Eight Facts About the Racial Excellence Gap

Uditi Karna* ⓡ Andrew Simon‡ ⓡ Min Sok Lee§ ⓡ John A. List¶
ⓡ Haruka Uchida‖

September 2024

Abstract

Rising educational disparities and concomitant wealth inequities have been observed around the globe over the past several decades. While those left behind have rightfully received considerable attention, much less scrutiny has been given to educational discrepancies in the deep right tail. Using rich administrative longitudinal data, we document an "excellence gap," and provide seven key facts. The seven 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.

*Columbia University; uk2172@columbia.edu
‡University of Virginia; arsimon@virginia.edu
§University of Chicago; mslee@uchicago.edu
¶University of Chicago, Australian National University, NBER, RFF, and IZA; jlist@uchicago.edu
‖University of Chicago; uchida@uchicago.edu.

Author order was randomized using the American Economic Association author randomization tool. We thank Derek Neal and Pablo Peña for helpful comments, and Steven Hemelt and Brian Jacob for sharing their code. We are grateful to Gabrianna Barcelo, Charlie Brobst, Joseph Cipriano, Zayne El-Kaissi, Bettina Hammer, Ariel Listo, Amani Mryan, Francesca Pagnotta, Lina Ramirez, Steven Shi, Reetika Sharma, Clare Smith, and Jeffrey Wang for excellent research assistance with the literature review. This research result 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.

Introduction

A key contributor to rising social and wealth inequities in the past several generations relates to educational disparities. This fact has motivated modern societies to focus on closing opportunity gaps in education. In the U.S., for example, attempts to close racial achievement gaps have typically come in the form of advancing "one-size-fits-all" accountability policies, such as No Child Left Behind (NCLB) (Hanushek et al., 2019; Kober, 2001; Williams, 1996; Ladson-Billings, 2006; Magnuson and Waldfogel, 2005; Darling-Hammond, 2007; Gamoran, 2008; Murnane, 2007). 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 most likely to improve aggregate scores (Paige, 2006; Gillborn and Youdell, 1999; Booher-Jennings, 2005).

Such triaging potentially creates a nefarious externality across demographic groups when applied across school districts, serving to increase racial gaps in the right tail of the distribution. Such racial excellence gaps—the difference in the percentage of students across racial groups that reach the top levels of overall performance—can have broad deleterious social effects. For the purposes of this study, we compare underrepresented racial groups in the highest level of education and STEM fields (Black, Hispanic, and other students, hereafter, "minority") to "non-minority" (White or Asian) students.[1] For instance, there will be a dearth of minority representation in scarce, high-level jobs, such as business executives, 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 dearth of evidence concerning the excellence gap, and we

[1] In terms of representation, Asian students are also a minority. We show analyses by these finer race categories in the Appendix.

do not understand its causes and consequences at a deep enough level to advance useful policy advice (Plucker et al., 2010). Our paper fills this gap. We provide systematic, longitudinal evidence of the excellence gap by following students from 3rd grade through high school. We use comprehensive administrative data from the state of Michigan to study trends in racial differences in who reaches the highest levels of academic achievement over time to establish several excellence gap facts.

We first document the extent of the excellence gap and its drivers. We then examine variability of excellence gaps across schools. We additionally consider whether and how public policies, like accountability programs, affect the dynamics of the excellence gap as students progress through their education. 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.

More specifically, we document eight facts about the racial excellence gap. First, the gap is evident early on, as non-minority students in Michigan are around 3 times more likely to be in the top decile of math achievement by the 3rd grade compared to minority students. Second, the excellence gaps steadily grow with age, widening by roughly 10 percent from 3rd to 8th grade. Third, we examine the extent to which an individual student's socioeconomic status (SES) and features of the school environment explain the excellence gap. Proxies for an individual student's SES explain around a third of the gap. Additional controls for the school environment explain a little over 10 percent the excellence gap. Fourth, we find that the magnitude of the math and reading racial excellence gaps differ by gender, but that the gap widens much more for female students from 3rd to 8th grade in both subjects. Fifth, we document that there is considerable heterogeneity in the excellence gaps across schools, and that the within-school math and reading gaps are highly correlated.

Sixth, we provide evidence that accountability policies in Michigan had minimal effects on the excellence gap. The accountability policy required schools below a performance ranking threshold to undergo district-led changes to reduce the within-school gap between high and

low achieving students, regardless of race. The policy allows us to estimate the causal effect of accountability policies on the excellence gap based on a regression discontinuity design. Our results indicate the policy had minimal impact, with some suggestive evidence that it led to educational triage, and mostly harmed non-minority students.

Seventh, we document that mobility in and out of the right tail of achievement is critically linked to race both during grades 3 through 8 and in later years of life. We find that while nearly half of non-minority students who reach the top decile of achievement in 3rd grade remain in the top decile in 8th grade, only around 35% of minority students remain. Likewise, among those who did not reach the top decile in 3rd grade, around 7% of non-minorities enter the top decile in 8th grade, while only around 2% of minorities do. These dynamics continue into later-life academic achievements, including whether a student takes advanced placement courses or attains top performance on the SAT. Similar to our previous results, differences in SES explain much of these gaps.

Finally, we show that school factors and quality explain a substantially larger portion of the racial difference in the *persistence* of excellence rather than *entry* into excellence. For example, we find that variation in school factors and quality explains over 30% of the racial difference in persistence in math excellence from 3rd to 8th grade. This is in stark contrast to explaining entry rates where school quality explains only around 12% of the racial difference in reaching the top decile by the 8th grade.

Our main contribution is to provide a comprehensive documentation of the excellence gap over time using longitudinal, statewide data following students from 3rd grade through high school. This extends past work which has been more limited in scope based on available survey data (e.g. Wyner et al., 2007; Plucker et al., 2010; Hoxby and Avery, 2013; Meinck and Brese, 2019; Rambo-Hernandez et al., 2019). We additionally consider the potential drivers of the gap and show that racial differences in socioeconomic status help explain a sizable portion of the excellence gap. This acknowledges the important role of poverty (Plucker et al., 2013) and suggests that family resources, potentially including how parents

spend their time with their children, are important (Guryan et al., 2008). We also find that school factors play a notable but small role, somewhat contrasting with Hanushek and Rivkin (2009). However, about half of the gap is explained by factors not covered in our data.

We then further contribute to the growing body of evidence, recently reviewed in Meyer et al. (2024), 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 of the excellence gap which can be helpful for understanding which policies may work, especially since we find that school-specific factors only explain about 10% of the gaps. Second, we estimate the causal effects of Michigan accountability policies on the excellence gap and educational triage. Within the taxonomy from Meyer et al. (2024), adapted from Plucker et al. (2017) and Plucker and Peters (2020), this analysis specifically adds to the mixed evidence on the effects of K-12 school accountability support as a state and non-district intervention. Our results indicate that there is no causal effect of the Michigan accountability policies on the excellence gap, building on descriptive quantitative and qualitative work[2]. Our results additionally highlight the need to find policies that jointly pursue the goals of equity and excellence.

The remainder of our paper is organized as follows. Section I summarizes our administrative data and illustrates the setting. Section II presents our methodology and the seven facts about the racial excellence gap. Section III concludes.

I Data and Measurement

We use statewide administrative data from the Michigan Education Research Institute (MERI). The data include all public school students in the state of Michigan. We focus on student-level demographics, standardized measures of achievement from grades 3 through

[2]Kim et al. (2016); Hamilton et al. (2018); Kaul and Davis (2018); Hodges and Lamb (2019); Peters et al. (2019); Rambo-Hernandez et al. (2019); Turner and Spain (2020); Mun et al. (2021); Lockhart et al. (2022).

8, 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. As such, 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 signals 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, are in households receiving food or cash assistance, are homeless, are migrant, or are in foster care.[3] [4]

Given the data structure,[5] we construct two subsamples. First, we create a "main sample," consisting of students who began 3rd 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 through grades 3 to 8 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 3rd grade within the years 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.[6] Based on the structure of the data, we acknowledge two important limitations. First, the data initially captures students in 3rd grade and therefore does not allow us to measure the extent disparities arise earlier (as in Fryer Jr and Levitt, 2013). Second, our analysis excludes private school students who make up 7% of enrollments

[3] https://medc.miedresearch.org/dataset/k-12-student-demographic-and-enrollment-data

[4] We observe whether a student is determined as 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 8th grade. We therefore use an indicator for whether the student is economically disadvantaged in 8th grade as our main measure of socioeconomic status. In the Appendix, we show additional robustness checks when measuring socioeconomic status as the number of times a student is determined economically disadvantaged, conditional on the total number of times this information is available for the student.

[5] For the cohorts of students who began 3rd grade within the years 2007 and 2012, the data contain achievement assessment scores from 3rd grade through 8th grade. For the cohorts of students who began 3rd grade within the years 2007 and 2009, the data further link to information on high school outcomes.

[6] We include students who switch to home-schooling.

in Michigan (White and DeGrow, 2016).[7]

To understand the representativeness of our sample, we follow the logic in List (2020) and explore how observables that are correlated with preferences, beliefs, and constraints vary across our setting and target settings of interest. To do so, we present summary statistics as well as the relevant averages in the United States in Table 1. We find that Michigan is largely representative of the country in terms of demographics, school investments, and test scores. However, Michigan requires, and pays, for high school students to take a college entrance exam (Hyman, 2017). This is quite useful for our purposes, as it means that nearly all students take the SAT, compared to only about half of students nationally. Since Michigan was an early adopter of this policy, we believe that the high school gaps we document are likely smaller than those of states without this policy. However, many more states now have similar requirements. To provides a better background to understand the entire population of interest, Appendix A provides additional descriptive statistics for both the "main sample" and "high school sample".

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 too different from the national average. For example, Michigan does not have a voucher program (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 do, respectively), and administrators are required to have field experience and completed a preparation program that aligns with state standards (38 and 48 do). However, Michigan requires relatively few credits for high school graduation, and uses the SAT for high school math and English assessments rather than relying on state tests.[8]

[7]However, based on the publicly available national data from the National Assessment of Educational Progress (NAEP), the authors 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 calculation, using only public school students may lead us to slightly overstate the magnitude of the excellence gap.

[8]See https://nces.ed.gov/programs/statereform/index.asp for more information.

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

Notes: 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. 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). $*$ 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.

Measuring the Excellence Gap

As our measure of academic achievement, we use assessment scores from the Michigan Educational Assessment Program (MEAP) for all years until 2014, and the Michigan Student Test of Educational Progress (M-STEP) for younger cohorts we observe after 2014. The MEAP was the standardized test administered to all elementary and middle-school students

in Michigan until the 2013-14 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 3 through 8, measuring performance on state standards in English language arts, math, 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.

The excellence gap is the difference in the likelihood of reaching the highest levels of achievement in the entire state on the standardized math assessment for minority students compared to non-minority students.[9] Our main measure of the excellence gap focuses on whether a student reaches the top decile of achievement, following Rambo-Hernandez et al. (2019) and Plucker et al. (2010).[10] We additionally consider who reaches the top five and one percent.

II Empirical Results

In this section, we summarize our empirical results as seven key facts concerning the excellence gap. We begin by exploring temporal aspects of the gap, proceed to discuss mobility features by race, and conclude by discussing potential policy solutions.

1. The excellence gap emerges no later than third grade

Figure 1a illustrates the likelihoods of minority and non-minority students reaching "excellence" in math, defined as scoring in the top decile of achievement in standardized exams (see Figure A2 for reading). The top panel shows the fraction of students achieving excellence across the 3rd-8th grades. The bottom panel presents the percentage change relative to 3rd

[9]More specifically, we compare Black and Hispanic students, who are underrepresented in higher education and STEM fields and, on average, have low likelihoods of reaching the top decile with White and Asian students who have much higher likelihoods of reaching the top decile. We use terms "minority" and "non-minority," respectively, to capture disadvantaged students based on race and educational attainment but acknowledge that Asian students are, in fact, minority students.

[10]Due to ties in test scores, we do not always have a threshold score that defines exactly the top 10%. Table A2 shows the fraction of students categorized as excellent in each grade and subject.

grade. Here, we define the gap as the difference in the likelihood of reaching the top decile of achievement between Black or Hispanic students and White or Asian students:

$$\text{Excellence Gap} = \mathbb{P}(\text{Top Decile}_i|\text{Minority}_i) - \mathbb{P}(\text{Top Decile}_i|\text{Non-minority}_i) \quad (1)$$

The figures show that for the youngest students in our data, 3rd graders, the excellence gap already exists. Indeed, the data suggest a large and significant disparity: non-minority students are three times more likely to be in the top decile compared to minority students. While the magnitude of the gap is smaller for reading test scores than math test scores, non-minority students are still more than 2.5 times more likely to reach the top decile in reading than minority students (Figure A2). This result shows that the difference in the likelihood of reaching the top decile of achievement between Black or Hispanic students and White or Asian students is evident, and large, at young ages.[11] The magnitude of the racial disparity in excellence is around 80% of the disparity in excellence by socioeconomic status (A1).

