely representative of the USA in terms of demographics, school investments and test scores. However, Michigan requires and pays for high school students to take a college entrance exam³⁴. This is useful for our purposes, as it means that nearly all students take the SAT, compared with only about half of students nationally. Since Michigan was an early adopter of this policy, we believe that the excellence gaps among high school students that we document are probably smaller (closer to zero in magnitude) than those in states that do not have this policy. However, many more states now have similar requirements. Supplementary Table 1 provides additional descriptive statistics.

The excellence gap emerges by third grade

Figure 1a illustrates the likelihoods of minority and non-minority students reaching ‘excellence’—defined as scoring in the top decile of achievement in standardized exams—in maths from the third to eighth grades (results for reading are shown in Supplementary Fig. 2). Following past work on the excellence gap^{4,5}, we define the racial excellence gap as the difference in the likelihood of reaching the top decile of achievement between Black or Hispanic students versus white or Asian students. We find that the excellence gap already exists for the youngest students in our data (third graders). Indeed, the data suggest a large and significant disparity: non-minority students are 3 times more likely to be in the top decile compared with minority students (two-tailed *t*-test of indicator for minority on likelihood in top 10%,

t -statistic = -34.50 ; $P < 0.001$; $n = 534,122$). Although the magnitude of the gap is smaller for reading test scores than maths test scores (7.2 versus 8.8 percentage points), non-minority students are still more than 2.5 times more likely to reach the top decile in reading than minority students (two-tailed t -test, t -statistic = -40.78 ; $P < 0.001$; $n = 534,122$) (Supplementary Fig. 2). These magnitudes are around 80% of the disparity in excellence by socioeconomic status (Supplementary Fig. 1). Although our data begin at grade 3, past evidence shows that the gap emerges even earlier, including as early as 2 years of age^{35,36}.

The excellence gap persists throughout schooling years

Figure 1a and Extended Data Fig. 1 capture our second fact: the excellence gap in maths increases with grade level from 8.0 percentage points (two-tailed t -test of indicator for minority on likelihood in top 10%, t -statistic = -34.50 ; $P < 0.001$; $n = 534,122$) among third graders to 9.2 percentage points (two-tailed t -test of indicator for minority on likelihood in top 10%, t -statistic = -23.44 ; $P < 0.001$; $n = 534,122$) among eighth graders. This widening of the gap is modest in size, with about half of this small change occurring between third and fourth grades (two-tailed t -test of indicator for minority on likelihood in top 10%, t -statistic = -32.81 ; $P < 0.001$; $n = 534,122$; Supplementary Table 3). This increase arises owing to differential rates of entry and exit across grades: Asian students are increasingly more likely, whereas Black students are less likely, to be in the top decile in higher grades (Supplementary Fig. 3).

One potential reason for these across-grade patterns is that students from minority groups are more likely to switch schools, particularly across transition grades, and this disrupts learning. Indeed, across grades four to eight, students from minority groups are more than four percentage points more likely to be in a different school than in the previous year than non-minority students (two-tailed t -test, t -statistic = 9.28 , $P < 0.001$, $n = 2,670,610$), from a baseline of 30%. However, accounting for differential rates of school switching minimally changes our estimates of the excellence gaps (Supplementary Table 3).

To provide deeper insights into this result, we show the achievement gaps by grade in Supplementary Fig. 6. Past work has explored the lack of growth in racial achievement gaps³⁷. However, the achievement gap and the excellence gap underscore two very distinct dimensions of racial gaps in education. Achievement gaps highlight differences near the centre of each group's distribution. Our measure of excellence examines specifically the composition of the right tail of the distribution of academic achievement.

Socioeconomic status and schools help explain excellence gaps

We conduct a Gelbach decomposition³⁸ to understand the role of each covariate in explaining the excellence gap. We find that racial differences in socioeconomic status explain roughly one third of the excellence gap in maths, whereas racial differences in schools attended explain around 12% (-2.53 and -1.09 percentage points for socioeconomic status and racial differences, respectively; two-tailed t -test, t -statistic = -37.18 , -6.44 , respectively; $P < 0.001$ for both, $n = 3,204,732$; Fig. 1b). The patterns are similar for reading achievement. Because our measures of economic disadvantage are coarse, our analyses are likely to provide lower bounds on the extent to which socioeconomic status explains the excellence gap. We interpret the school fixed effects to capture a large set of factors, such as spending policy, teacher quality and peer effects, and, moreover, they could be correlated with county-wide and city-wide features. School fixed effects are likely to capture a variety of local features both within and outside of schools.

In sum, these estimates provide support for our third fact, as these factors jointly explain much of the gap. The variation explained by socioeconomic status and school captures both the causal impact

of these characteristics as well as other potential correlating factors. Future work should explore the causal mechanisms through which these attributes explain achievement differences.

We contrast these findings on the excellence gap with patterns that describe racial disparities in other parts of the achievement distribution. We conduct the same decomposition exercise for racial differences in surpassing lower levels of achievement: reaching above the median or bottom 10% of achievement (Supplementary Fig. 7). We find that variation in schools explains a substantially large portion of these lower-level racial differences; 40% for the bottom 10% of students and 25% for the bottom 50% of students in maths (two-tailed t -test, t -statistic = -18.06 and -15.98 , respectively; $P < 0.001$ and $n = 3,204,732$ for both). For comparison, socioeconomic status explains 17% of the difference for the bottom 10% of students and 23% for the bottom 50% of students in maths (two-tailed t -test, t -statistic = -48.86 and -54.54 , respectively; $P < 0.001$ and $n = 3,204,732$ for both). This suggests that the drivers of the excellence gap are not necessarily the same drivers of disparities in lower levels of academic performance.

Finally, we estimate the main specification (equation (2)), but with school-by-cohort-by-grade fixed effects to address concerns that school quality and neighbourhood traits change across cohorts. The results minimally change (Supplementary Fig. 4).

Female students drive racial excellence gap growth

The direction of gender differences in the racial excellence gap depends on the subject (Extended Data Fig. 2). Across grades, the racial excellence gap in maths is larger among male students than among female students (for instance, 9.9 percentage points for male students versus 8.5 percentage points for female students in eighth grade; two-tailed t -test of gender difference, t -statistic = 8.69 ; $P < 0.001$; $n = 534,122$), whereas the racial excellence gap in reading is larger among female students (for instance, 6.3 percentage points for male students versus 8.9 percentage points for female students in eighth grade; two-tailed t -test of gender difference, t -statistic = -14.98 ; $P < 0.001$; $n = 534,122$). These patterns are probably driven by the fact that male students are more likely to reach excellence in maths than female students, whereas the gender difference is reversed for reading.

By contrast, in both subjects, the change in the excellence gap from third to eighth grade is substantially larger among girls than among boys, with the gap among girls growing by 1.7 percentage points in maths and 1.4 percentage points in reading (two-tailed t -test, t -statistic = 3.88 and 4.09 , respectively; $P < 0.001$ for both, $n = 537,218$), whereas there is no evidence that the gap grows among boys for both subjects (two-tailed t -test, t -statistic = 1.31 and -0.28 , respectively; $P = 0.190$ and $P = 0.777$, respectively; $n = 531,026$). These results differ from previous work on the left tail of the achievement distribution that finds that economically disadvantaged Black boys are more likely to fall behind as they progress through school³⁹. In Supplementary Information, section E, we show that these gender differences are explained by baseline differences in achievement, particularly for maths.

Excellence gaps vary across schools

We next investigate how racial differences in excellence vary across schools. Figure 2 illustrates the distributions of school-level maths and reading excellence gaps, showing that the average gap masks substantial heterogeneity. We reject the null hypothesis that excellence gaps are homogeneous across schools in both subjects (Pearson chi-squared test, $\chi^2 = 29,698.2$, $11,727.9$ for maths and reading, respectively; $P < 0.001$ for both; $n = 1,940,195$). Although excellence gaps vary across grades, this across-school heterogeneity is not driven by differences in grades associated with schools: accounting for grades minimally changes the results (for example, we draw the same conclusion when subsetting to third and fourth graders: Pearson chi-squared test, $\chi^2 = 10,034.8$ and

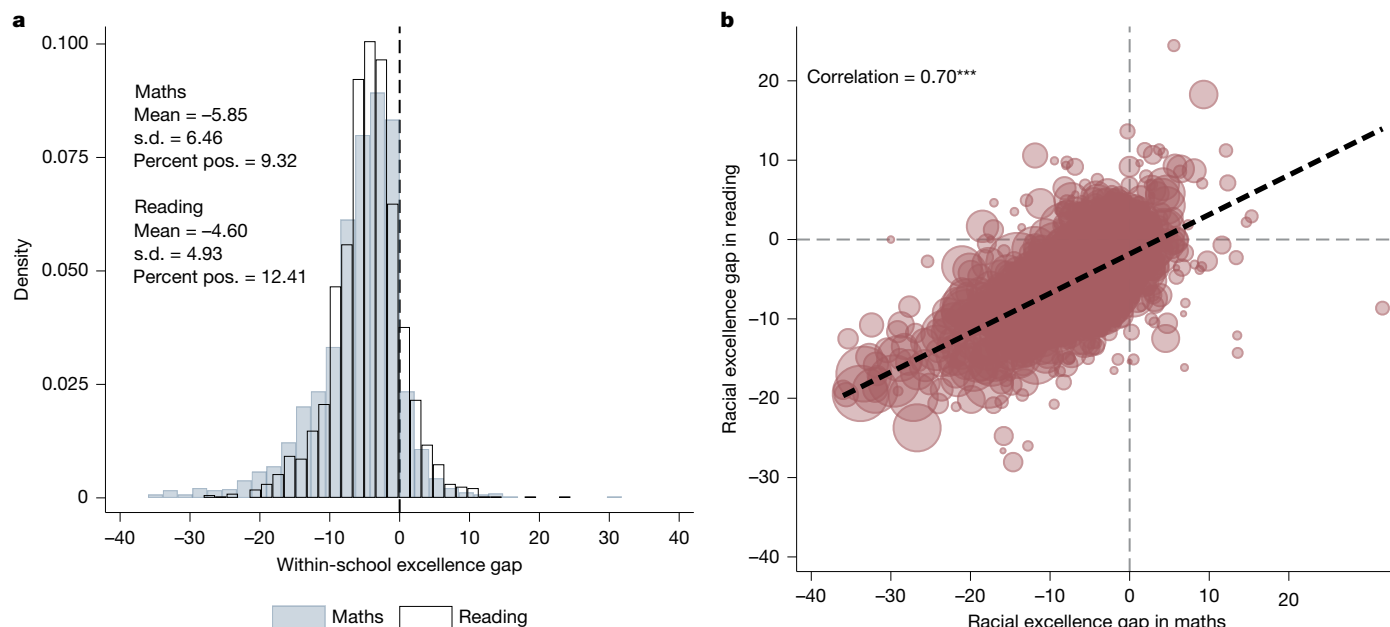


Fig. 2 | Excellence gaps across schools. **a**, The distribution of the maths and reading excellence gaps across schools. Percent pos., the percentage of schools that have a non-negative excellence gap in the corresponding subject. Mean and s.d. of the school-level excellence gap for maths and reading are indicated. **b**, Relationship between a school's excellence gap in maths and in reading. Positive values of the excellence gap are in favour of minorities. The unit of observation is a school. These measures pool students across years. The linear

fit weights schools by the number of students who take a standardized exam in the school. Excellence is defined as reaching the top decile on the maths or reading statewide standardized exam, MEAP for all years until 2014 and M-STEP for all years after 2014. The analysis uses our main sample, as defined in Methods, which consists of 534,122 students in total: 136,137 classified as racial minority students and 397,985 classified as non-minority students.