2. The excellence gap slightly widens throughout early years of schooling, in both math and reading

Figure 1a also provides insights on our second fact—the excellence gap in math increases steadily with school-grade from 8.3 percentage points among 3rd graders to 9.2 percentage points among 8th graders. This widening of the gap is modest in size, with most of this slight change occurring between 3rd and 4th grades.[12] The increasing excellence gap observed happens due to entry and exit: Asian students are quite likely to grow into the top decile, whereas Black students exit from the top decile (Figure A3).[13]

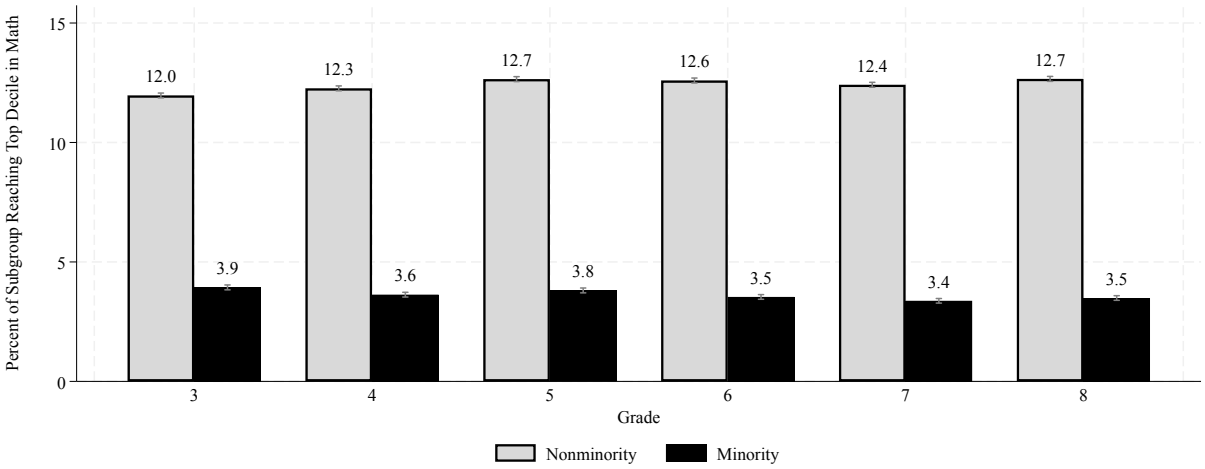
[11] We suspect that the gap emerges even earlier. Fryer Jr and Levitt (2013) find gaps among two and four year olds in a national sample, and smaller scale data sets on early childhood education, such as List et al. (2021), show gaps as early as three years old.

[12] Table A3 tests for differences in the excellence gap across grades.

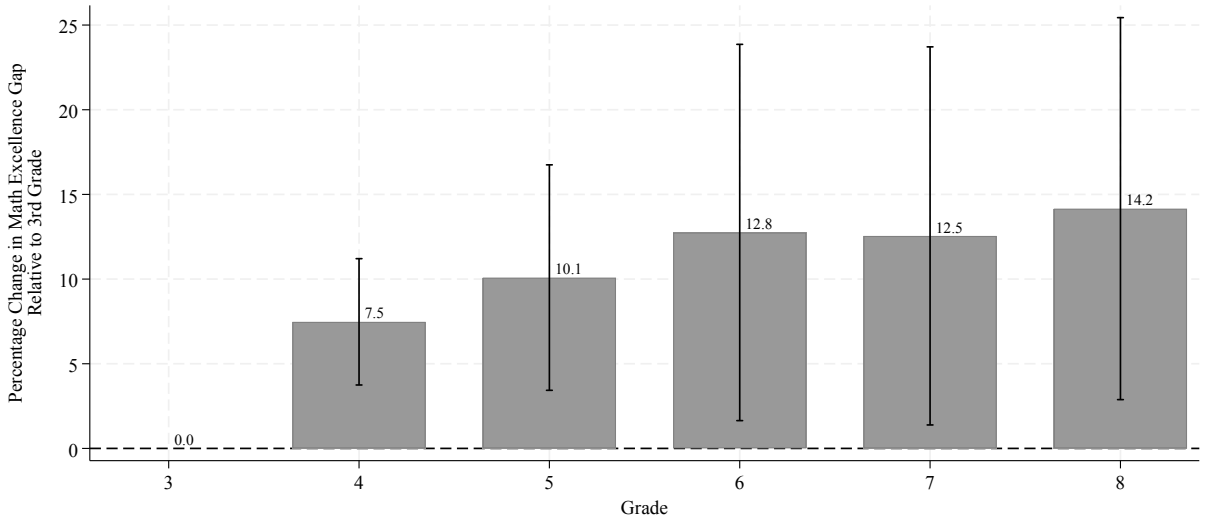
[13] One potential reason for these across-grade patterns is that minorities are more likely to switch schools particularly across transition grades, and this disrupts learning. Indeed, across grades 4 through 8, minority students are over 4 percentage points more likely to be in a different school than the prior year than non-

Figure 1: The Math Excellence Gap by Grade

(a) Likelihoods of Reaching Excellence



(b) Percentage Change in Excellence Gap From 3rd Grade



Notes: These figures display (a) the likelihood that a minority (Black or Hispanic) student or non-minority (White or Asian) student reaches excellence, and (b) 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 MEAP (for all years until 2014) or M-STEP (for all years after 2014). As noted in Section I, due to ties in test scores, we do not always have a threshold score that defines exactly the top 10%. Table A2 shows the fraction of students categorized as excellent in each grade and subject. Analysis uses the "main sample", as defined in Section I.

minority students (from a baseline of 71 percent). However, we find little evidence supporting the notion that differential rates of school-switching drives the widening of the excellence gap. While minorities are more likely to switch schools than non-minorities, accounting for changes in schools does not qualitatively

Figure 2a illustrates the likelihood that a non-minority student or minority student reaches excellence, averaged over grades 3 through 8. In math, non-minority students are on average around 9 percentage points, or 3 times, more likely to be in the top decile than minority students. As the Appendix shows, while the magnitude of the gap is slightly smaller for reading test scores than for math, these gaps arise in both areas. These patterns persist whether under alternative definitions of excellence, such as when considering reaching the top 5% or 1% (see Figure A26).

To provide deeper insights into this result, we show the achievement gaps by grade in Appendix C. These data reveal growth in the gap in math achievement with student age. However, the achievement gap and the excellence gap underscore two very distinct dimensions of racial gaps in education. Achievement gaps highlight differences at the center of each group's distribution. Our measure of excellence examines specifically the composition of the right tail of academic achievement.

The data also reveal an interesting gender effect: the excellence gap in math is around 2 percentage points smaller among female students than male students, and the excellence gap in reading is around 1 percentage point smaller among female students than male students. This is an interesting effect in isolation, and does not interact with race.

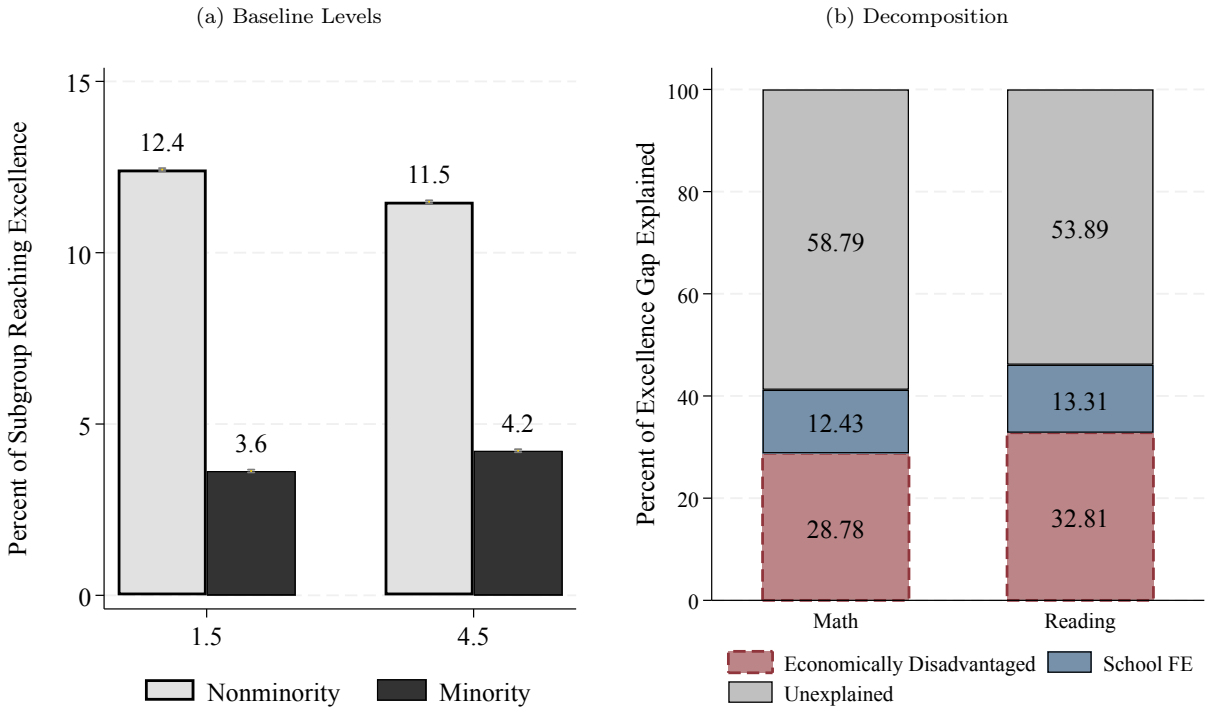
3. Nearly a third of the excellence gap is explained by socioeconomic status, and over a tenth is explained by school quality

Next, we explore the extent to which variation in students' socioeconomic status and heterogeneity in schools explain the excellence gap. Our main specification estimates, for student i of cohort c in grade g at school s and of socioeconomic status $e(i)$:

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

change the pattern (Table A3).

Figure 2: The Excellence Gap



Notes: These figures display (a) the percentage of minority and non-minority groups that reach the top decile on the MEAP (for all years until 2014) or M-STEP (for all years after 2014), and (b) the Gelbach decomposition of each minority and non-minority difference. These measures average over students in grades 3 through 8. "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 households receiving food or cash assistance, is homeless, is migrant, or is in foster care. This indicator is interacted with grade and cohort. "School FE" indicates the inclusion of school fixed effects. Analysis uses the "main sample", as defined in Section I.

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 (Excellence Gap $= \beta_1 = \mathbb{P}(Y_{i,g}|M_i = 1) - \mathbb{P}(Y_{i,g}|M_i = 0)$).

We conduct a Gelbach decomposition to understand the role of each covariate in explaining the excellence gap. 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.[14]

We plot our estimates in Figure 2b. Racial differences in socioeconomic status explains roughly a third of the excellence gap in math, while racial differences in schools attended explains around 12 percent. The patterns are quite similar across both math and reading achievement. However, we stress that our measures of being economically disadvantaged are coarse and so our analyses provide lower bounds on the importance of socioeconomic status for explaining 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, could be correlated with county-wide and city-wide features. In fact, the school fixed effects likely capture a variety of local features both within and outside of the school.

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 describing racial dispar-

[14]We are only able to include an indicator for socioeconomic status based on the available data. While our results with this variable indicate that socioeconomic status matters, the results likely understate its importance. We show in Figure A5 the same analyses as in Figure 2b, but using the number of times that a student is observed as being economically disadvantaged, conditional on the number of times that this information is available.

ities in other parts of the achievement distribution. In particular, we conduct the same decomposition exercise for racial differences in surpassing lower levels of achievement: reaching above the median or top 90% of achievement (Figure A7). We find that variation in schools explains a substantially larger portion of these lower-level racial differences than they do in explaining the excellence gap. In comparison, socioeconomic status explains a smaller portion of the variation. This highlights that the drivers of the excellence gap are not necessarily the same drivers of disparities in lower levels of academic performance.

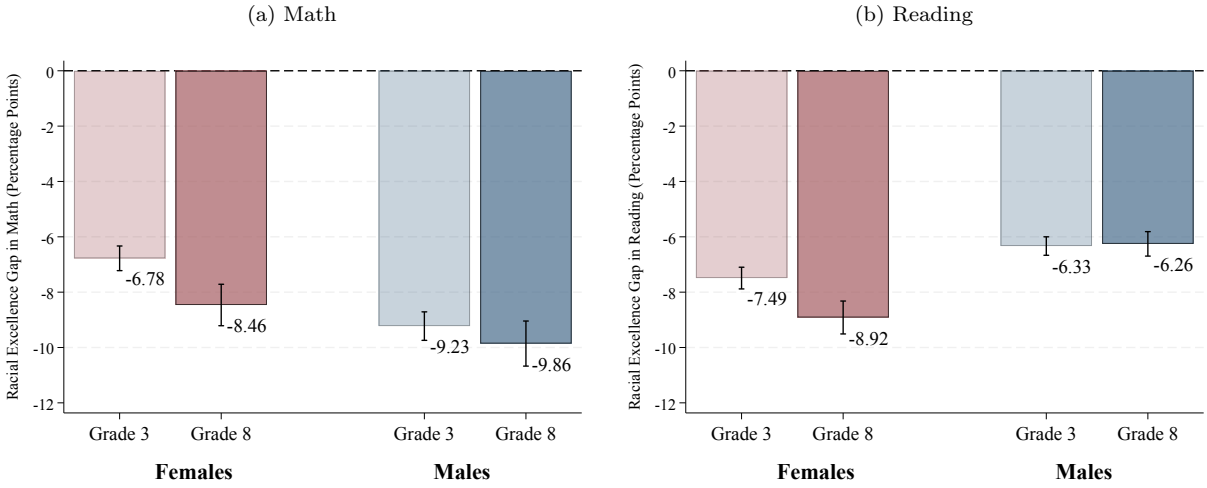
Finally, we have so far considered time-invariant school fixed effects. This may underestimate the role of school traits on the excellence gap if school quality and neighborhood traits vary over time. To address this concern, we estimate the main specification (Eq 2) but with school-by-year-by-cohort fixed effects, and find that the excellence gap coefficient minimally changes (Figure A4).

4. The magnitude of the math and reading racial excellence gaps differ across by gender, but the growth from 3rd to 8th grade in both subjects is driven by female students

We now consider the extent to which the excellence gap differs for female and male students. We separately estimate the excellence gap by subject and sex for 3rd and 8th grade students and present the results in Figure 3. Interestingly, the direction of gender differences depends on the subject. Across grades, the racial excellence gap in math is larger among male students than female students, while the racial excellence gap in reading is larger among female students.

In contrast, growth in the racial excellence gap with age also differs by gender but in the same direction across subjects: in both subjects, the change in the excellence gap is substantially larger among girls than among boys. These results are in contrast to previous work studying the lower tail of the achievement distribution that finds that (economically disadvantaged) black boys are more likely to fall behind as they progress through school

Figure 3: Excellence gaps by gender



Notes: These figures show for female and male students the proportion of each minority group that reaches excellence. Analysis uses the "main sample", as defined in Section I.

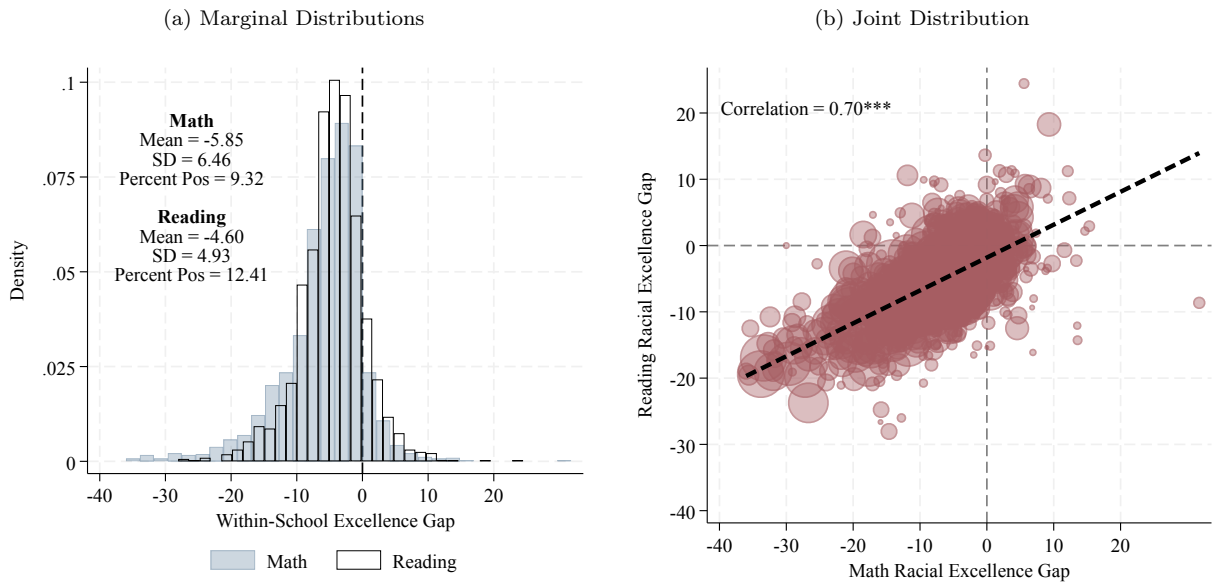
(Autor et al., 2023). In Appendix E, we show that particularly for math, these gender differences are largely explained by baseline differences in achievement.

5. There is heterogeneity in the excellence gap across schools, and the gaps across subjects is highly correlated

We next investigate how racial differences in excellence vary across schools. We calculate school-specific excellence gaps in math and reading by calculate excellence gaps for each school-year-subject in the data, and then average over years in each subject.[15] Figure 4 illustrates their distributions. We find that the average excellence gap masks heterogeneity and reject the null hypothesis that excellence gaps are homogeneous across schools (Pearson Chi-squared test, $p < 0.001$). Although excellence gaps vary across grades (Section II), this across-school heterogeneity is not driven by differences in grades associated with schools: accounting for grades minimally changes the results. While minority students are around 5 to 6 percentage points less likely to reach excellence in the average school, this gap widens to around 10 to 13 percentage points for schools in the 10th percentile of the excellence

[15]We remove school-years that have less than 10 minority or non-minority test-takers.

Figure 4: Excellence Gaps Across Schools



Notes: These figures display (a) the distribution of the math and reading excellence gaps across schools and (b) the relationship between a school's excellence gap in math and in reading. Positive values of the excellence gap are in favor of minorities. A unit of observation is a school. These measures pool students across years. The linear fit in panel (b) weights schools by the number of students who take a standardized exam in the school. The mean and standard deviation (SD) of the school-level excellence gap is written in (a). "Percent Pos" refers to the percentage of schools that have a non-negative excellence gap in the corresponding subject. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

gap. We estimate a standard deviation of around 5-6 percentage points in both subjects.[16] We describe our procedures in further detail in Appendix D. Interestingly, we find that our observed school-grade excellence gaps are quite stable over time (Figure A8).

However, despite across-school variation in the magnitude of the excellence gap, less than 10% of all schools exhibit a positive excellence gap in math. This pattern persists across grades (Figure A9). Similarly, less than 13% of all schools have a positive excellence gap in reading. In other words, non-minority students are more likely to reach excellence than minority students in most schools.

[16]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 Appendix D, we follow the steps of Kline et al. (2022) to estimate the distribution of latent school-levels excellence gaps, which treats yearly school-level gaps as noisy estimates of a school's latent gap.

We find that a school's excellence gaps across subjects are positively linked: the correlation between a school's math and reading excellence gap is 0.70 (Figure 4b). We find similar patterns when measuring excellence gaps as proportional differences in the rates of reaching excellence (Figure A10).

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					-11.05***	-1.36**					-8.32***
	(0.78)					(0.95)	(0.61)					(0.78)
Frac. Econ Disad.		10.49***				-2.29**		4.32***				-1.82*
		(0.82)				(1.11)		(0.58)				(1.04)
Average Math Score (std)			-14.78***			-13.57***			-1.22			-0.34
			(1.18)			(1.10)			(0.89)			(0.82)
Average Reading Score (std)			7.69***			0.57			-3.34***			-8.80***
			(1.21)			(1.31)			(0.99)			(1.14)
N Test-takers (log)				-0.47		0.12				0.33		0.39
				(0.38)		(0.31)				(0.26)		(0.26)
Location: Suburb					0.19	0.82***					0.55*	0.74***
					(0.44)	(0.30)					(0.29)	(0.24)
Location: Rural					1.48***	0.60*					0.83***	0.33
					(0.44)	(0.31)					(0.29)	(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

Notes: This table correlates 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. Standard errors in parentheses, clustered at school level. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I. * $p < .1$, ** $p < .05$, *** $p < .01$

From regressing a school's excellence gap on its demographic composition and average student achievement, we find in our data that the schools that display the most negative excellence gaps are schools with a larger fraction of minority and economically disadvantaged students students (Table 2).[17] Schools located in suburbs tend to exhibit more positive excellence gaps, after accounting for differences in achievement level and demographic composition.[18]

[17]We show in Appendix A4 that similar patterns emerge when predicting whether a school has a positive excellence gap, meaning that minority students have higher rates of reaching excellence than non-minority students.

[18]In our data, around 45 percent of schools are classified as located in a suburb, 26 percent as rural, and 4 percent are missing a classification.

6. Past policies targeting achievement gaps have failed to reduce the excellence gap

The first four facts perhaps raise more new questions than they answer. For example, how malleable is the excellence gap? And, are the excellence gap dynamics affected by school or state policies? These are two critical policy questions that naturally arise after establishing an excellence gap. Our fifth fact begins to unpack the dynamics of the excellence gap in relation to public policies. More specifically, and in light of our results that the school fixed effects explain a large percent of the excellence gap, we aim to further understand the extent that the excellence gap from third to eighth grade is driven by differences in school resources. To complete this task, we examine whether and to what extent changes in state accountability policies affect the composition of which students reach excellence.

Past work finds that test-based accountability polices that target average student achievement or proficiency can cause schools to redirect resources away from the highest achieving students (e.g. Neal and Schanzenbach, 2010; Hemelt and Jacob, 2020). 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 students (e.g. Jackson and Mackevicius, 2024) or schools are more likely to redirect resources away from high-achieving minority students than high-achieving non-minority students because of differences in social capital (Reardon et al., 2019), then the excellence gap will widen.

We investigate the effects of a post-NCLB policy in Michigan that was designed specifically with the goal to reduce (non-race-based) within-school inequality in highly disadvantaged schools. Starting in 2012, the lowest performing schools were given a "Priority" designation, and required to make significant 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 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 Hemelt and Jacob (2017, 2020), who 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%), 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. More information is available from the state of Michigan.[19]

The assignment mechanisms for the Focus and Priority designations allow us to estimate the causal effect of this accountability policy using a regression discontinuity (RD) design, where the school performance index (for Priority) and composite gap index (for Focus) are the running variables. We estimate the causal effect of being assigned to a Focus or Priority school on the excellence gap in the first 4 years of implementation, following the RD design in Hemelt and Jacob (2020, 2017). They find that the Focus designation shrank the within-school math achievement distribution (e.g. difference between school's 90th and 10th percentiles) in the short run, but that this effect dissipated quickly. They also find suggestive evidence of increased teacher departures from Focus schools, but not Priority schools. They do not consider excellence (90th percentile across all schools) or differential effects by race, which is the focus of our analysis.

We assess the suitability of the regression discontinuity research design. First, we show that the running variables and their respective cutoffs indeed generate substantial discontinu-

[19] "Top-to-Bottom Ranking, Priority, Focus and Rewards Schools Identification Business Rules."

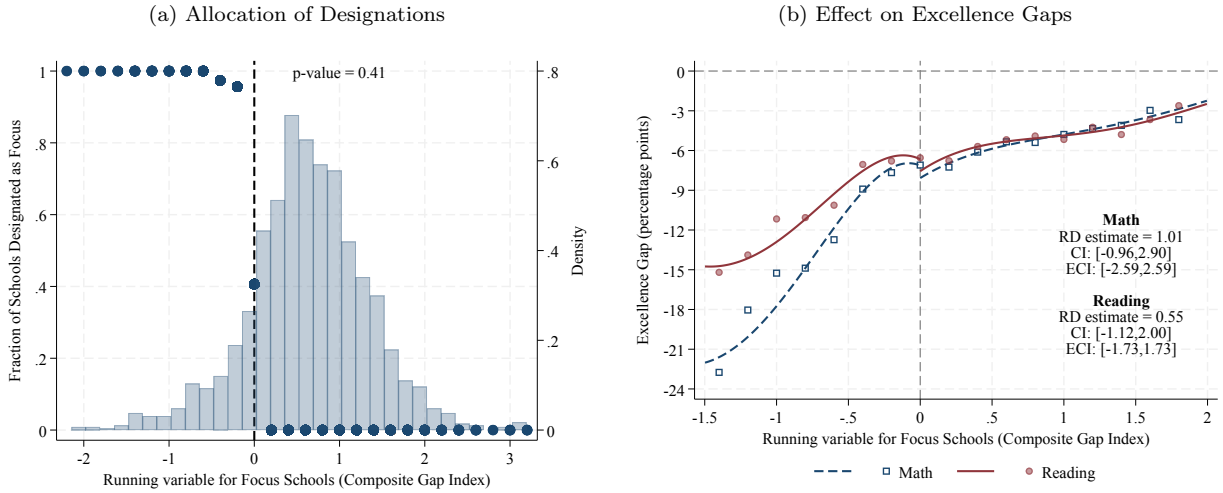
ities in whether a school is deemed as Priority or Focus (Figures A12a and 5a, respectively). Second, we investigate the densities of the running variables and find no strong evidence of discontinuities in density at the cutoffs using a test of manipulation following Cattaneo et al. (2020). Third, we conduct numerous placebo tests, using pre-determined school characteristics as outcomes, and find little evidence for imbalance (Figures A13 and A14). This garners support for our identification assumption that there are no imbalances on unobservables that affect our outcomes of interest. Taken 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. We provide more information about the policy and the research design in Appendix F.