3472.6 for maths and reading, respectively; $P < 0.001$ for both; $n = 1,189, 1,198$). While students from minority groups are around 5 percentage points less likely to reach excellence in the average school, this gap widens to around 10 to 13 percentage points in schools in the 10th percentile of the excellence gap. We estimate a s.d. of 5–6 percentage points in both subjects. Notably, our observed school-grade excellence gaps are quite stable over time (Supplementary Fig. 8).

However, despite across-school variation in the magnitude of the excellence gap, less than 10% of all schools exhibit a positive excellence gap in maths (fraction of schools with positive maths gap; two-sided t -test, bootstrap t -statistic = 14.34; $P < 0.001$; 95% confidence interval = [8.04, 10.59]; $n = 1,975$). This pattern persists across grades (Supplementary Fig. 9). Similarly, less than 13% of all schools have a positive excellence gap in reading (fraction of schools with positive reading gap; two-sided t -test, bootstrap t -statistic = 14.34; $P < 0.001$; 95% confidence interval = [10.95, 13.86]; $n = 1,975$). In other words, non-minority students reach excellence at higher rates than minority students in the majority of schools.

We find that a school's excellence gaps across subjects are positively linked: the correlation between a school's maths and reading excellence gap is 0.70 (Pearson's correlation, t -statistic = 43.33; $P < 0.001$; $n = 1,975$) (Fig. 2b). We find similar patterns when measuring excellence gaps as proportional differences in the rates of reaching excellence (Supplementary Fig. 10).

From regressing a school's excellence gap on its demographic composition and average student achievement, we find that the schools that display the most negative excellence gaps are those with a larger fraction of minority students and economically disadvantaged students (two-tailed t -tests, t -statistic = -11.62 and -2.07, respectively; and $P < 0.001$ and $P = 0.039$, respectively; $n = 1,975$) (Extended Data Table 2), and we find similar patterns for predicting whether a school has a non-negative excellence gap (Supplementary Table 4). Schools located in suburbs (around 45% of schools in our data) tend to exhibit more positive excellence gaps (two-tailed t -test, t -statistic = 2.77;

$P = 0.006$; $n = 1,975$), after accounting for differences in achievement level and demographic composition.

Past policies failed to reduce excellence gaps

We now consider whether and how public policies may contribute to or reduce the racial excellence gap. More specifically, and in light of our results that the school fixed effects explain a meaningful portion of the excellence gap, we aim to further understand the extent to which the excellence gap from third to eighth grade is driven by differences in school resources by examining the causal effect of state accountability policies on which students reach excellence.

Past work finds that test-based accountability policies that target average student achievement or proficiency can cause schools to redirect resources away from the highest-achieving students^{3,6}. In that case, the benefits for some students come at the expense of others. If the resources have a larger impact on low-income or minority students⁴⁰, or if schools are more likely to redirect resources away from high-achieving minority students than high-achieving non-minority students because of differences in social capital⁴¹, then the excellence gap will widen.

We investigate the effects of a post-NCLB policy in Michigan that was designed specifically with the goal of reducing non-race-based within-school inequality in highly disadvantaged schools. Starting in 2012, the lowest-performing schools were given a 'Priority' designation, and were required to make considerable changes to the school organization and instruction. Changes were made at the discretion of the individual schools, and ranged from replacing the principal and increasing professional development to formally closing and reopening the school. Priority schools were required to use 10% of their Title I funding to support the implementation. Schools not given a Priority designation could be deemed 'Focus' schools. These schools were required to undergo district-led changes. They were assigned a district improvement facilitator and received guidance on and toolkits for

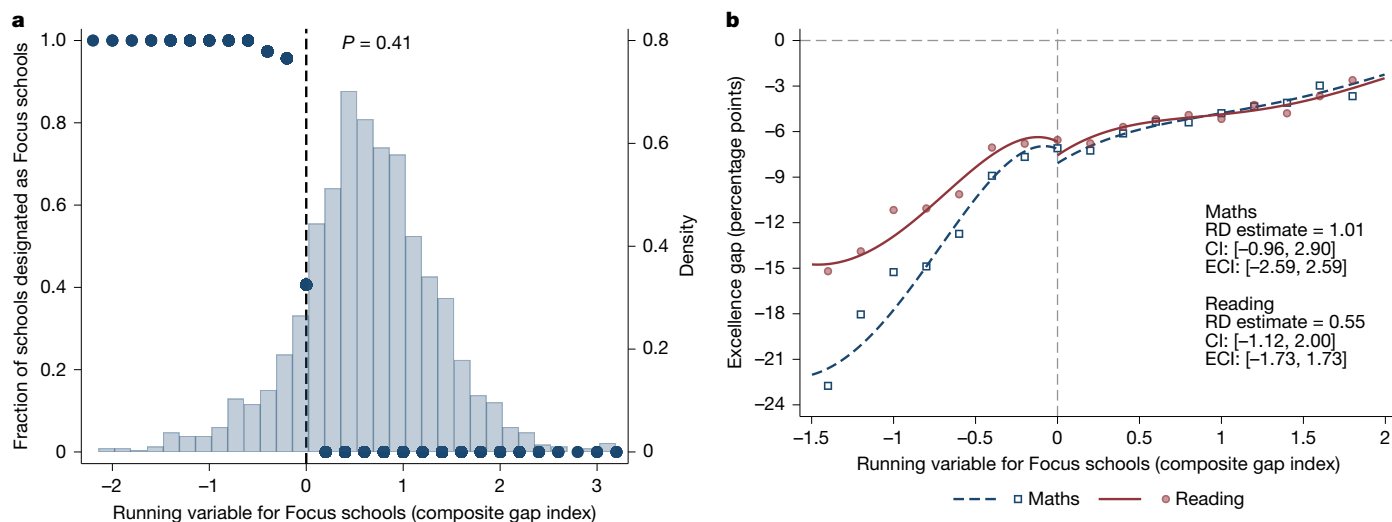


Fig. 3 | Effect of Focus designations on excellence gaps. a, The distribution of the school-level composite gap index, which is the running variable used to determine Focus designation. The scatter plot (left vertical axis) bins schools by every 0.2 units of the index, and the shaded bins correspond to a histogram (right vertical axis) of the running variable. The P value corresponds to a test of manipulation using local polynomial density estimation⁴³ against the null hypothesis of equality at the cutoff. **b**, Effect of Focus school designations on school-level racial excellence gaps in maths and reading. Schools to the left of the Focus cutoff (those with negative values of the running variable) are designated as Focus schools. Lines represent local linear fits⁴⁸ and the scatter plot bins observations by 0.2 units of the running variable. Positive values of

the excellence gap are in favour of minorities. Observations are at the school year level, and each regression pools observations from the first, second, third and fourth year after policy designation. Excellence is defined as reaching the top decile on the maths or reading statewide standardized exam, MEAP for all years until 2014 and M-STEP for all years after 2014. CI refers to the robust bias-corrected 95% confidence interval⁴⁹. ECI is the equivalence test 95% confidence interval centred around 0 (ref. 50). All specifications cluster at the school level. The sample consists of 1,745 unique schools, whose achievement measures are constructed using 1,739,566 unique student-year observations: 356,570 correspond to minority students, and 1,382,996 correspond to non-minority students.

how to reduce within-school gaps. Schools identified as Priority or Focus retained that designation for four years. Our analysis of these policies directly builds on refs. 6,42, which also provide more details on the policy.

Michigan schools were identified as Priority or Focus schools using a top-to-bottom ranking. A school was designated as a Priority school if it was in the bottom 5% of a school performance index that the state created using measures on achievement gaps (25% weight), student achievement (50%) and student growth (25%) over time. A school that was not assigned Priority status was designated as a Focus school if its composite gap index, which consisted of measures of achievement gaps in five subjects (mathematics, reading, science, social studies and writing), fell below the score corresponding to the tenth percentile of Title I schools.

The assignment mechanisms for the Focus and Priority designations enable us to estimate the causal effect of this accountability policy using a regression discontinuity design, in which the school performance index (for Priority) and composite gap index (for Focus) are the running variables. We estimate the causal effect of being assigned the Focus or Priority designation on the excellence gap in the first four years of implementation following the regression discontinuity design in past work^{6,42} that finds that the Focus designation shrank the within-school maths achievement distribution (the difference between the 90th and 10th percentiles) in the short run, but that this effect dissipated quickly. We build on those works by analyzing the impacts on excellence (90th percentile across all schools) and differential effects by race.

We assess the suitability of the regression discontinuity research design for Focus and Priority designations. First, we verify that the running variables and their respective cutoffs indeed generate substantial discontinuities in whether a school is deemed as Focus or Priority (Fig. 3a and Supplementary Fig. 12a; two-tailed test of first stage coefficient, t -statistic $> 10^{14}$ for Priority and t -statistic = 59.69 for Focus; $P < 0.001$ for both; $n = 1,745$ for Focus and 1,797 for Priority). Second, we investigate the densities of the running variables⁴³ and

find no strong evidence of discontinuities in density at the cutoffs using a test of manipulation (two-tailed t -test, robust t -statistic = 0.84, 0.82; $P = 0.410$ and 0.402; $n = 1,745$, 1,797 for Focus and Priority, respectively). Third, we conduct numerous placebo tests, using pre-determined school characteristics (baseline maths excellence gap, reading excellence gap, fraction minority, fraction economically disadvantaged, log enrolment and urbanicity) as outcomes, and find little evidence for imbalance (Supplementary Figs. 13 and 14; For Priority; two-tailed robust t -statistic = 0.51, 0.61, 0.54, 0.09, 0.02 and 1.80 for each listed baseline characteristic, respectively; $P = 0.682$, 0.337, 0.709, 0.453, 0.671, 0.382; $n = 6,592$; For Focus; two-tailed robust t -statistic = 0.67, 0.61, 0.30, 0.46, 0.25 and 0.65, respectively; $P = 0.500$, 0.841, 0.580, 0.343, 0.810, 0.629; $n = 6,737$). This garners support for our identification assumption that there are no imbalances on unobservables that affect our outcomes of interest. Together, we confirm that the setting is appropriate for a regression discontinuity design. Our estimates therefore capture the local average treatment effect of each designation on the excellence gap, and the likelihood of reaching excellence by race, by comparing schools with test scores indexes just below the cutoff to schools just above the cutoff. We provide more information about the policy and the research design in Methods.

We consider the effect of the accountability policies on school-level excellence gaps, as computed in previous sections. Figure 3 provides a summary for Focus and Supplementary Fig. 12 summarizes for Priority (Supplementary Table 7). We find that neither a Priority nor Focus designation significantly impacted school-level maths and reading excellence gaps (two-tailed t -tests, all robust P values greater than 0.325; $n = 6,858$ for Priority and $n = 6,721$ for Focus). However, we acknowledge that the point estimate on maths is about 10% of the excellence gap and similar in magnitude to the growth in the excellence gap between third and eighth grades. The confidence intervals for Priority designation are large, given that relatively few schools were assigned a Priority status (5% of all schools by construction).

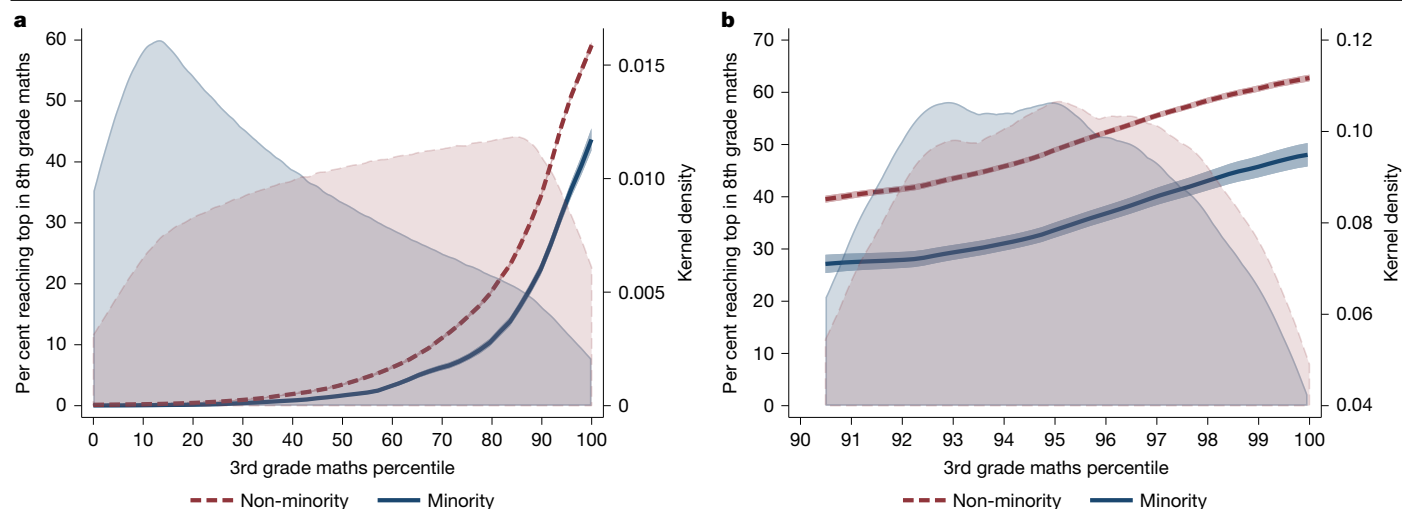


Fig. 4 | Eighth grade excellence by third grade rank. Proportion of students who reach the top decile of eighth grade achievement according to their third grade achievement (lines, left axis) and the distribution of third grade achievement (histograms, right axis) for all third grade students (**a**) and for students who reached the top 10% in third grade (**b**). Excellence is defined as reaching the top decile on the maths statewide standardized exam, MEAP for

Null effects on the excellence gap may mask co-movements in subgroups' likelihoods of reaching excellence. However, we find little evidence to support the notion that the policies improved rates of excellence for either subgroup (Supplementary Fig. 15). We find suggestive evidence that the Priority designation worsened non-minority students' likelihood of achieving excellence in maths by 0.33 percentage points, although this is only significant at the 10% level (two-tailed t -test, robust 95% confidence interval: $[-4.20, 0.32]$; robust $P = 0.092$; $n = 6,858$). Point estimates suggest a larger reduction of 0.77 percentage points for the Focus designation on non-minority students, but it is not statistically significant (two-tailed t -test, robust 95% confidence interval: $[-2.93, 1.48]$; robust $P = 0.519$; $n = 6,724$). Overall, this suggests that accountability policies have at best small effects on the excellence gap or educational triage. Note that this is not a construct of how excellence is defined: we define excellence at the state level, whereas the regression discontinuity compares schools just above or below the designation cutoff.

Together with the evidence from past work, these results suggest that changes in school environment can influence students' academic trajectories, but policies must be properly designed to reduce disparities in top achievement.

Mobility in and out of excellence differs by race

Our prior facts focused on excellence gaps from third to eighth grade. Taking advantage of our longitudinal data, we now consider individual students' trajectories to understand to what extent high-achieving third grade students persist in the top decile and to what extent lower achieving students move up into excellence. To examine mobility, we continue with an exploration of data from grades 3 to 8, then we extend the analysis to later years.

Grades 3 to 8

Figure 4 plots the likelihood that an individual student in each third grade decile is in the top decile in eighth grade, as well as the distributions of third grade achievement by race (see also Supplementary Fig. 19). Of note, at each third grade achievement level, minority students are less likely to be in the top decile by the time they reach the eighth grade compared with non-minority students with identical

all years until 2014 and M-STEP for all years after 2014. Shaded areas around the line graphs correspond to 95% confidence intervals around the mean. Shaded areas reflect the density of students at each achievement level. The analysis uses our main sample, as defined in Methods, which consists of 534,122 students in total: 136,137 classified as racial minority students and 397,985 classified as non-minority students.

third grade test performance (joint F -test of significant differences across each decile of third grade performance; $F = 43.72$; $P < 0.001$; $n = 534,122$). Among those who reach the top decile of achievement in third grade, nearly half of non-minorities stay in the top decile in eighth grade, whereas only around 35% of minority students remain in the top decile. In other words, minority students who reach the top decile in third grade are about 15 percentage points more likely to fall out of the top decile than their top-achieving non-minority counterparts (two-tailed t -test, t -statistic = -13.63 ; $P < 0.001$; $n = 52,988$). In a similar vein, we find large racial differences in whether a student who did not reach excellence in third grade is able to achieve it in eighth grade. Whereas 7% of non-minority students who did not reach the top decile in third grade ended up reaching it in eighth grade, only around 2% of minority students accomplished this feat (two-tailed t -test, t -statistic = -24.14 ; $P < 0.001$; $n = 481,134$).

Part of these differences in persistence comes from the fact that even among students in the top decile, minority students have lower average test scores—that is, they are more likely to be in the 90th–94th percentiles, whereas non-minority students are more likely to be in the 96th–99th percentiles. Notably, at every percentile of third grade achievement from 90th to 99th, non-minority students are also much more likely to remain in the top decile than their minority peers with identical test scores by about around 14 percentage points (joint F -test of significant differences across each percentile from 90th to 100th of third grade performance; $F = 19.95$; $P < 0.001$; $n = 52,988$; Fig. 4b). Therefore, differences in the distribution of third grade achievement by race alone do not fully explain differential persistence. We find the same patterns with reading (Supplementary Fig. 16).

When discussing these results, especially our results on persistence conditioned on a student's third grade score, it is important to acknowledge that test scores on any given day are noisy measures of a student's latent skill, and there are concerns that the noise is especially large in younger grades, which may affect our ability to study these dynamics³⁷. We describe and explore these concerns further in Supplementary Information, section G.1.

These patterns in the excellence gap are qualitatively the same across females and males (Supplementary Fig. 11 and Supplementary Table 6). We explore this further in Supplementary Information, section E. When

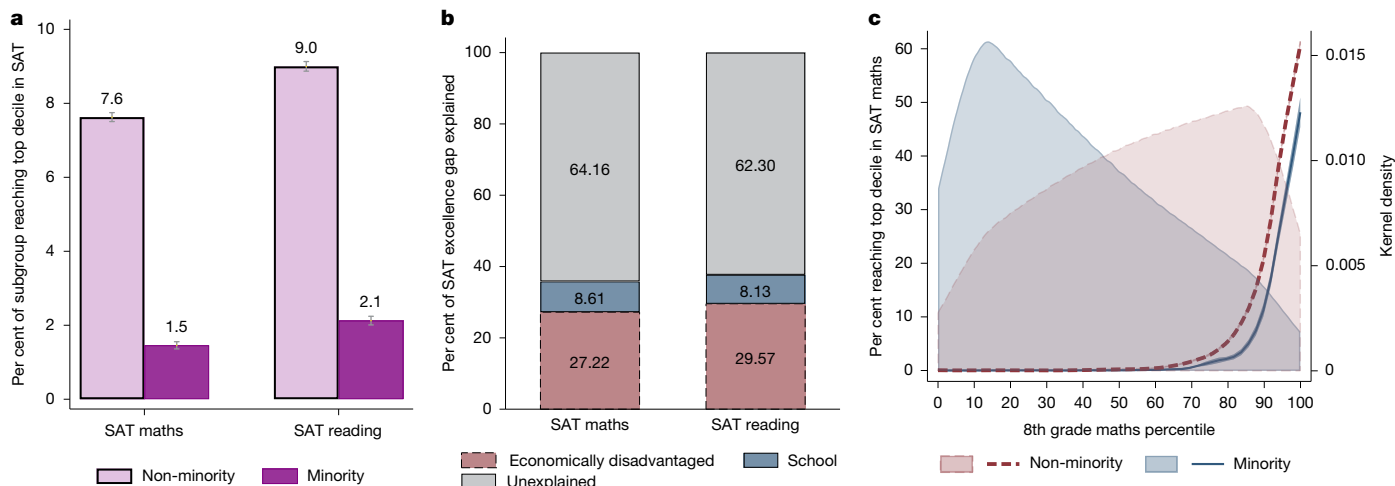


Fig. 5 | Excellence in college readiness. a, Racial differences in reaching the top decile on the SAT Math and Evidence-Based Reading and Writing tests. Data are mean $\pm$ 95% confidence interval. **b**, Gelbach decomposition of the minority and non-minority difference in reaching the top decile on the SAT Math and Evidence-Based Reading and Writing tests. **c**, Proportion of students who reach the top decile on the maths SAT (lines, left axis) across the eighth grade maths test score distribution (histograms, right axis). Shaded areas around the line

considering finer race categories, the largest differences are between Asian and Black students (Supplementary Fig. 18).