We consider the effect of the accountability policies on the within-school excellence gap, as computed in previous sections. Figure 5 summarizes for Focus, and Figure A12 for Priority. We find that neither a Priority nor Focus designation significantly impacted school-level math and reading excellence gaps. However, we acknowledge that the point estimate on Math is about 10% of the excellence gap, and similar in magnitude to the growth in the excellence gap between 3rd and 8th 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).[20]

Null effects on the excellence gap may mask co-movements in subgroups' likelihoods of reaching excellence. However, we find little evidence supporting the notion that the policies improved rates of excellence for either subgroup (Figure A15). We find suggestive evidence that the Priority designation worsened non-minority students' likelihood of achieving excellence in math by 0.33 percentage points, though this is only significant at the 10% level. Point estimates suggest a larger reduction for Focus designation on non-minority students,

[20]We follow Cattaneo et al. (2019) and present point estimates together with bias-corrected confidence intervals. For this reason, the point estimates are not in the center of the confidence intervals.

Figure 5: Effect of Focus Designations on Excellence Gaps



Notes: This figure shows the effects of (a) distribution of the school-level composite gap index, which is the running variable used to determine Focus designation and (b) the effect of Focus school designations on racial excellence gaps in math and reading. Schools to the right of the Focus cutoff are designated as Focus schools. The scatterplot in panel (a) bins schools by every 0.2 units of the index. The p-value in (a) corresponds to a test of manipulation using local polynomial density estimation by Cattaneo et al. (2020), against the null hypothesis of equality at the cutoff. The lines in (b) represent local-linear fits following Calonico et al. (2017), and the scatterplot bins observations by 0.2 units of the running variable. Positive values of the excellence gap are in favor 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 math MEAP (for all years until 2014) or M-STEP (for all years after 2014). CI are robust bias-corrected 95% confidence intervals (Calonico et al., 2014). ECI are equivalence-test 95% confidence intervals, following Hartman (2021). Both cluster at the school level.

but it is not statistically significant. Overall, this suggests that accountability policies have at best small effects on the excellence gap or educational triage.[21]

Taken 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.

7. Mobility in and out of the right tail of achievement is critically linked to race

Our first five facts focused on excellence gaps from 3rd through 8th grade. Taking advantage of our longitudinal data, we now consider individual students' trajectories to understand to what extent high achieving 3rd grade students persist in the top decile, and lower achieving

[21]Note that this is not by construction 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.

student move up into excellence. To examine mobility, we continue with an exploration of data from grades 3-8, then we extend the analysis to later years.

Grades 3-8

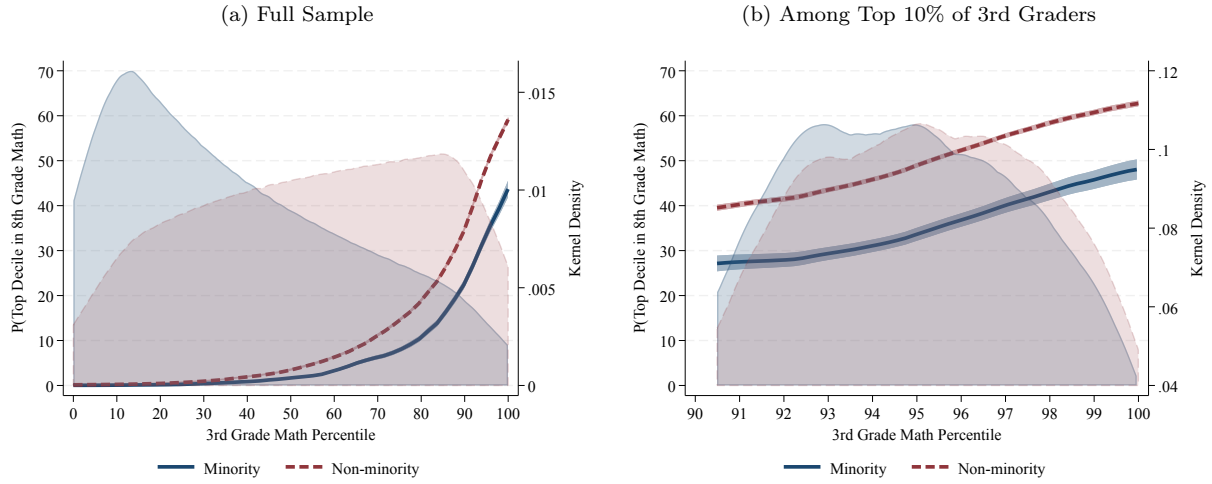
Figure 6 plots the likelihood that an individual student in each 3rd grade decile is in the top decile in 8th grade, as well as the distributions of 3rd grade achievement by race. Strikingly, at each 3rd grade achievement level, minority students are less likely to be in the top decile by the time they reach the 8th grade compared to non-minority students with *identical* 3rd grade test scores. Among students who reach the top decile of achievement in 3rd grade, while nearly half of non-minorities stay in the top decile in 8th grade, only around 35% of minority students remain. In other words, minority students who reach the top decile in 3rd grade are about 17 percentage points more likely to fall out of the top decile than their top achieving non-minority counterparts. Part of this difference 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 90-94th percentiles, while non-minority students are more likely to be in the 96-99th percentiles (Panel B of Figure 6).

Interestingly, at every percentile of 3rd grade achievement from 90 to 99, non-minority students are also much more likely to remain in the top decile than their minority peers with identical test scores (Panel B of Figure 6). Therefore, differences in the distribution of 3rd grade achievement by race alone only minimally explain differential persistence. We find the same patterns with reading (Figure A16).

When discussing these results, especially our results on persistence that condition on a student's 3rd grade score, it is important to acknowledge that test scores on any given day are noisy measures. We describe and explore these concerns in Appendix G.1.

In a similar vein, we find large racial differences in whether a student who did not reach excellence in 3rd grade is able to achieve it in 8th grade. While 7% of non-minority students who did not reach the top decile in 3rd grade end up reaching it in 8th grade, only around

Figure 6: 8th Grade Excellence by 3rd Grade Rank



Notes: These figures show the proportion of students who reach the top decile of the 8th grade achievement by 3rd grade achievement, and the distribution of 3rd grade achievement. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Figure (a) shows for each 3rd grade decile group, while (b) shows for each percentile in the top 10 percent. Shaded areas around lines correspond to the 95% confidence intervals. Shaded areas reflect the density of students at each achievement level. Analysis uses the "main sample", as defined in Section I.

2% of minority students accomplish that feat.

These patterns in the excellence gap are qualitatively the same across females and males (Figure A11). We explore this further in Appendix E. When considering finer race categories, the largest differences are between between Asian and Black students (Figure A18).[22]

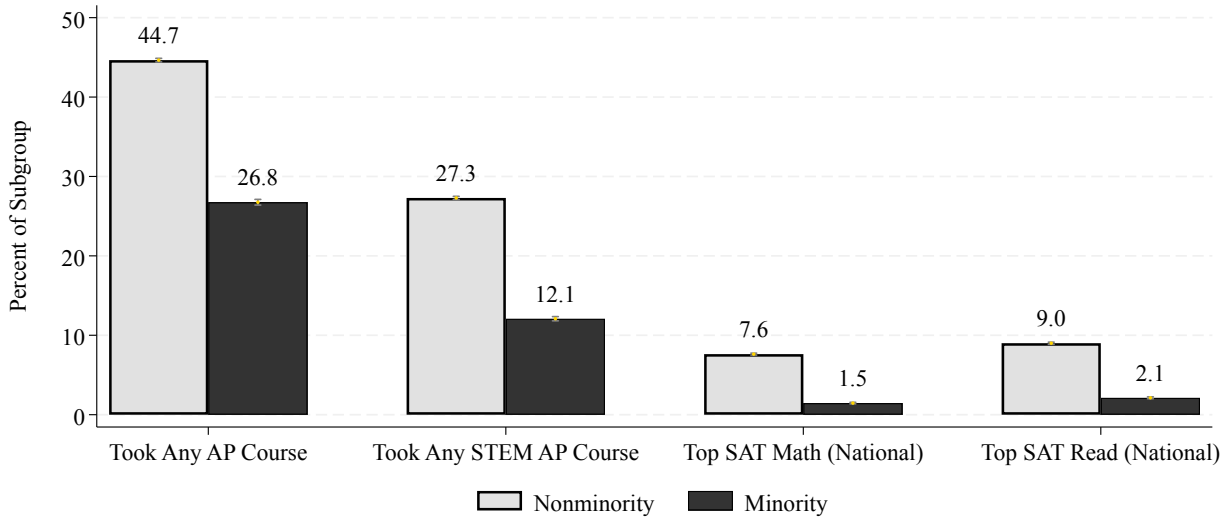
Persistence into College Readiness

We next examine racial differences in several measures of college readiness. We have measures for taking at least one AP course, taking at least one STEM AP course, or scoring in the top decile on the math or reading section of the SAT. For these, we use the "high school sample", as defined in Section I.

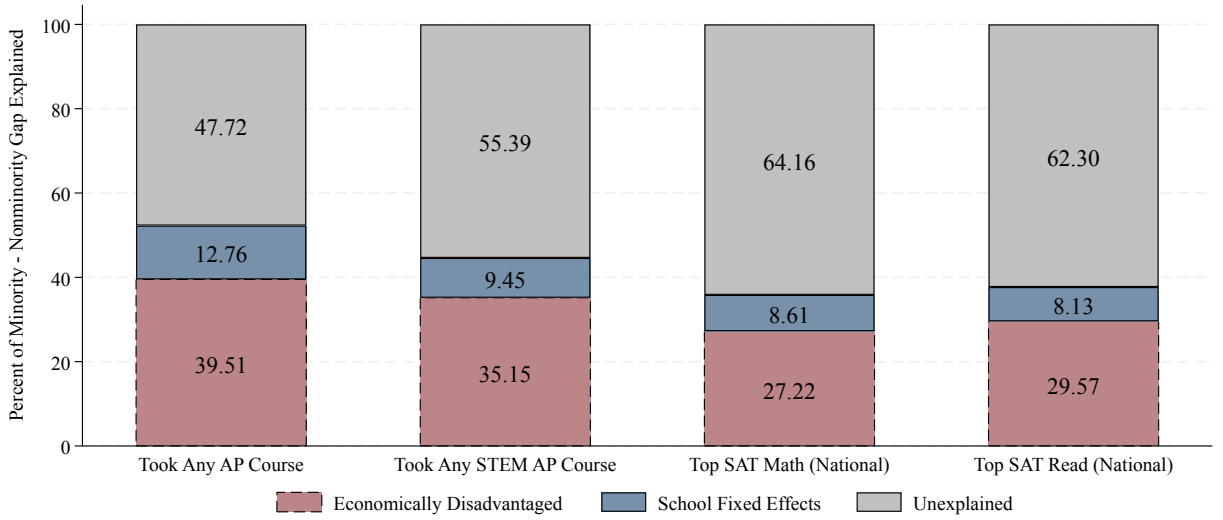
[22]Note that part of the racial difference in the likelihood of falling out of the top decile is, in some sense, mechanical due to the underlying distribution of 3rd grade achievement. As we show Figure 6b, top decile minority students are over-represented in the lower percentiles (e.g. 90) compared to non-minority students. Since a larger share of minority students are marginal to the top decile, they are more likely to fall out of the top decile in 8th grade. Nevertheless, this does not explain our result. Recall that at every percentile of 3rd grade achievement from 90 to 99, non-minority students are much more likely to remain in the top decile than their minority peers with the same scores.