Persistence into college readiness

We next examine racial differences in several measures of college readiness: taking at least one Advanced Placement (AP) course, taking at least one AP course in a STEM field, or scoring in the top decile on the maths or reading section of the SAT. We find considerable racial differences in these outcomes (Fig. 5a for SAT and Extended Data Fig. 3 for AP). Minority students are substantially less likely to take an AP course, take a STEM AP course, and score in the top decile in the SAT for both maths and reading than their non-minority counterparts (two-tailed t -tests, t -statistic = -20.09 , -22.32 , -17.76 , -24.53 ; all $P < 0.001$; $n = 244,750$). Although a slightly larger percentage of non-minorities take the SAT than minorities (98.6% versus 95.8%, respectively; two-tailed t -test, t -statistic = -16.32 ; P value on racial difference < 0.001 ; $n = 244,750$), both groups have a very high rate of taking the exam because the state requires that high school students take it³⁴.

We find that racial differences in student socioeconomic status explain 30–40% of racial disparities in college readiness (Fig. 5b for SAT and Extended Data Fig. 3b for AP) (for taking AP, taking STEM AP, top in SAT maths and top in SAT reading: -7.07 , -5.35 , -1.68 and -2.03 percentage points, respectively; two-tailed t -tests, t -statistic = -28.84 , -25.05 , -19.05 , -23.11 ; all $P < 0.001$; $n = 244,750$). Although school fixed effects are also significant predictors of the gaps, they generally explain only around 10% of the racial disparities in college readiness (for taking AP, taking STEM AP, top in SAT maths and top in SAT reading: -2.28 , -1.44 , -0.531 and -0.56 percentage points, respectively; two-tailed t -test, t -statistic = -2.74 , -2.44 , -2.22 , -2.62 ; $P < 0.001$, $P = 0.015$, $P = 0.027$, $P = 0.009$; $n = 244,750$). In sum, student socioeconomic status explains much more of the excellence gap variation than the school environment.

These racial disparities in college readiness persist across the eighth grade achievement distribution (Fig. 5c, Extended Data Fig. 4 and Supplementary Fig. 23). Even among students who reached the top decile in eighth grade, 44% of non-minority students reach the top decile in the maths SAT, whereas only 30% of minority students make that level of excellence (two-tailed t -test, t -statistic = -9.88 ; $P < 0.001$; $n = 27,086$). Among eighth graders who did not reach excellence, non-minority

students are more than 6 times more likely to achieve top performance on the SAT than minority students (2.05 and 0.34 percentage points for non-minority and minority, respectively; two-tailed t -test, t -statistic = -18.85 ; $P < 0.001$; $n = 217,664$). These patterns also extend to reading (Supplementary Fig. 17).

Accounting for excellence gaps during grades three to eight explains some of the college readiness gaps (Supplementary Fig. 21). With regards to AP course taking, racial excellence gaps in eighth grade explain around 8–13% of the total gap, and accounting for third to seventh grades explains around 20%. The excellence gap explains a substantially higher fraction of the racial disparity in SAT scores, with eighth grade excellence gaps explaining around 24–30% of the total gap, and excellence gaps in previous years explaining around 50% of the gap.

Accounting for excellence gaps during grades three to eight explains some of the college readiness gaps (Supplementary Fig. 21). With regards to AP course taking, racial excellence gaps in eighth grade explain around 8–13% of the total gap, and accounting for third to seventh grades explains around 20%. The excellence gap explains a substantially higher fraction of the racial disparity in SAT scores, with eighth grade excellence gaps explaining around 24–30% of the total gap, and excellence gaps in previous years explaining around 50% of the gap.

Socioeconomic status and schools help explain gaps in mobility

Finally, we investigate the extent to which socioeconomic status and school factors and quality explain mobility in and out of the deep right tail. We focus first on transitions from third to eighth grade. By decomposing the eighth grade excellence gap among those who reached excellence in third grade maths, we find that variation in school quality explains more than 30% (4.93 percentage points; two-tailed t -test; t -statistic = -5.65 ; $P < 0.001$; $n = 52,988$) of the racial difference in persistence in excellence in maths from third to eighth grade (Fig. 6a). This is in contrast to the racial difference in entry into eighth grade maths excellence: among students who were not in the top decile for third grade maths, school quality explains only around 12% (0.65 percentage points; two-tailed t -test, t -statistic = -3.46 ; $P = 0.001$; $n = 481,134$) of the racial difference in reaching the top decile by eighth grade. These patterns are qualitatively similar for reading (Fig. 6a). In Supplementary Fig. 20, we separate out the sample of students who did not reach excellence in third grade and find that these patterns persist in these finer bins.

Notably, these patterns persist into college readiness. In this case, school quality explains a substantial portion of the racial differences in all four of our measures of college readiness among students who were in the top decile of eighth grade maths (for example, around

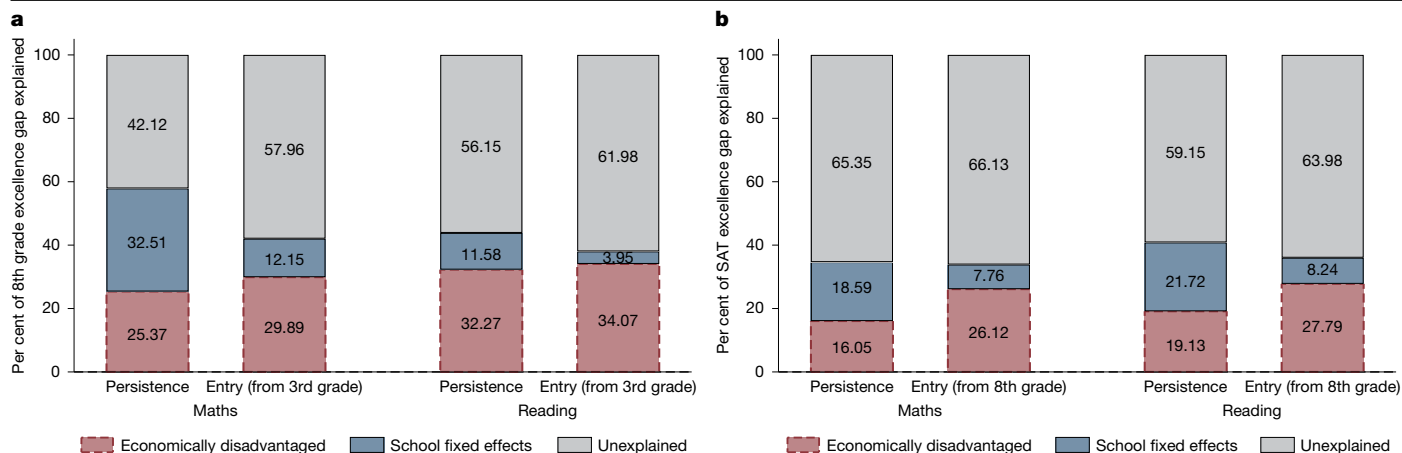


Fig. 6 | Explaining racial differences in persistence and entry of excellence. Gelbach decomposition of the minority and non-minority difference in reaching excellence in eighth grade maths or reading, conditional on having reached it or not in the corresponding subject during third grade (**a**) and reaching the top decile in SAT Math (left two bars) or Evidence-Based Reading and Writing (right two bars) tests, conditional on having reached excellence or not in the corresponding subject during eighth grade (**b**). Persistence refers to whether a student who had reached excellence in that subject in the specified

previous year also reaches it in the subsequent outcome of interest; entry refers to whether a student who had not reached excellence in the specified prior year reaches it in the subsequent outcome of interest. Analysis in **a** uses our main sample, as defined in Methods, which consists of 534,122 students in total: 136,137 classified as racial minority students and 397,985 classified as non-minority students. Analysis in **b** uses our high school sample, as defined in Methods, which consists of 244,750 students in total: 58,370 classified as racial minority students and 186,380 classified as non-minority students.

20% for reaching the top decile in SAT maths or reading) (for taking AP, taking STEM AP, top in SAT maths and top in SAT reading: -3.00 , -2.81 , -2.56 and -3.13 percentage points, respectively; two-tailed t -test, t -statistic = -2.94 , -2.58 , -2.63 , -3.66 ; $P = 0.003$, 0.010 , 0.009 , $P < 0.001$; $n = 27,086$). By contrast, among students who were not in the top decile of eighth grade maths, schools explain a small fraction of the racial differences in college readiness and for two of the outcomes we fail to reject zero at conventional levels of significance (for taking AP, taking STEM AP, top in SAT maths and top in SAT reading: -1.26 , -0.62 , -0.13 , -0.18 percentage points, respectively; two-tailed t -test, t -statistic = -1.55 , -1.19 , -1.86 , -1.76 ; $P = 0.121$, 0.234 , 0.064 , 0.079 ; $n = 217,664$) (Fig. 6b for SAT, Extended Data Fig. 5 for AP, Supplementary Fig. 22 by top in reading and Supplementary Fig. 24 across eighth grade achievement). In sum, these results match those from above and provide support for the fact that school quality explains a substantially larger portion of the racial difference in the persistence of excellence rather than entry into the top tier.

Conclusion

In modern economies characterized by increasing returns to human capital and ‘winner-take-all’ markets, small departures in the right tail of the achievement distribution can lead to considerable surplus creation. A natural corollary is that large welfare losses to an economy can result because of deterioration of talent of the population in the right tail. For the USA, which plans to compete in an increasingly global marketplace in which high-level maths, science and language skills are crucial⁴⁴, the lack of attention to high-achieving students should be of higher concern. This attribute highlights a major oversight for an education policy that intends to achieve educational equity by focusing on other parts of the achievement distribution. Although studying proficiency gaps and excellence gaps are both important, excellence gaps are equally concerning because both gaps are growing rapidly in size, and the groups that are failing to reach high levels of achievement are the fastest growing segments of the US population⁵.

Importantly, we argue that policies that aim to reduce excellence gaps should have more prominence than they do currently, alongside reforms targeting proficiency, as society needs both. The needs

of every student must be met to fully achieve equity, and in turn, to bring greater efficiency in policy to reach longer-term economic goals.

We contribute to this effort by documenting the excellence gap from third grade to college readiness and exploring its key drivers. In doing so, our goal is to strengthen the current documentation on the excellence gap by measuring statewide gaps across cohorts over time. The existence and persistence of the excellence gap throughout a student’s life demonstrate the need for policymakers and scholars to focus their efforts on efficiency through education policy. Furthermore, it is crucial to further explore mechanisms of bridging the excellence gap and what interventions may reduce it²². Our study suggests, perhaps indirectly, that rather than using state policies, school-level policies, such as gifted education or extracurricular programmes (for example, in refs. 45–47), may be effective.

Overall, we view this work as an important step towards systematically documenting the excellence gap. We do so using administrative data from one state, Michigan. The policies and systems in place in Michigan may have specific implications for the excellence gap that are not true everywhere. On the basis of past work (for example, in ref. 22), we may expect these patterns to be broadly representative in the USA. The replication of these analyses across states and countries will reinforce our understanding and better inform the key moderators. We encourage researchers to further document the excellence gap across contexts.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-08536-0>.

1. Hanushek, E. A., Peterson, P. E., Talpey, L. M. & Woessmann, L. The achievement gap fails to close. *Educ. Next* **19**, 8–17 (2019).
2. Gillborn, D. & Youdell, D. *Rationing Education: Policy, Practice, Reform, and Equity* (McGraw-Hill Education, 1999).
3. Neal, D. & Schanzenbach, D. W. Left behind by design: proficiency counts and test-based accountability. *Rev. Econ. Stat.* **92**, 263–283 (2010).

4. Plucker, J. A., Burroughs, N. & Song, R. *Mind the (Other) Gap! The Growing Excellence Gap in K-12 Education* (Center for Evaluation and Education Policy, Indiana Univ., 2010).
5. Rambo-Hernandez, K. E., Peters, S. J. & Plucker, J. A. Quantifying and exploring elementary school excellence gaps across schools and time. *J. Adv. Acad.* **30**, 383–415 (2019).
6. Hemelt, S. W. & Jacob, B. A. How does an accountability program that targets achievement gaps affect student performance? *Educ. Finance Policy* **15**, 45–74 (2020).
7. Kober, N. *It Takes More than Testing: Closing the Achievement Gap. A Report of the Center on Education Policy* (Center on Education Policy, 2001).
8. Williams, B. *Closing the Achievement Gap: A Vision for Changing Beliefs and Practices* (ASCD, 1996).
9. Ladson-Billings, G. From the achievement gap to the education debt: understanding achievement in us schools. *Educ. Res.* **35**, 3–12 (2006).
10. Magnuson, K. A. & Waldfogel, J. Early childhood care and education: effects on ethnic and racial gaps in school readiness. *Future Child.* **15**, 169–196 (2005).
11. Darling-Hammond, L. Race, inequality and educational accountability: the irony of ‘no child left behind’. *Race Ethn. Educ.* **10**, 245–260 (2007).
12. Gamoran, A. *Standards-Based Reform and the Poverty Gap: Lessons for “No Child Left Behind”* (Brookings Institution Press, 2008).
13. Murnane, R. J. Improving the education of children living in poverty. *Future Child.* **17**, 161–182 (2007).
14. Paige, R. No Child Left Behind: The ongoing movement for public education reform. *Harvard Educ. Rev.* **76**, 461–473 (2006).
15. Booher-Jennings, J. Below the bubble: “educational triage” and the Texas accountability system. *Am. Educ. Res. J.* **42**, 231–268 (2005).
16. Wyner, J. S., Bridgeland, J. M. & Dilulio Jr, J. J. *Achievementtrap: How America is Failing Millions of High-achieving Students from Lower-Income Families* (Civic Enterprises, 2007).
17. Hoxby, C. M. & Avery, C. *The Missing “One-offs”: The Hidden Supply of High-Achieving, Low Income Students* (National Bureau of Economic Research, 2013).
18. Meinck, S. & Brese, F. Trends in gender gaps: using 20 years of evidence from TIMSS. *Large Scale Assess. Educ.* **7**, 8 (2019).
19. Plucker, J., Hardesty, J. & Burroughs, N. *Talent on the Sidelines: Excellence Gaps and America's Persistent Talent Underclass* (Center for Education Policy Analysis, Univ. of Connecticut, 2013).
20. Guryan, J., Hurst, E. & Kearney, M. Parental education and parental time with children. *J. Econ. Perspect.* **22**, 23–46 (2008).
21. Hanushek, E. A. & Rivkin, S. G. Harming the best: how schools affect the black-white achievement gap. *J. Pol. Anal. Manag.* **28**, 366–393 (2009).
22. Meyer, M. S., Shen, Y. & Plucker, J. A. Reducing excellence gaps: a systematic review of research on equity in advanced education. *Rev. Educ. Res.* **94**, 33–72 (2024).
23. Plucker, J. A., Peters, S. J. & Schmalensee, S. Reducing excellence gaps: a research-based model. *Gifted Child Today* **40**, 245–250 (2017).
24. Plucker, J. A. & Peters, S. J. *Excellence Gaps in Education: Expanding Opportunities for Talented Students* (Harvard Education Press, 2020).
25. Kim, J., Watkins, S. & Yoon, S. W. Social justice alert: future innovators and leaders fall behind in academic achievement. *Educ. Leader. Rev.* **17**, 129–141 (2016).
26. Hamilton, R. et al. Disentangling the roles of institutional and individual poverty in the identification of gifted students. *Gifted Child Q.* **62**, 6–24 (2018).
27. Kaul, C. R. & Davis, B. K. How the state education agencies addressed gifted education in the title II sections of their ESSA state plans. *Gifted Child Today* **41**, 159–167 (2018).
28. Hodges, J. & Lamb, K. Washington's high-ability programs during the no child left behind era. *J. Educ. Gifted* **42**, 283–302 (2019).
29. Peters, S. J., Gentry, M., Whiting, G. W. & McBee, M. T. Who gets served in gifted education? demographic representation and a call for action. *Gifted Child Q.* **63**, 273–287 (2019).
30. Turner, E. O. & Spain, A. K. The multiple meanings of (in)equity: remaking school district tracking policy in an era of budget cuts and accountability. *Urban Educ.* **55**, 783–812 (2020).
31. Mun, R. U., Ezzani, M. D., Lee, L. E. & Ottwein, J. K. Building systemic capacity to improve identification and services in gifted education: a case study of one district. *Gifted Child Q.* **65**, 132–152 (2021).
32. Lockhart, K., Meyer, M. S. & Crutchfield, K. A content analysis of selected state plans for gifted and talented education. *J. Adv. Acad.* **33**, 3–42 (2022).
33. List, J. A. *Non est Disputandum de Generalizability? A Glimpse Into the External Validity Trial* (National Bureau of Economic Research, 2020).
34. Hyman, J. Act for all: the effect of mandatory college entrance exams on postsecondary attainment and choice. *Educ. Fin. Pol.* **12**, 281–311 (2017).
35. Fryer, R. G. & Levitt, S. D. Testing for racial differences in the mental ability of young children. *Am. Econ. Rev.* **103**, 981–1005 (2013).
36. List, J. A., Pernaudet, J. & Suskind, D. L. Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. *Nat. Commun.* **12**, 1367–1378 (2021).
37. Bond, T. N. & Lang, K. The black-white education scaled test-score gap in grades K-7. *J. Hum. Resour.* **53**, 891–917 (2018).
38. Gelbach, J. B. When do covariates matter? and which ones, and how much? *J. Labor Econ.* **34**, 509–543 (2016).
39. Autor, D., Figlio, D., Karbownik, K., Roth, J. & Wasserman, M. Males at the tails: how socioeconomic status shapes the gender gap. *Econ. J.* **133**, 3136–3152 (2023).
40. Jackson, C. K. & Mackevicius, C. L. What impacts can we expect from school spending policy? evidence from evaluations in the United States. *Am. Econ. J.* **16**, 412–446 (2024).
41. Reardon, S. F., Kalogrides, D. & Shores, K. The geography of racial/ethnic test score gaps. *Am. J. Sociol.* **124**, 1164–1221 (2019).
42. Hemelt, S. W. & Jacob, B. *Differentiated Accountability and Education Production: Evidence from NCLB Waivers* (National Bureau of Economic Research, 2017).
43. Cattaneo, M. D., Jansson, M. & Ma, X. Simple local polynomial density estimators. *J. Am. Stat. Assoc.* **115**, 1449–1455 (2020).
44. Hess, F. M. & Rotherham, A. J. Can NCLB survive the competitiveness competition? *Sch. Bus. Aff.* **74**, 8 (2008).
45. Card, D. & Giuliano, L. Universal screening increases the representation of low-income and minority students in gifted education. *Proc. Natl Acad. Sci. USA* **113**, 13678–13683 (2016).
46. Assouline, S. G., Ihrig, L. M. & Mahatmya, D. Closing the excellence gap: investigation of an expanded talent search model for student selection into an extracurricular stem program in rural middle schools. *Gifted Child Q.* **61**, 250–261 (2017).
47. Ihrig, L. M., Lane, E., Mahatmya, D. & Assouline, S. G. Stem excellence and leadership program: increasing the level of stem challenge and engagement for high-achieving students in economically disadvantaged rural communities. *J. Educ. Gifted* **41**, 24–42 (2018).
48. Calonico, S., Cattaneo, M. D., Farrell, M. H. & Titiunik, R. rdrobust: Software for regression-discontinuity designs. *Stata J.* **17**, 372–404 (2017).
49. Calonico, S., Cattaneo, M. D. & Titiunik, R. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82**, 2295–2326 (2014).
50. Hartman, E. Equivalence testing for regression discontinuity designs. *Polit. Anal.* **29**, 505–521 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Methods

Ethics approval

This research is approved by the University of Chicago Social and Behavioral Sciences Institutional Review Board.

Data

We use statewide administrative data from the Michigan Education Research Institute (MERI). The data include all public school students in the state of Michigan. The broad educational policy context in Michigan also matters for the external validity of our results. Based on the State Educational Practices (SEP) report by the National Center for Education Statistics, Michigan policies are not very different from the national average. For example, Michigan does not have a voucher programme (only 15 of the 50 US states do), teachers are required to have passed a basic skills exam and subject matter exams, and have an assessment of teaching performance (39, 46 and 26 states do, respectively), and administrators are required to have field experience and completed a preparation programme that aligns with state standards (38 and 48 states do, respectively). However, Michigan requires relatively few credits for high school graduation, and uses the SAT for high school maths and English assessments rather than relying on state tests (further information is available at <https://nces.ed.gov/programs/staterreform/index.asp>).

We focus on student-level demographics, standardized measures of achievement from grades three to eight, and outcomes during high school, including AP course taking and SAT scores. These data are unique in that they allow us to track the entire population of students in Michigan across multiple cohorts over time and to follow these students longitudinally to examine high school outcomes. Thus, we examine statewide patterns to compute a measure of the excellence gap at the grade-by-cohort level. This allows us a glimpse into the persistence of the gap through high school and differences in college readiness. The data also include various student-level demographics, such as gender, race and whether a student is economically disadvantaged, which is defined by whether the student is determined to be eligible for free or reduced-price meals, is in a household receiving food or cash assistance, is homeless, are migrant, or is in foster care (<https://medc.miedresearch.org/dataset/k-12-student-demographic-and-enrollment-data>).