Figure 7: Gaps in College Readiness

(a) Baseline Levels



(b) Decomposition



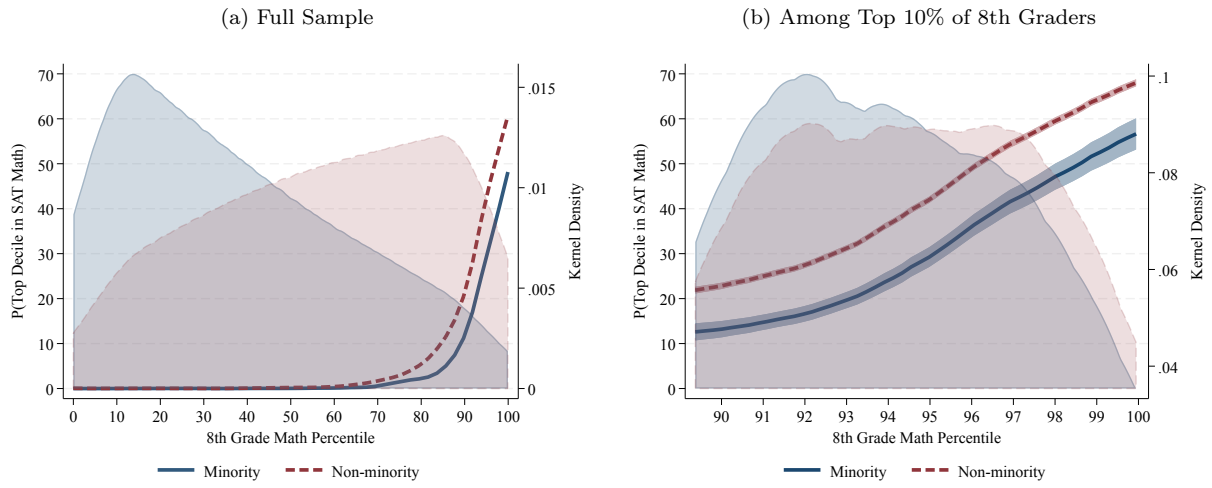
Notes: This figure displays (a) the percentage of each race/ethnicity group that takes at least one AP course, takes at least one STEM AP Course, reaches the top decile of SAT Math, and reaches the top decile of SAT Evidence-Based Reading & Writing, and (b) the Gelbach decomposition of each minority non-minority difference. For SAT standards, we use the top decile of nationwide test-takes, using the 90th percentile score reported by College Board for the year that the exam was taken. "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 the "high school sample", as defined in Section I.

Figure 7 illustrates considerable racial differences in these outcomes. Minority students are substantially less likely to take an AP course, take a STEM AP course, take the SAT,

and score in the top decile in the SAT for both math and reading than their non-minority race counterparts. While a slightly larger percentage of non-minorities take the SAT than minorities, both groups have an incredibly high rate of taking the standardized exam. One reason for this is the state's offering of the exam to all public school juniors since 2007 (Hyman, 2017).

Following the same decomposition steps as in Section II, we find that racial differences in student socioeconomic status explain around 27-45 percent of racial disparities in college readiness (Figure 7b). While school fixed effects help explain some gaps, they generally explain less than 10 percent of the gaps. In sum, student socioeconomic status explains much more of the excellence gap variation than school fixed quality.

Figure 8: Excellence in the SAT Math Section by 8th Grade Rank



Notes: These figures show 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. Figure (a) shows for each 8th grade decile group, while (b) show for each percentile in the top 10 percent. Analysis uses the "high school sample", as defined in Section I.

These racial disparities in college-readiness exist even when honing in on students who reached excellence in prior years (Figure 8). Even among students who reached the top decile in 8th grade, 34% of non-minorities reach the top decile in the SAT while only 20% of minorities make that level of excellence. Among 8th graders who did not reach excellence, non-minority students are more than 4 times more likely to achieve top performance on

the SAT than minority students. These differences extend to excellence in reading as well (Figure A17), and are statistically significant at conventional levels.

Accounting for excellence gaps during grades 3 through 8 explains some of the college readiness gaps (Figure A21). With regards to AP course-taking, racial excellence gaps in 8th grade explain around 8 to 13 percent of the total gap, and accounting for grade 3 through 7 explains around 20 percent. The excellence gap explains a substantially higher fraction of the racial disparity in SAT scores, with 8th grade excellence gaps explaining around 24-30 percent of the total gap, and excellence gaps in prior years explaining around 50 percent.

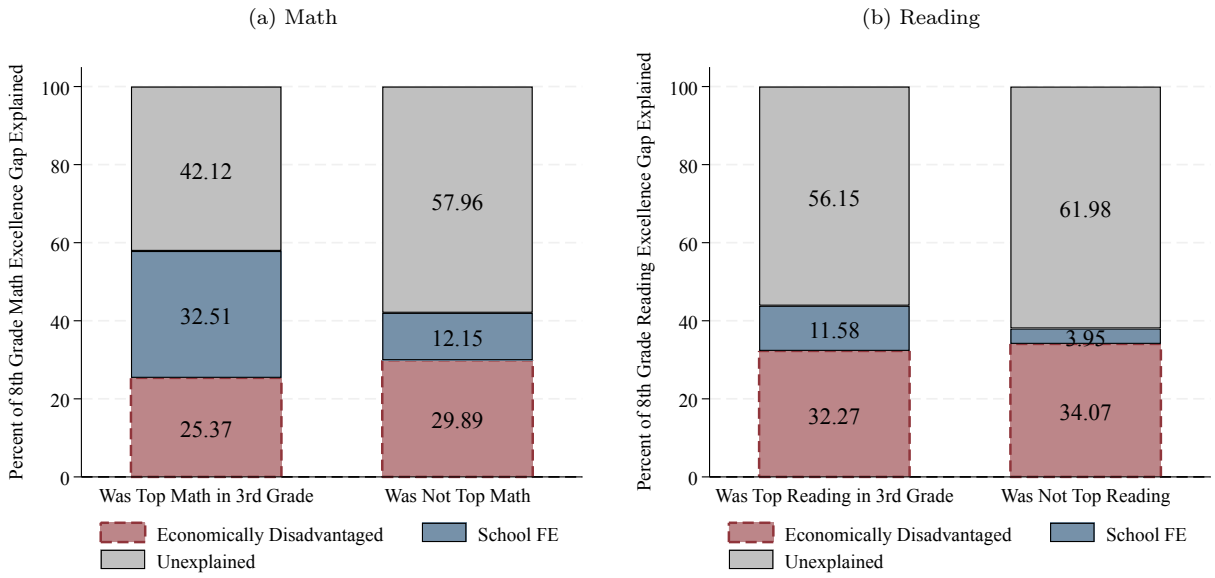
8. School quality explains a substantially larger portion of the racial difference in the *persistence* of excellence rather than *entry* into excellence

Finally, we investigate the extent to which socioeconomic status and school factors and quality explain mobility in and out of the deep right tail. For math test scores, we find that variation in school quality explains over 30% of the racial difference in persistence in excellence from 3rd to 8th grade (Figure 9a). This is in stark contrast to the racial difference in entry into 8th grade excellence: among students who were not in the top decile for math in the 3rd grade, school quality explains only around 12% of the racial difference in reaching the top decile by 8th grade.

For reading, school quality explains a smaller portion of the racial disparity in the persistence of excellence— around 12% (Figure 9b). However, we again find that school quality explains substantially more of the variation in persistence than entry: school quality explains around 4% of the racial difference in whether a student who was not in the top decile of reading in 3rd grade enters the top in 8th grade. In Appendix Figure A20, we separate out the sample of students who did not reach excellence in 3rd grade and find that these patterns persist in these finer bins.

Importantly, these patterns persist into college-readiness. In this case, school quality explains substantially more of the racial differences in all four of our measures of college

Figure 9: Explaining Racial Differences in Persistence and Entry into 8th Grade Excellence



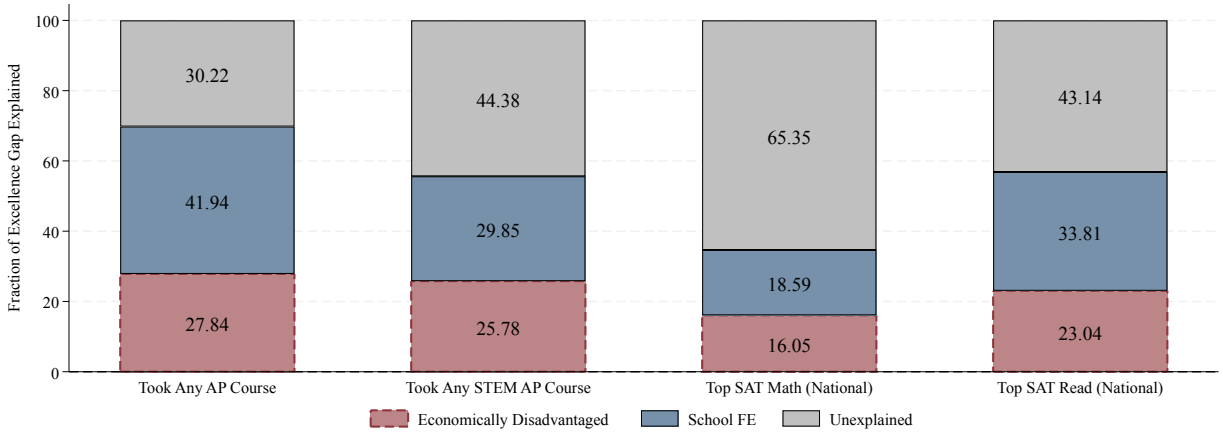
Notes: These figures display the Gelbach decomposition in the 8th grade excellence gap in (a) math test scores (b) reading test scores. Both figures split the sample by whether a student achieved excellence in that subject in 3rd grade. For example, the leftmost bar of panel (a) illustrates the Gelbach decomposition for the racial difference in persistence of excellence: whether a student reaches the top decile of math in 8th grade, among students who scored in the top decile of math in 3rd grade. "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 households receiving food or cash assistance, is homeless, is migrant, or is in foster care. "School FE" indicates the inclusion of school fixed effects. Analysis uses the "main sample", as defined in Section I.

readiness among students who were in the top decile of 8th grade compared to students who were not in the top decile, whether we look at excellence in math or reading (Figure 10 for math and Figure A22 for reading).[23] All told, 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.

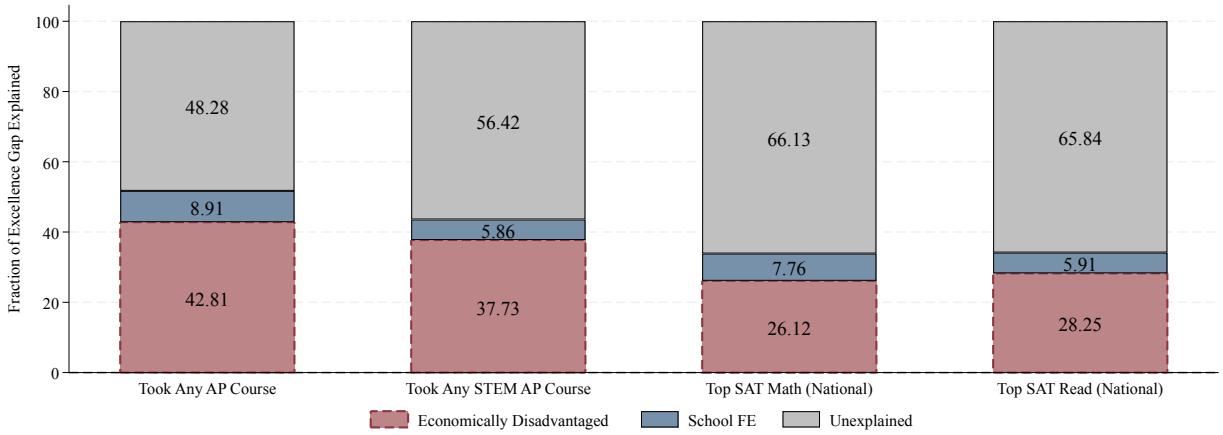
[23]Figure A24 bins students by decile in 3rd grade.

Figure 10: Explaining Racial Differences Persistence and Entry in College Readiness (Math)

(a) Persistence (Among Those Top in 8th Grade Math)



(b) Entry (Among Those Not Top in 8th Grade Math)



Notes: This figure displays the Gelbach decomposition of each minority non-minority difference in whether a student takes at least one AP course, takes at least one STEM AP Course, reaches the top decile of SAT Math, and reaches the top decile of SAT Evidence-Based Reading & Writing, among (a) students who achieved excellence in 8th grade math and (b) students who did not achieve excellence in 8th grade math. For SAT standards, we use the top decile of nationwide test-takes, using the 90th percentile score reported by College Board for the year that the exam was taken. "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 the "high school sample", as defined in Section I.

III Conclusion

In modern economies characterized by increasing returns to human capital and "winner take all" markets, small departures in the right tail can lead to considerable surplus creation. A

natural corollary is that large welfare losses to an economy can result because of deterioration of right tail talent. This fact highlights a major oversight of education policy intending to achieve educational equity by focusing on other parts of the achievement distribution. In addition, while studying both proficiency gaps and excellence gaps is important, excellence gaps are equally as concerning because they are both growing rapidly in size, and because the groups failing to reach high levels of achievement are the fastest growing segments of the U.S. population (Rambo-Hernandez et al., 2019). Importantly, we do not argue that researchers and policymakers should substitute away from remedial reforms in favor of excellence reforms, but rather that policies to reduce excellence gaps should have more prominence than they do currently, as society needs both.

In a macro sense, our results directly link to competitiveness on the global scale. The lack of attention to high-achieving students should be of higher concern if the U.S. plans to compete in an increasingly global marketplace where high-level math, science, and language skills are crucial (Hess and Rotherham, 2008). This calls attention to the equity-efficiency tradeoff that policymakers face. Ensuring minimum proficiency of every student can lead us towards educational equity, though, this proficiency standard is insufficient as a measure of educational quality or economic competitiveness (Rambo-Hernandez et al., 2019). We must meet the learning needs of every student to fully achieve equity, and in turn, bring greater efficiency in policy to reach longer-term economic goals.

We contribute to this effort by documenting the excellence gap from 3rd grade through college 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, and extending our reach to study the persistence of the gap through the key measures of college readiness. 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 it (Meyer et al., 2024). Our study

perhaps indirectly suggests that rather than using state policies, school level policies, like gifted education or extracurricular programs (Card and Giuliano, 2016; Assouline et al., 2017; Ihrig et al., 2018, e.g.), may be effective.

Overall, we view this work as an important step to systematically document 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. Based on past work (e.g. Meyer et al., 2024), we may expect these patterns to be broadly representative in the United States. The replication of these analyses across states and countries will strengthen our understanding and better inform the key moderators. We encourage researchers to further document the excellence gap across contexts.

References

Assouline, S. G., L. M. Ihrig, and D. Mahatmya (2017). 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 Quarterly 61*(3), 250–261.

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2023). Males at the tails: How socioeconomic status shapes the gender gap. *The Economic Journal 133*(656), 3136–3152.

Booher-Jennings, J. (2005). Below the bubble:"educational triage" and the texas accountability system. *American educational research journal 42*(2), 231–268.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Card, D. and L. Giuliano (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences 113*(48), 13678–13683.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'no child left behind'. *Race ethnicity and education 10*(3), 245–260.

Dinno, A. (2023). *TOSTREGRESS: Stata module to perform equivalence test with two one-sided tests for regression coefficients*. Portland State University. Version 1.2.3.

Fryer Jr, R. G. and S. D. Levitt (2013). Testing for racial differences in the mental ability of young children. *American Economic Review 103*(2), 981–1005.

Gamoran, A. (2008). *Standards-Based Reform and the Poverty Gap: Lessons for" No Child Left Behind"*. Brookings Institution Press.

Gillborn, D. and D. Youdell (1999). *Rationing education: Policy, practice, reform, and equity*. McGraw-Hill Education (UK).

Guryan, J., E. Hurst, and M. Kearney (2008). Parental education and parental time with children. *Journal of Economic perspectives 22*(3), 23–46.

Hamilton, R., D. B. McCoach, M. S. Tutwiler, D. Siegle, E. J. Gubbins, C. M. Callahan, A. V. Brodersen, and R. U. Mun (2018). Disentangling the roles of institutional and individual poverty in the identification of gifted students. *Gifted Child Quarterly 62*(1), 6–24.

Hanushek, E. A., P. E. Peterson, L. M. Talpey, and L. Woessmann (2019). The achievement gap fails to close. *Education Next 19*(3), 8–17.

Hanushek, E. A. and S. G. Rivkin (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of policy analysis and management 28*(3), 366–393.

Hartman, E. (2021). Equivalence testing for regression discontinuity designs. *Political Analysis 29*(4), 505–521.

Hemelt, S. W. and B. Jacob (2017). Differentiated accountability and education production: Evidence from nclb waivers. Technical report, National Bureau of Economic Research.

Hemelt, S. W. and B. A. Jacob (2020). How does an accountability program that targets achievement gaps affect student performance? *Education Finance and Policy 15*(1), 45–74.

Hess, F. M. and A. J. Rotherham (2008). Can nclb survive the competitiveness competition? *School Business Affairs 74*(2), 8.

Hodges, J. and K. Lamb (2019). Washington's high-ability programs during the no child left behind era. *Journal for the Education of the Gifted 42*(4), 283–302.

Hoxby, C. M. and C. Avery (2013). The missing "one-offs": The hidden supply of high-achieving, low income students. Technical report.

Hyman, J. (2017). Act for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy 12*(3), 281–311.

Ihrig, L. M., E. Lane, D. Mahatmya, and S. G. Assouline (2018). Stem excellence and leadership program: Increasing the level of stem challenge and engagement for high-achieving students in economically disadvantaged rural communities. *Journal for the Education of the Gifted 41*(1), 24–42.

Jackson, C. K. and C. L. Mackevicius (2024). What impacts can we expect from school spending policy? evidence from evaluations in the united states. *American Economic Journal: Applied Economics 16*(1), 412–446.

Kaul, C. R. and B. K. Davis (2018). How the state education agencies addressed gifted education in the title ii sections of their essa state plans. *Gifted Child Today 41*(3), 159–167.

Kim, J., S. Watkins, and S. W. Yoon (2016). Social justice alert: Future innovators and leaders fall behind in academic achievement. *Education Leadership Review 17*(1), 129–141.

Kline, P., E. K. Rose, and C. R. Walters (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics 137*(4), 1963–2036.

Kober, N. (2001). It takes more than testing: Closing the achievement gap. a report of the center on education policy.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in us schools. *Educational researcher 35*(7), 3–12.

List, J. A. (2020). Non est disputandum de generalizability? a glimpse into the external validity trial. Technical report, Working paper.

List, J. A., J. Pernaudet, and D. L. Suskind (2021). Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. *Nature Communications 12*(5765), 1367–1378.

Lockhart, K., M. S. Meyer, and K. Crutchfield (2022). A content analysis of selected state plans for gifted and talented education. *Journal of Advanced Academics 33*(1), 3–42.

Magnuson, K. A. and J. Waldfogel (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The future of children*, 169–196.

Meinck, S. and F. Brese (2019). Trends in gender gaps: using 20 years of evidence from timss. *Large-Scale Assessments in Education 7*(1), 1–23.

Meyer, M. S., Y. Shen, and J. A. Plucker (2024). Reducing excellence gaps: A systematic review of research on equity in advanced education. *Review of Educational Research 94*(1), 33–72.

Michelmore, K. and S. Dynarski (2017). The gap within the gap: Using longitudinal data to understand income differences in educational outcomes. *AERA Open 3*(1), 2332858417692958.

Mun, R. U., M. D. Ezzani, L. E. Lee, and J. K. Ottwein (2021). Building systemic capacity to improve identification and services in gifted education: A case study of one district. *Gifted Child Quarterly 65*(2), 132–152.

Murnane, R. J. (2007). Improving the education of children living in poverty. *The future of children*, 161–182.

Neal, D. and D. W. Schanzenbach (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics 92*(2), 263–283.

Paige, R. (2006). No child left behind: The ongoing movement for public education reform. *Harvard Educational Review 76*(4), 461–473.

Peters, S. J., M. Gentry, G. W. Whiting, and M. T. McBee (2019). Who gets served in gifted education? demographic representation and a call for action. *Gifted Child Quarterly 63*(4), 273–287.

Plucker, J., J. Hardesty, and N. Burroughs (2013). Talent on the sidelines: Excellence gaps and america's persistent talent underclass. *Center for Education Policy Analysis, University of Connecticut*.

Plucker, J. A., N. Burroughs, and R. Song (2010). Mind the (other) gap! the growing excellence gap in k-12 education. *Center for Evaluation and Education Policy, Indiana University*.

Plucker, J. A. and S. J. Peters (2020). *Excellence gaps in education: Expanding opportunities for talented students*. Harvard Education Press.

Plucker, J. A., S. J. Peters, and S. Schmalensee (2017). Reducing excellence gaps: A research-based model. *Gifted Child Today 40*(4), 245–250.

Rambo-Hernandez, K. E., S. J. Peters, and J. A. Plucker (2019). Quantifying and exploring elementary school excellence gaps across schools and time. *Journal of Advanced Academics 30*(4), 383–415.

Reardon, S. F., D. Kalogrides, and K. Shores (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology 124*(4), 1164–1221.

Turner, E. O. and A. K. Spain (2020). The multiple meanings of (in) equity: Remaking school district tracking policy in an era of budget cuts and accountability. *Urban Education 55*(5), 783–812.

White, R. and B. DeGrow (2016). *A Survey of Michigan's Private Education Sector. A Mackinac Center Report.* ERIC.

Williams, B. (1996). *Closing the achievement gap: A vision for changing beliefs and practices.* ASCD.

Wyner, J. S., J. M. Bridgeland, and J. J. DiIulio Jr (2007). Achievementrap: How america is failing millions of high-achieving students from lower-income families. *Civic Enterprises*.

Appendix

A Summary Statistics by Race, Gender, and SES **36**

B The Excellence Gap **37**
B.1 Measuring Economically Disadvantaged Status 43

C Achievement at other points of the distribution **44**

D Excellence Gap Across Schools **46**

E Racial Excellence Gaps and Gender **51**

F Accountability Policy **52**

G Persistence in the Excellence Gap **58**
G.1 Measurement error 58
G.2 College readiness 60

H College Readiness **64**

I Alternative Definitions of the Excellence Gap **68**

A Summary Statistics by Race, Gender, and SES

Table A1: Summary Statistics

	Panel A: Main Sample									
		By Race					By Gender		By SES	
	All	White	Black	Hispanic	Asian	Other	Male	Female	Not Econ Disadv.	Econ Disadv.
% of data		71.7	15.3	6.5	2.8	3.7	49.7	50.3	55.6	44.4
Math Percentile Gr. 3	50.5	54.7	33.0	40.5	67.2	47.4	52.3	48.7	58.6	40.4
Math Percentile Gr. 8	50.5	54.8	31.0	40.5	75.1	46.4	50.1	50.8	59.9	38.7
Rdg Percentile Gr. 3	50.5	54.2	36.3	40.2	59.2	49.2	48.7	52.3	58.0	41.1
Rdg Percentile Gr. 8	50.5	54.0	34.4	42.4	69.8	48.1	47.0	53.8	58.7	40.1
N	534122	383222	81871	34588	14763	19678	265513	268609	297231	236891
	Panel B: High School Sample									
		By Race					By Gender		By SES	
	All	White	Black	Hispanic	Asian	Other	Male	Female	Not Econ Disadv.	Econ Disadv.
% of data		73.5	14.6	5.9	2.7	3.3	48.8	51.2	58.0	42.0
Math Percentile Gr. 3	51.9	55.8	34.2	41.6	66.9	49.0	54.3	49.5	59.0	42.0
Math Percentile Gr. 8	52.2	56.2	32.6	42.3	74.8	48.6	53.0	51.5	60.4	40.9
Rdg Percentile Gr. 3	51.8	55.0	38.6	41.7	59.6	50.4	50.4	53.1	58.3	42.8
Rdg Percentile Gr. 8	52.0	55.2	36.4	43.9	68.7	50.3	49.6	54.3	58.9	42.5
Took SAT	98.1	98.6	95.8	97.8	99.3	96.8	97.9	98.2	99.1	96.7
SAT Pctile (Within Data)	53.1	57.4	31.8	42.1	76.0	49.7	53.3	52.9	62.0	40.5
Top Decile SAT Math (Natl)	6.2	6.6	0.5	1.8	35.3	4.8	7.7	4.7	9.6	1.3
Top Decile SAT Reading (Natl)	7.4	8.3	1.0	2.7	29.6	6.1	7.4	7.3	11.1	2.2
Took AP Course	40.4	43.6	23.0	32.0	74.8	33.8	36.1	44.5	52.1	24.2
No. AP Courses Taken	2.4	2.6	1.0	1.6	7.0	1.9	2.2	2.6	3.3	1.1
No. STEM AP Courses Taken	0.9	0.9	0.3	0.5	3.3	0.7	0.9	0.8	1.2	0.3
N	244750	179870	35854	14323	6510	8193	119441	125309	141941	102809

Notes: This table displays summary statistics for the (a) main sample (b) high school sample, as defined by Section I. "Econ Disadv" includes all students who are classified as economically disadvantaged by Michigan in 8th grade, 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. Math and reading percentiles are calculated from MEAP (for all years until 2014) or M-STEP (for all years after 2014). "Within Data" indicates that the respective SAT or PSAT percentiles were calculated using the data in the sample. "Natl" indicates that the percentile cutoffs were determined by College Board at the national level.