We observe whether a student is determined to be economically disadvantaged starting from 2012. In other words, the earliest grade that we observe economically disadvantaged status for all cohorts in our data is the eighth grade. We therefore use an indicator for whether the student is economically disadvantaged in eighth grade as our main measure of socioeconomic status. In the Supplementary Information, we show additional robustness checks for measuring socioeconomic status as the number of times a student is determined as economically disadvantaged, conditional on the total number of times this information is available for the student (Supplementary Fig. 5).

For the cohorts of students who began third grade between 2007 and 2012, the data contain achievement assessment scores from third grade to eighth grade. For the cohorts of students who began third grade between 2007 and 2009, the data further link to information on high school outcomes. Given this data structure, we construct two analysis samples. First, we create a ‘main sample’, consisting of students who began third grade within the years 2007 and 2012, and progressed through grades on time. We subset to students for whom we observe test scores for each grade from grades three to eight to be able to follow individual students over time.

Second, we create a ‘high school sample’, which subsets from the main sample to students who began third grade between 2007 and 2009, and who are observed reaching the 11th grade in Michigan. In other words, students who move out of Michigan during the data time frame are excluded from our analyses. (We include students who switch to home schooling). Based on the structure of the data, we acknowledge two

important limitations. First, the data initially captures students in third grade and therefore does not allow us to measure the extent disparities arise earlier (as in ref. 35). Second, our analysis excludes private school students, who make up 7% of enrolments in Michigan⁵¹. (However, based on the publicly available national data from the National Assessment of Educational Progress (NAEP), we calculate relatively small differences, within race, across public and private schools, in achieving high test scores, especially for white students who are over-represented in private schools. Based on these calculations, using only public school students may lead us to slightly overstate the magnitude of the excellence gap). Supplementary Table 1 presents summary statistics.

Outcome variable

As our measure of academic achievement, we use assessment scores from the MEAP for all years until 2014, and the M-STEP for younger cohorts we observe after 2014. The MEAP was the standardized test administered to all elementary (5–11 years of age) and middle school (11–14 years of age) students in Michigan until the 2013–2014 school year. It tested mathematics, reading, science, writing and social studies. After 2014, the MEAP was replaced by the M-STEP, a standardized test administered to students in grades three to eight, measuring performance on state standards in English language arts, maths, science and social studies. Going forward, for reading test scores, we use scores from MEAP reading when available, and otherwise M-STEP English language arts.

We define academic excellence as reaching the top 10% in the statewide standardized test, within a grade-year. Within each grade-year, we first rank all students separately by their maths and reading (or other high school) exam scores. For ties where students have the same score, we use the average rank (based on Stata’s `egen rank` command). Supplementary Table 2 shows the per cent ‘excellence’ after breaking ties. We show results for using a 5% and 1% cutoff in the Supplementary Figs. 25–27. Supplementary Table 2 shows the fraction of students categorized as excellent in each grade and subject.

For gaps in college readiness, we observe students’ performance on the SAT maths and reading, enrolment in an AP course, and enrolment in an AP course in a STEM field. STEM AP courses are any AP course in the data with a subject area classified as mathematics, life and physical sciences, computer and information sciences, and engineering and technology. Specifically, this consists of AP Calculus AB, AP Calculus BC, AP Statistics, AP Biology, AP Chemistry, AP Physics B, AP Physics C, AP Environmental Science, AP Computer Science A and AP Computer Science AB.

Following Remark 2.1 in ref. 52, we define the family of null hypotheses of interest in a manner that does not require multiple testing corrections.

Defining racial gaps

Following past work on the excellence gap^{4,5}, we define the racial excellence gap as the difference across minority (Black or Hispanic) and non-minority students’ rates of reaching excellence. The racial gap is therefore given by the following expression:

$$\text{Excellence gap} = P(\text{Top decile}_i | \text{Minority}_i) - P(\text{Top decile}_i | \text{Non-minority}_i) \quad (1)$$

We use terms ‘minority’ and ‘non-minority’, respectively, to capture disadvantaged students on the basis of race and educational attainment but acknowledge that Asian students are, in fact, a racial minority.

Measuring racial excellence gaps

Our first analyses compare differences in academic achievements by student demographics (with the focus on race), using two-tailed *t*-tests to compare the likelihood of reaching excellence for minority versus non-minority students. Our main analyses use ordinary least squares

for multivariate comparisons, including testing whether subgroup differences change with students' grade level.

Measuring drivers of excellence gaps

To understand the role of various covariates in explaining racial differences in achievement, we conduct Gelbach decompositions³⁸. Conceptually, this approach treats each factor as an 'omitted variable' in the relationship between excellence and race and measures the bias that would result if the factor were excluded. This allows us to disentangle the impact on the excellence gap of each factor. This approach is of particular value when our observables are correlated; for example, school fixed effects are likely endogenous to our measure of economically disadvantaged such that the difference in the point estimates of the excellence gap with and without controlling for the school fixed effects is likely also capturing differences in economic status. We therefore estimate the unconditional size of the excellence gap and compare the magnitude to estimates that additionally control socioeconomic status and school fixed effects.

To implement this method, we use the `b1x2` command in Stata. In practice, it estimates a series of the following equation with and without the additional control for socioeconomic status and school fixed effects. Specifically, for student i of cohort $c(i)$ in grade g at school s and of socioeconomic status $e(i)$:

$$Y_{i,g} = \alpha + \beta_1 M_i + \gamma_{g,c(i)} + \lambda_{e(i),s} + \epsilon_{i,g} \quad (2)$$

where $Y_{i,g}$ is an indicator for whether i reaches the top decile in grade g , M_i is an indicator for whether the student is a minority, and $\lambda_{e(i),s}$ captures school fixed effects and controls for student socioeconomic status interacted with grade and cohort. β_1 is our main coefficient of interest, which measures the percentage point difference between the likelihood minority and non-minority reach the highest levels of achievement, that is, the excellence gap.

Measuring school-level excellence gaps

To compare school-level heterogeneity in racial excellence gaps, we calculate the excellence gaps in maths and reading for each school. We construct this measure by calculating excellence gaps for each school year observation in the data, meaning for each school s and year y :

$$\hat{\theta}_{s,y} = \frac{1}{|N_{s,y}|} \sum_{j \in N_{s,y}} I\{\text{Excellence}\}_i - \frac{1}{|M_{s,y}|} \sum_{j \in M_{s,y}} I\{\text{Excellence}\}_i \quad (3)$$

where $M_{s,y}$ is the set of all minority students in school s in year y and $N_{s,y}$ is the set of all non-minority students in school s in year y . We omit school years associated with fewer than ten minority or ten non-minority test takers. We then aggregate these estimates to the school level:

$$\hat{\theta}_s = \frac{1}{|R_s|} \sum_{y \in R_s} \hat{\theta}_{s,y} \quad (4)$$

where R_s is the set of years that school s is observed in our data. We estimate school-specific excellence gaps, $\hat{\theta}_s$, for maths and reading. Figure 2 shows the distribution of these gaps. We first test for heterogeneity across schools using a Pearson chi-squared test: our test statistic is $\sum_s \frac{(\hat{\theta}_s - \bar{\theta})^2}{e_s^2}$, where $\bar{\theta} = \frac{1}{N} \sum_s \hat{\theta}_s$ with N as the total number of schools in our data, and e_s is the standard error associated with our school-specific estimate of the excellence gap, $\hat{\theta}_s$ presented in the first row of Supplementary Table 5. This test is under the assumption that the distribution of each school's excellence gap is normally distributed, centred around the across-school population mean with variance e_s^2 .

Our school-specific measures of observed excellence gaps face minimal sampling error for the corresponding school year, given our use of comprehensive administrative data. In Supplementary Information, section D, we follow the steps of ref. 53 to estimate the distribution of

latent school-level excellence gaps, which treats yearly school-level gaps as noisy estimates of a school's latent gap.

The effect of accountability policies

To understand the impact of past policies on the racial excellence gap, we use a regression discontinuity that leverages the discrete threshold that was used to designate policy actions towards schools in Michigan starting in 2012.

Michigan schools were identified as Priority or Focus schools using a top-to-bottom ranking. A school was designated as a Priority school if it was in the bottom 5% of a school performance index that the state created using measures on achievement gaps (25% weight), student achievement (50%), student growth (25%) over time. A school not assigned Priority status was designated as a Focus school if its composite gap index, which was comprised of measures of achievement gaps in 5 subjects (mathematics, reading, science, social studies and writing), fell below the score corresponding to the 10th percentile of Title I schools. The assignment mechanisms for the Focus and Priority designations allow us to estimate the causal effect of this accountability policy using a regression discontinuity design, where the school performance index (for Priority) and composite gap index (for Focus) are the running variables.

We construct the composite gap index cutoff following the steps outlined by the Michigan Department of Education (https://www.michigan.gov/-/media/Project/Websites/mde/Year/2012/08/01/201112_School_Rankings_Business_Rules.pdf?rev=9f4770c935da470aa8dff4ef8f03194e). We get the total number of Title I schools in Michigan (2,775), a value used when computing the composite gap index cutoff, from publicly available data by the National Center for Education Statistics.

In our analyses, we follow the sample restrictions outlined by ref. 6. We exclude schools that serve high school students because the ranking for these schools include graduation rates. Specifically, we remove schools who offer grades higher than 9th grade, and schools whose highest offered grade is below fourth grade. We also exclude schools that closed during the summer during when the policy was announced, and schools that were identified as 'persistently low achieving' or schools that received a School Improvement Grant prior to 2012 because these schools did not face the same designation procedures as other schools⁴². Finally, following ref. 6, we exclude schools who had fewer than 50 test takers in the year prior to the policy to remove schools for special populations. We are ultimately left with a sample of 1,797 unique schools, whose achievement measures are constructed using 1,783,497 unique student-year observations: 393,891 correspond to minority students, and 1,389,606 to non-minority students.

We assess the suitability of the regression discontinuity research design. First, we test that the running variables and their respective cutoffs indeed generate substantial discontinuities in whether a school is deemed as Priority or Focus. Second, we investigate the densities of the running variables and find no strong evidence of discontinuities in density at the cutoffs using a test of manipulation following ref. 43. Third, we conduct numerous placebo tests using predetermined school characteristics as outcomes, which provide support towards our identification assumption that there are no imbalances on unobservables that affect our outcomes of interest.

In the regression discontinuity specification, we pool observations from one and four years following policy designation, following the regression discontinuity design in past work studying this policy^{6,42}. We focus on the effect of the designations from the first year of the policy implementation, which was in the summer of 2012. We cluster standard errors at the school level. Our estimates therefore capture the local average treatment effect of each designation on the excellence gap, and the likelihood of reaching excellence by race, by comparing schools with test scores indexes just below the cutoff to schools just above. For estimation, we implement the robust bias-corrected approach of ref. 49, using the `rdrobust` Stata package for the regression tables and

Article

the `rdplot` Stata command for the regression discontinuity figures. We include the following school-level controls for precision: fraction of economically disadvantaged students, indicator for elementary school, highest grade offered in school, log of total number of test takers in the school, average maths and reading standardized scores, and indicators for urbanicity (suburban and urban). These controls take values from the year prior to policy designation, 2011. Each control is interacted with year indicators. We weight our regressions that estimate the impacts on the excellence gap using the total number of test takers in the school in 2011. For the regressions that estimate impacts on minority (non-minority) students' likelihoods of reaching excellence, we weight observations by the total number of minority (non-minority) test takers in the school in 2011.

We additionally calculate equivalence test confidence intervals following ref. 50 to test for lack of effects. We use the code provided in ref. 54. To validate the regression discontinuity design, we implement tests of manipulation using the local polynomial density estimation of ref. 43. We follow ref. 55 and present point estimates together with bias-corrected confidence intervals. For this reason, the point estimates are not in the centre of the confidence intervals.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

This research used data structured and maintained by the MERI–Michigan Education Data Center (MEDC). MEDC data are modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information and opinions of the authors and

are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. Researchers interested in accessing these data can submit a request to MEDC through <https://medc.miedresearch.org>. There are no constraints to submit a proposal.

Code availability

Code used in this paper will be publicly available at <https://www.openicpsr.org/openicpsr/project/215681/version/V1/view>.

51. White, R. & DeGrow, B. *A Survey of Michigan's Private Education Sector. A Mackinac Center Report* (ERIC, 2016).
52. List, J. A., Shaikh, A. M. & Vayalinkal, A. Multiple testing with covariate adjustment in experimental economics. *J. Appl. Econometrics* **38**, 920–939 (2023).
53. Kline, P., Rose, E. K. & Walters, C. R. Systemic discrimination among large US employers. *Q. J. Econ.* **137**, 1963–2036 (2022).
54. Hartman, E. RDD equivalence test. Github https://github.com/ekhartman/rdd_equivalence/blob/master/RDD_equivalence_functions.R?raw=TRUE (2020).
55. Cattaneo, M. D., Idrobo, N. & Titiunik, R. *A Practical Introduction to Regression Discontinuity Designs: Foundations* (Cambridge Univ. Press, 2019).

Acknowledgements The authors thank D. Neal and P. Peña for helpful comments; S. Hemelt and B. Jacob for sharing their code; G. Barcelo, C. Brobst, J. Cipriano, Z. El-Kaissi, B. Hammer, A. Listo, A. Mryan, F. Pagnotta, L. Ramirez, S. Shi, R. Sharma, C. Smith and J. Wang for research assistance with the literature review.

Author contributions J.A.L., M.S.L. and H.U. conceived the research questions and acquired the data. U.K., J.A.L., A.S. and H.U. formulated the methodological approach, and U.K., A.S. and H.U. analysed the data. U.K., A.S., J.A.L. and H.U. wrote the manuscript. The authors are listed in alphabetical order.

Competing interests The authors declare no competing interests.

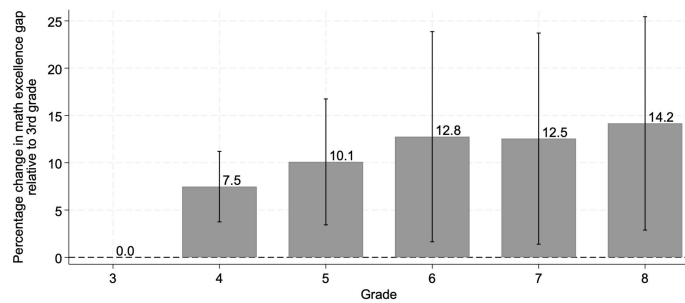
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-08536-0>.

Correspondence and requests for materials should be addressed to John A. List, or Haruka Uchida.

Peer review information Nature thanks Aline Bütikofer, Jonathan Plucker, Ariella Spitzer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

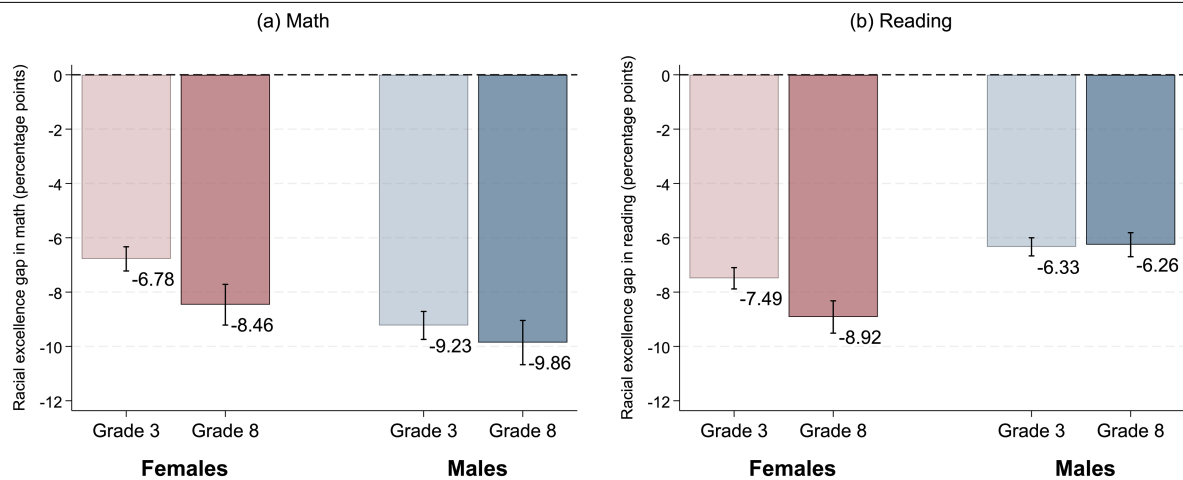
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | Percentage change in excellence gap from 3rd grade.

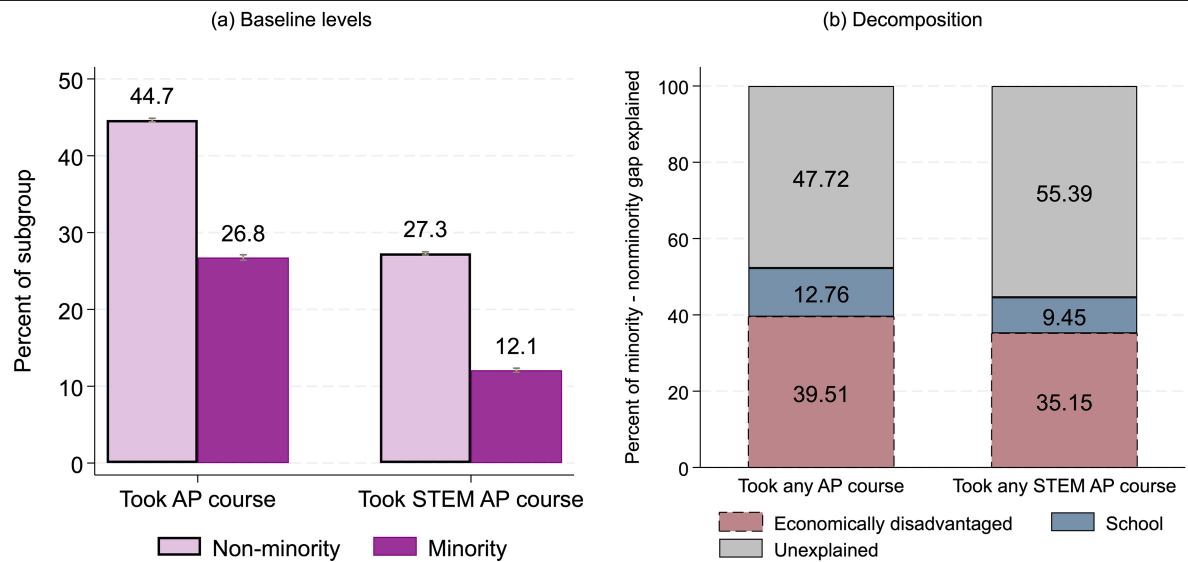
This figure displays the percentage change in this difference between minority status in reaching excellence, relative to the difference in the 3rd grade.

Excellence is defined as reaching the top decile on the math statewide standardized exam, MEAP for all years until 2014 and M-STEP for all years after 2014. Analysis uses our main sample, as defined in Methods, which consists of 534,122 students in total: 136,137 classified as a racial minority and 397,985 as a non-minority. Error bars reflect 95% confidence intervals around the mean.



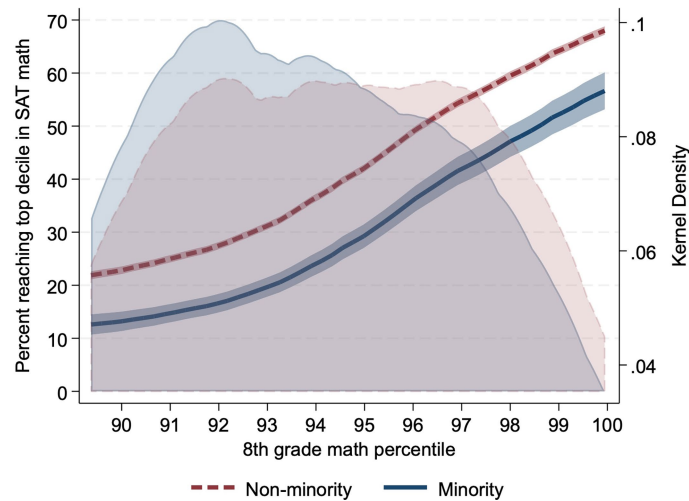
Extended Data Fig. 2 | Excellence gaps by gender. These figures show for female and male students the proportion of each minority group that reaches excellence in (a) math and (b) reading. Analysis uses our main sample, as

defined in Methods, which consists of 534,122 students in total: 136,137 classified as a racial minority and 397,985 as a non-minority. Error bars reflect 95% confidence intervals around the mean.