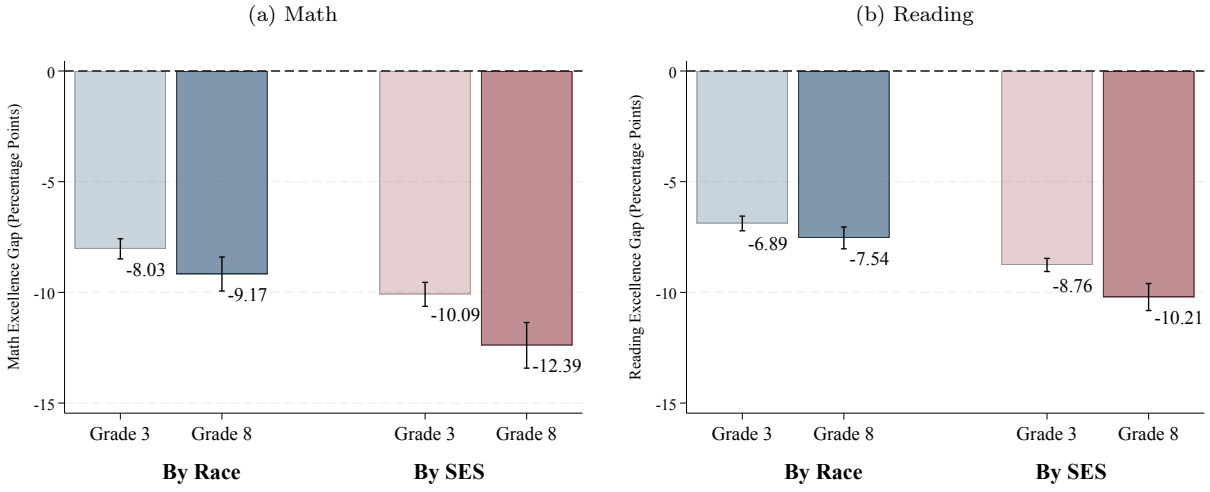
Table A2: Summary Statistics: Fraction categorized as excellent

	Grade					
	3	4	5	6	7	8
Math	9.92	10.06	10.40	10.29	10.11	10.32
Reading	9.63	9.05	9.26	10.11	9.86	9.90

Notes: This table displays summary statistics for the fraction of students who are categorized as being excellent, by subject and grade. The fractions are not exactly 10 because of ties in scores.

B The Excellence Gap

Figure A1: The Excellence Gap by Race and SES



Notes: These figures display differences across race and SES groups in the likelihood of achieving excellence in (a) math and (b) reading. Excellence is defined as reaching the top decile on the reading MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

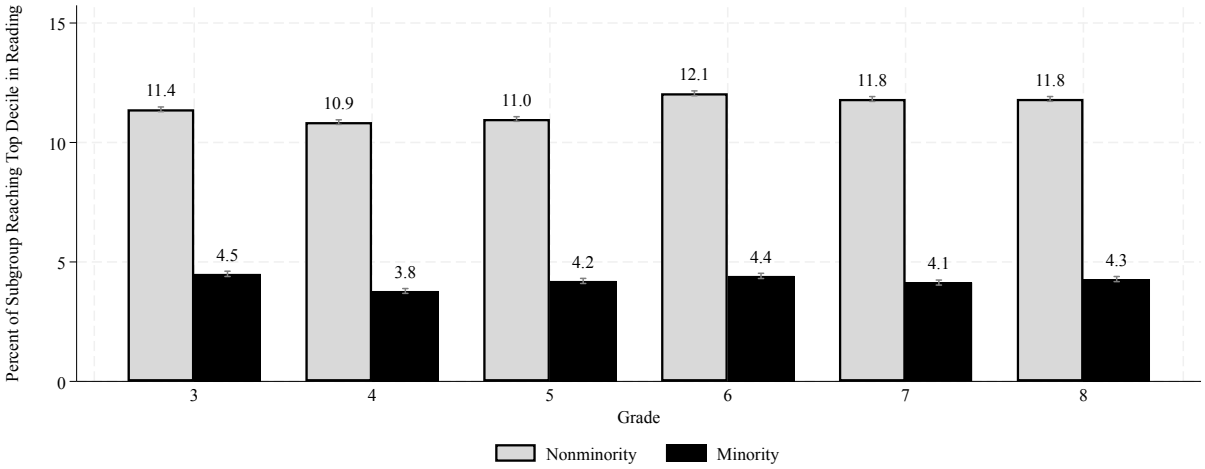
Table A3: Excellence Gap Across Grades

	Math						Reading					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Minority	-8.80***	-7.68***	-8.03***				-7.27***	-6.33***	-6.89***			
	(0.23)	(0.55)	(0.23)				(0.14)	(0.34)	(0.17)			
Grade		0.11						0.18**				
		(0.13)						(0.08)				
Minority × Grade		-0.20*						-0.17**				
		(0.11)						(0.07)				
Minority × Grade 3				-8.03***						-6.89***		
				(0.23)						(0.17)		
Minority × Grade 4			-0.60***	-8.63***	-8.43***	-7.86***			-0.18	-7.07***	-6.89***	-6.47***
			(0.15)	(0.26)	(0.26)	(0.29)			(0.12)	(0.15)	(0.15)	(0.17)
Minority × Grade 5			-0.81***	-8.84***	-8.77***	-8.24***			0.11	-6.78***	-6.71***	-6.33***
			(0.27)	(0.31)	(0.31)	(0.33)			(0.17)	(0.18)	(0.18)	(0.19)
Minority × Grade 6			-1.02**	-9.06***	-9.17***	-8.90***			-0.76***	-7.64***	-7.74***	-7.54***
			(0.44)	(0.39)	(0.39)	(0.40)			(0.27)	(0.23)	(0.23)	(0.23)
Minority × Grade 7			-1.01**	-9.04***	-8.94***	-8.44***			-0.80***	-7.68***	-7.59***	-7.23***
			(0.44)	(0.39)	(0.40)	(0.49)			(0.29)	(0.25)	(0.25)	(0.29)
Minority × Grade 8			-1.14**	-9.17***	-9.00***	-8.38***			-0.65**	-7.54***	-7.39***	-6.94***
			(0.45)	(0.39)	(0.40)	(0.52)			(0.29)	(0.25)	(0.25)	(0.30)
Same school as prior year					1.97***	2.18***					1.74***	1.89***
					(0.22)	(0.29)					(0.13)	(0.17)
Minority × Same school as prior						-0.73***						-0.54***
						(0.26)						(0.16)
Adjusted R^2	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01

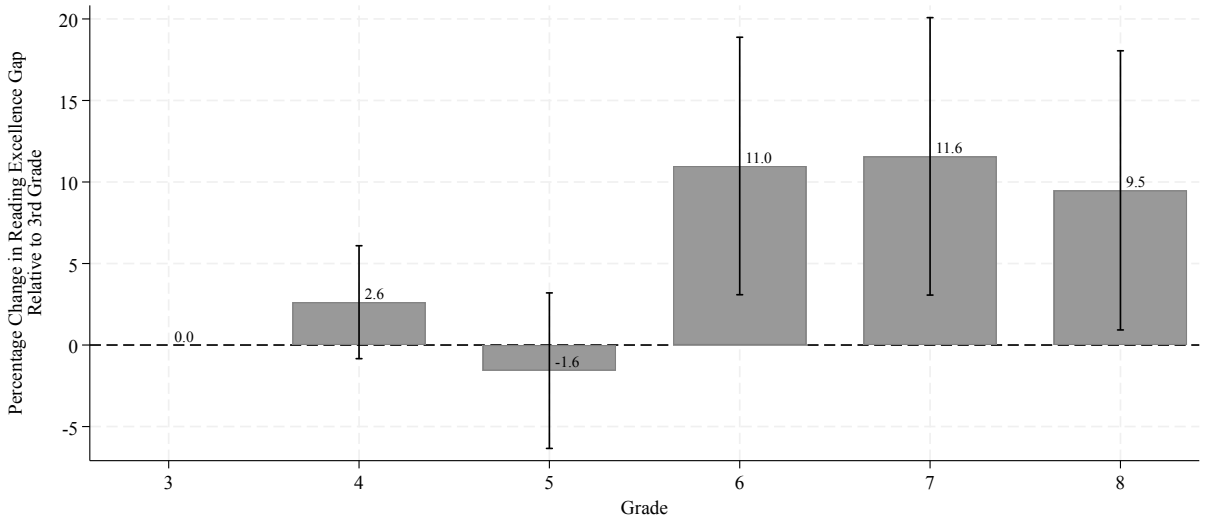
Notes: This table shows the difference in the percentage of minority and non-minority groups that reach excellence. Dependent variable is an indicator for whether the student reaches excellence, taking on a value of 100 if so and 0 otherwise. Regressions including a covariate for whether the student was in the same school as the prior year start from grade 4. Excellence is defined as reaching the top decile on the math or reading (given by the header) section of MEAP (for all years until 2014) or M-STEP (for all years after 2014). Standard errors in parentheses, clustered at the school-level. * $p < .1$, ** $p < .05$, *** $p < .01$

Figure A2: The Excellence Gap in Reading by Grade

(a) Likelihoods of Reaching Excellence



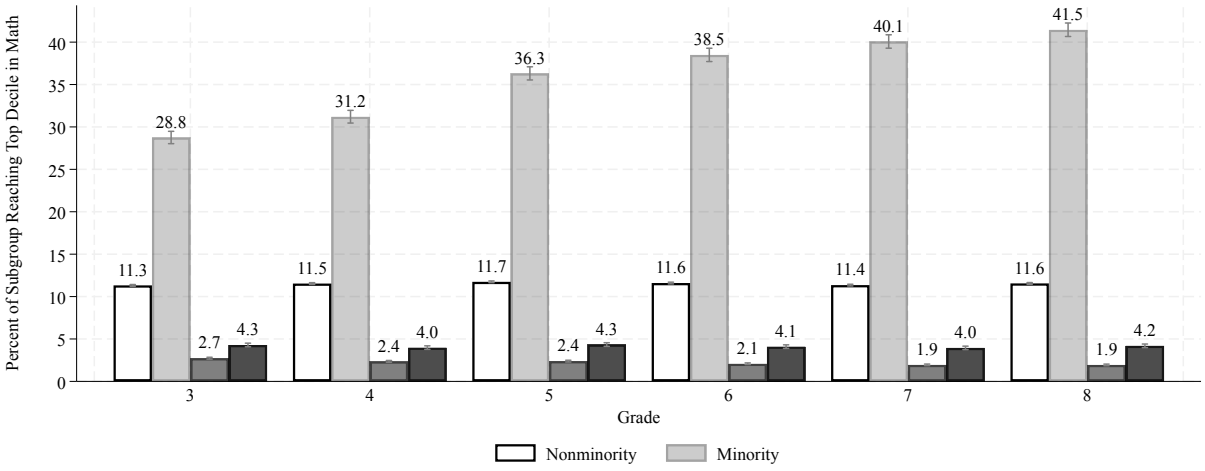
(b) Percentage Change in Excellence Gap from 3rd Grade



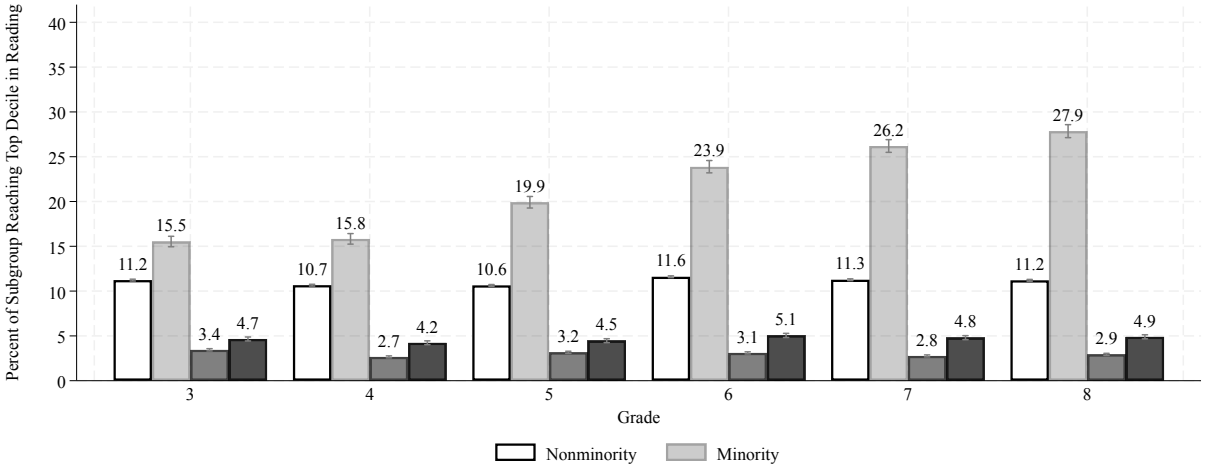
Notes: These figures display (a) the likelihood that a minority (Black or Hispanic) student or non-minority (White or Asian) student reaches excellence, and (b) the percentage change in this difference between minority status in reaching excellence, relative to the difference in the 3rd grade. Excellence is defined as reaching the top decile on the reading MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