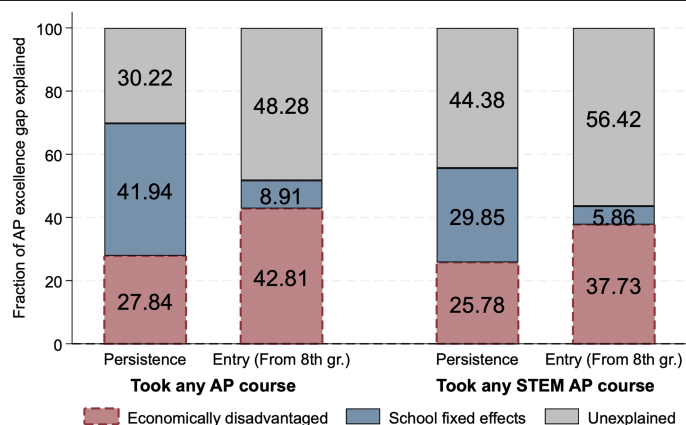
Extended Data Fig. 3 | Gaps in college readiness (AP coursework). This figure displays (a) the percentage of each race/ethnicity group that takes at least one AP course or takes at least one STEM AP Course and (b) the Gelbach decomposition of each minority non-minority difference. “Economically disadvantaged” is an indicator for whether a student is classified as economically disadvantaged by Michigan which is defined by whether the

student is determined to be eligible for free or reduced-price meals, are in households receiving food or cash assistance, are homeless, are migrant, or are in foster care. Analysis uses our high school sample, as defined in Methods, which consists of 244,750 students in total: 58,370 classified as a racial minority and 186,380 as a non-minority. Error bars reflect 95% confidence intervals around the mean.



Extended Data Fig. 4 | Excellence in the SAT math section by 8th grade rank.

This figure shows the proportion of students who score in the top decile of the SAT math section nationally, and the distribution of 8th grade achievement in math for each percentile in the top 10 percent. Shaded lines around the two lines correspond to the 95% confidence intervals and shaded areas reflect the density of students at each achievement level. Analysis uses our high school sample, as defined in Methods, which consists of 244,750 students in total: 58,370 classified as a racial minority and 186,380 as a non-minority.



Extended Data Fig. 5 | Explaining racial differences in persistence and entry in college readiness (AP coursework). This figure displays the Gelbach decomposition of each minority non-minority difference in whether a student takes at least one AP course or takes at least one STEM AP Course, among students who achieved excellence in 8th grade math (for Persistence) and students who did not achieve excellence in 8th grade math (for Entry). “Economically disadvantaged” is an indicator for whether a student is classified as economically disadvantaged by Michigan which is defined by whether the student is determined to be eligible for free or reduced-price meals, are in households receiving food or cash assistance, are homeless, are migrant, or are in foster care. Analysis uses our high school sample, as defined in Methods, which consists of 244,750 students in total: 58,370 classified as a racial minority and 186,380 as a non-minority.

Article

Extended Data Table 1 | Population characteristics

	Michigan	U.S.
Demographics		
% White	72	64
% Black	15	12
% Hispanic	7	16
% Asian	3	5
School Investments		
Pupil-Teacher Ratio	18	16
Per Pupil Expenditure	10,577	10,663
Mean Teacher Salary	57,442	54,965
Household Characteristics		
Mean Household Income	59,772	68,259
% of students eligible for FRPL	46	48
Test Scores: 8th Grade Math		
% Advanced	6	8
-White	6	11
-Black	0	2
-Hispanic	5	3
-Asian/Pacific Islander	31	22
College Readiness		
% of students who have taken the SAT	98*	49
% of students who have taken at least one AP course	40	38
% of students who enrolled in 2-yr or 4-yr college	64	63

*Since 2016-17, Michigan offered the SAT exam to all public school juniors for free of charge. The NCES reports that 100% of graduating seniors in Michigan took the SAT in 2016-17.

This table displays population characteristics of Michigan and the United States as a whole for the year 2010 or academic year 2010-11. Demographics and college readiness (SAT and AP) statistics come from Michigan administrative data used in this paper, students who began 3rd grade in public school within the years of 2007 and 2012 and progressed through grades 3 through 8 on time. Demographic data from the United States Census Bureau 2010 Decennial Census. School investment data, including % of students eligible for FRPL, from the National Center of Education Statistics' Digest of Education Statistics. Mean household income data from the 2010 American Community Survey. Test score data from the National Assessment of Educational Progress (NAEP) administered in 2011. All college readiness measures are for the 2016-2017 academic year. SAT data for the US is from the National Center for Education Statistics (NCES). AP enrollment data for the US is from the College Board, which refers to public school students only. College enrollment data is from the National Information Center for Higher Education Policymaking and Analysis (NCHEMS).

Extended Data Table 2 | Predicting excellence gaps across schools

	Math Excellence Gap						Reading Excellence Gap					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Frac. Minority	0.63 (0.78)					-11.05*** (0.95)	-1.36** (0.61)					-8.32*** (0.78)
Frac. Econ Disad.		10.49*** (0.82)				-2.29** (1.11)	4.32*** (0.58)					-1.82* (1.04)
Average Math Score (std)			-14.78*** (1.18)			-13.57*** (1.10)			-1.22 (0.89)			-0.34 (0.82)
Average Reading Score (std)			7.69*** (1.21)			0.57 (1.31)			-3.34*** (0.99)			-8.80*** (1.14)
N Test-takers (log)				-0.47 (0.38)		0.12 (0.31)				0.33 (0.26)		0.39 (0.26)
Location: Suburb					0.19 (0.44)	0.82*** (0.30)					0.55* (0.29)	0.74*** (0.24)
Location: Rural					1.48*** (0.44)	0.60* (0.31)					0.83*** (0.29)	0.33 (0.26)
N	1975	1975	1975	1975	1975	1975	1975	1975	1975	1975	1975	1975
Adjusted R ²	0.05	0.18	0.42	0.05	0.06	0.51	0.06	0.10	0.17	0.06	0.06	0.28

* $p < .1$, ** $p < .05$, *** $p < .01$

This table presents results of twelve OLS regressions to predict a school's excellence gap with its observable characteristics. Positive values of the excellence gap are in favor of minorities. Observation is at the school-level, weighted by the number of test-takers in that school across all years. Each regression includes indicators for each grade that a school teaches, and year that the school is observed in the data for. Excellence is defined as reaching the top decile on the math or reading statewide standardized exam, MEAP for all years until 2014 and M-STEP for all years after 2014. Analysis uses our main sample, as defined in Methods, which consists of 534,122 students in total: 136,137 classified as a racial minority and 397,985 as a non-minority. Standard errors in parentheses, clustered at the school-level. All tests are two-sided.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	The authors used student longitudinal data from Michigan (MERI). Below we include the statement from MERI about the data collection and access that is in the data availability statement of the manuscript. From their raw data, we cleaned it and performed the analysis using Stata 18 statistical software. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information and opinions of the authors and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. Researchers interested in accessing this data can submit a request to MEDC through medc.miedresearch.org.
Data analysis	The authors used Stata 18 statistical software to clean the raw data from the state of Michigan (MERI) and to perform the analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information and opinions of the authors and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof. Researchers interested in accessing this data can submit a request to MEDC through medc.miedresearch.org. There are no constraints to submit a proposal.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

In the manuscript, we report information about and results by student's gender. The data from the state of Michigan have a variable for gender, but not sex. Gender is reported by the student or family.

Reporting on race, ethnicity, or other socially relevant groupings

The data include information about the student's reported race. About 96% of students report their race to be white, Black, Hispanic, or Asian. Following past work on this topic, we group students by race into underrepresented minority students (Black and Hispanic) or non-minority (white or Asian) for our main analyses. We additionally present results by individual reported racial categories in the supplemental information.

Our research questions specifically pertain to the excellence gap, which is the racial difference between minority and non-minority students in reaching the highest levels of academic achievement. Race therefore plays a central role in our analysis, and we are explicitly interested in race itself---it is not a proxy for another variable.

In addition to race, the data from Michigan includes information about socioeconomic status. WE additionally consider the role of socioeconomic status in our analyses. Since our research question pertains to race, we only control for other variables to understand potential mediators.

Population characteristics

Our sample consists of 3-12th grade students in Michigan public schools. Based on reported race, the sample is 72% white, 15% Black, 7% Hispanic, and 3% Asian. 46% of students in the sample qualify for free or reduced price lunch, a measure of socioeconomic status. 50.3% are female.

Extended Data Table 1 provides more information about our sample of Michigan students and compares it to a National sample of students.

Recruitment

Individuals were not recruited by the researchers. The data are collected by the state of Michigan.

Ethics oversight

This research is approved by the Social and Behavioral Sciences Institutional Review Board (SBS IRB) of the University of Chicago.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Using rich administrative longitudinal data, we document an "excellence gap," and provide eight key facts. The eight facts provide vital temporal insights into the gap, link stability of excellence and student backgrounds, and assess the efficacy of previous public policies to curb the excellence gap. Building such scientific knowledge provides a crucial first step to target and eliminate excellence gaps, a necessary ingredient to broaden the racial pool of innovators and remove the glass ceiling in labor markets.

The data contain student level test scores (quantitative) and demographics.

Research sample	The research sample contains public school students in grades 3-12 in Michigan. The data is collected and maintained as student longitudinal dataset by the state of Michigan. We compare our sample to similar students in the US (Extended Data Table 1). Michigan has a larger share of white and Black students, and fewer Hispanic and Asian students, than the rest of the US. Average household income is also slightly lower. The school inputs in Michigan are roughly similar. The sample is important because Michigan is a large, mostly representative state that collects data on students from 3-12 grade. This allows for a systematic longitudinal study to measure racial gaps and their evolution over students' lives.
Sampling strategy	The data is collected by the state of Michigan for all students in public K-12 schools. We obtained data for all students in cohorts of students for whom we would be able to observe their entire progression through elementary and secondary school, corresponding to cohorts who enter 3rd grade between 08-07 and 08-12. We include all students in all public schools for three reasons. First, one of our contributions is to document the excellence gap with population level administrative data. Second, some of our analyses are at the school or school by cohort level. Finally, the accountability policy we study only applies to a subset of schools.
Data collection	The data is collected by the state of Michigan. The data contains state standardized test scores, and other national assessments (SAT, AP tests). The researchers were not involved in the data collection process.
Timing	We obtain data for students who start 3rd grade between 08-07 and 08-12. The analysis follows these students through high school
Data exclusions	From the state longitudinal student data from Michigan, our sample includes all students who are in the data each year from 3-8th grade. Students who are not in the data each year are excluded. This reduces our sample from 674,851 to 534,122. We then link this sample to their additional high school (9-12th grade) outcomes. The data is limited to Michigan public schools. We do this to limit changes in sample composition over time since we are interested in how gaps evolve as students progress through school.
Non-participation	Our data consists of the population of students in Michigan public schools in grades 3-12.
Randomization	Our analysis contains two relevant groupings. First, in all of our analyses, we group students by reported race. This is central to our research question. Section, in Section II, Fact 6, we additionally group students based on whether their school was subject to additional accountability policies. Assignment to the accountability policies is determined by a cut-off rule based on the school's ranking on a performance index. This setting, therefore, allows us to estimate the causal effect of the accountability policy using a regression discontinuity design.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.