Figure A3: Fraction of Students Reaching the Top Decile

(a) Math

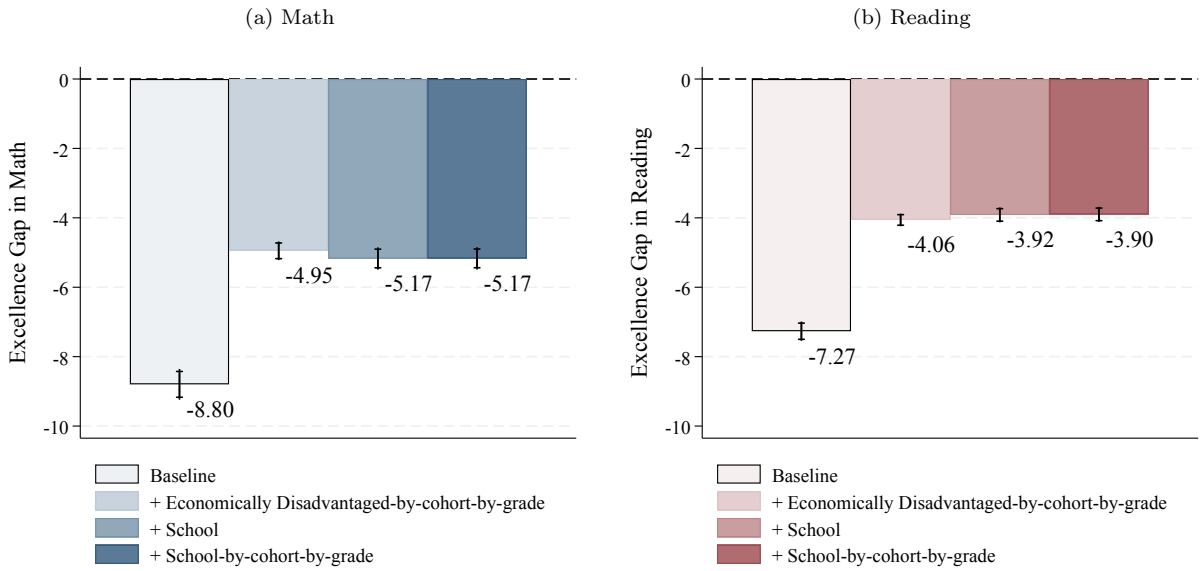


(b) Reading



Notes: These figures display the fraction of students that reach the top decile on the MEAP (for all years until 2014) or M-STEP (for all years after 2014) (a) math section and (b) reading section, by each racial group. Analysis uses the "main sample", as defined in Section I.

Figure A4: Conditional Racial Excellence Gap

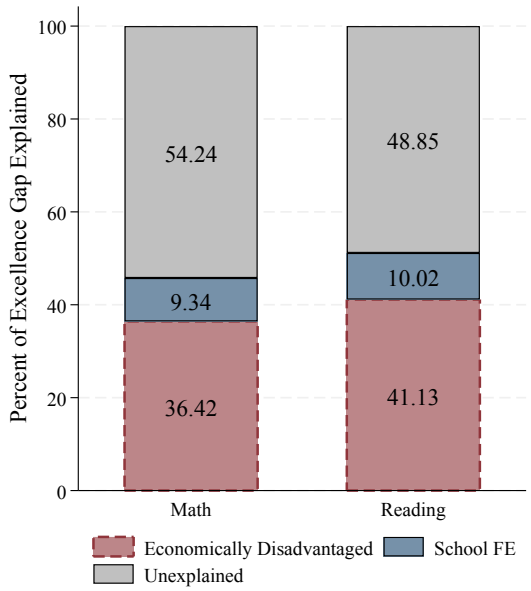


Notes: These figures display differences across minority and non-minority students in the rates of reaching excellence, conditional on controls described by the figure legend. Excellence is defined as reaching the top decile on the reading MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

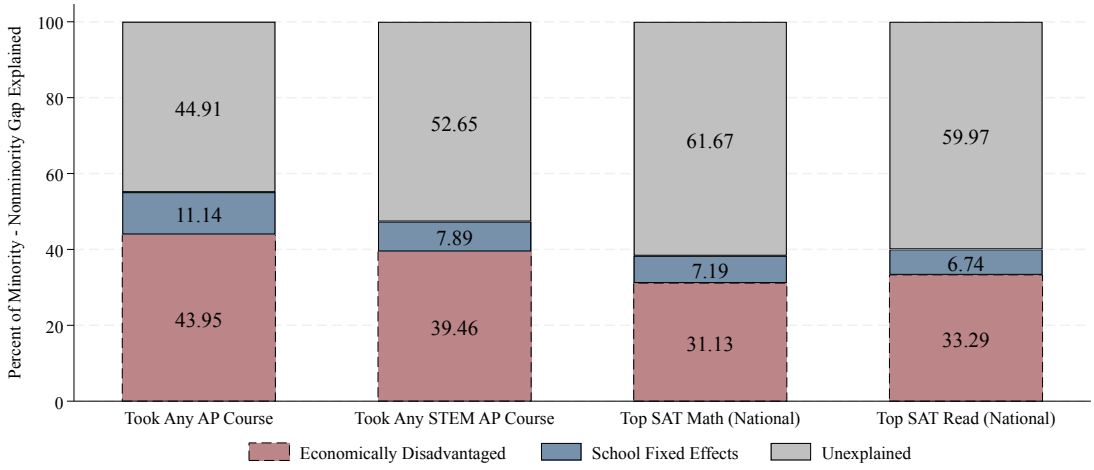
B.1 Measuring Economically Disadvantaged Status

Figure A5: Explaining Excellence Gaps with Number of Times Economically Disadvantaged

(a) Grade 3-8



(b) College Readiness

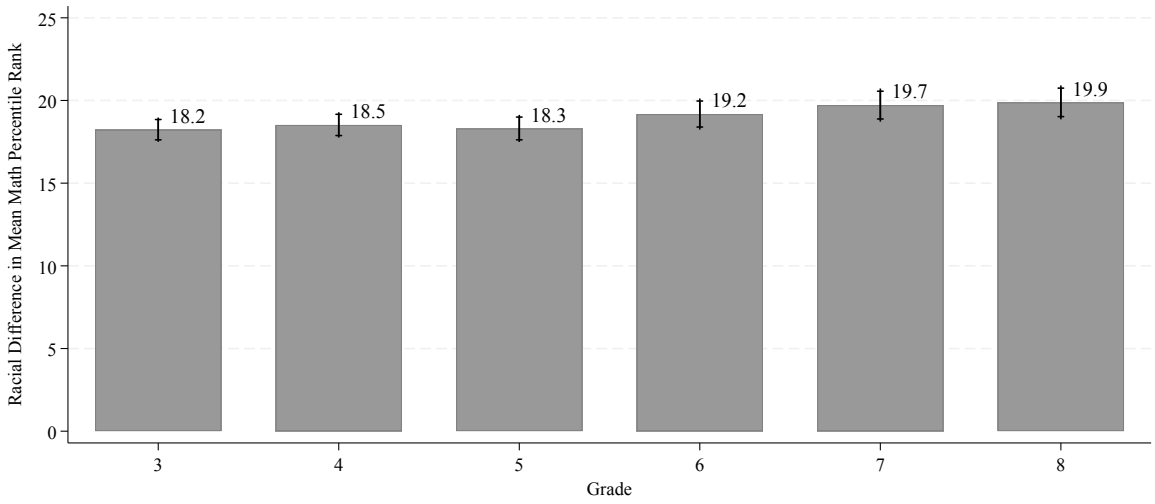


Notes: These figures display the Gelbach decomposition of each minority and non-minority difference in (a) reaching the top decile on the MEAP (for all years until 2014) or M-STEP (for all years after 2014), and (b) reaching various college-readiness outcomes. (a) averages over students in grades 3 through 8. "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 households receiving food or cash assistance, is homeless, is migrant, or is in foster care. In this figure, both the total number of times that a student is determined as economically disadvantaged and the total number of times that this information is available for the student are interacted with grade and cohort following Michelmore and Dynarski (2017). "School FE" indicates the inclusion of school fixed effects. (a) uses the "main sample" and (b) the "high school sample", as defined in Section I.

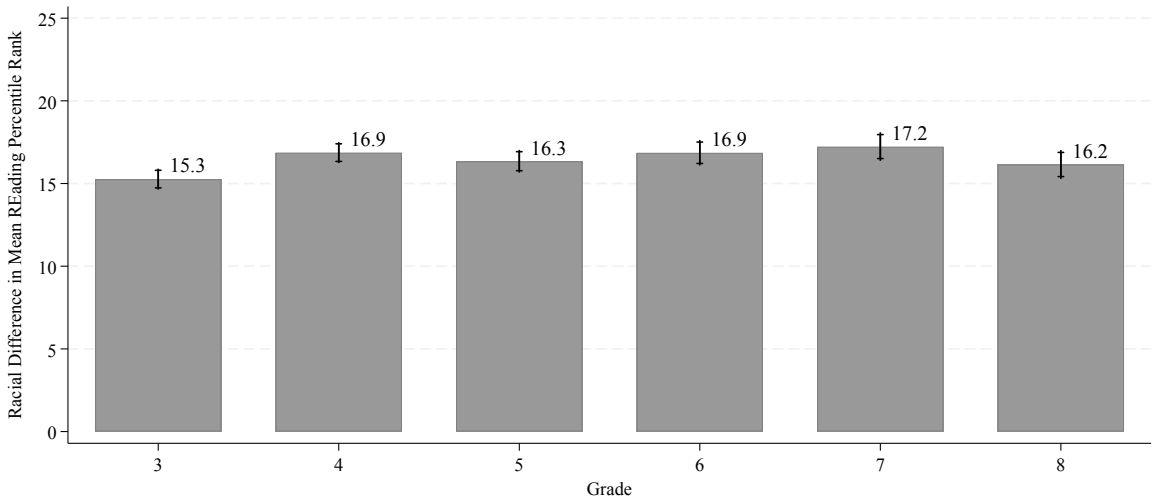
C Achievement at other points of the distribution

Figure A6: The Achievement Gap

(a) Math



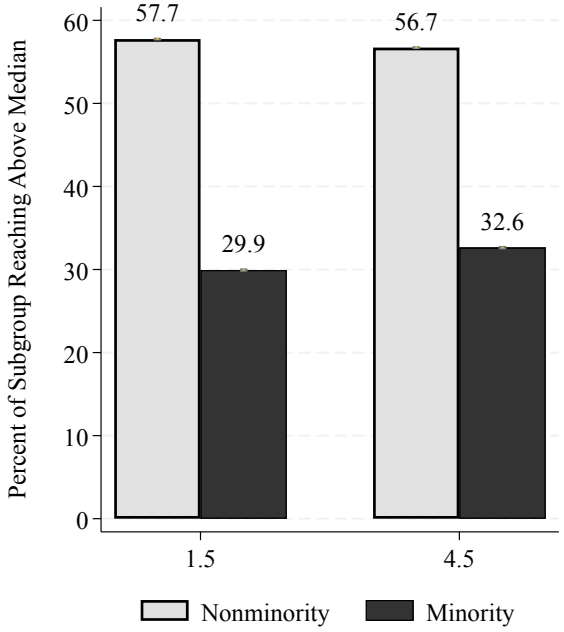
(b) Reading



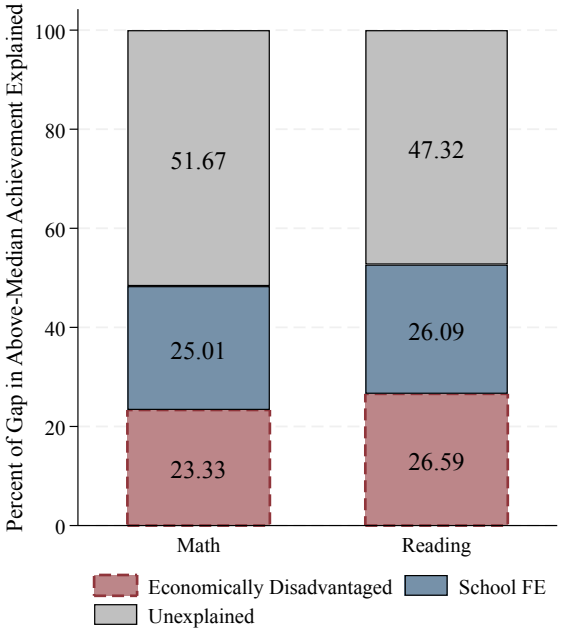
Notes: This figure displays the difference in the average (a) math and (b) reading percentile among minorities and non-minority students from 3rd through 8th grade, using the MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

Figure A7: Gaps in Achievement at the Median and Bottom 10%

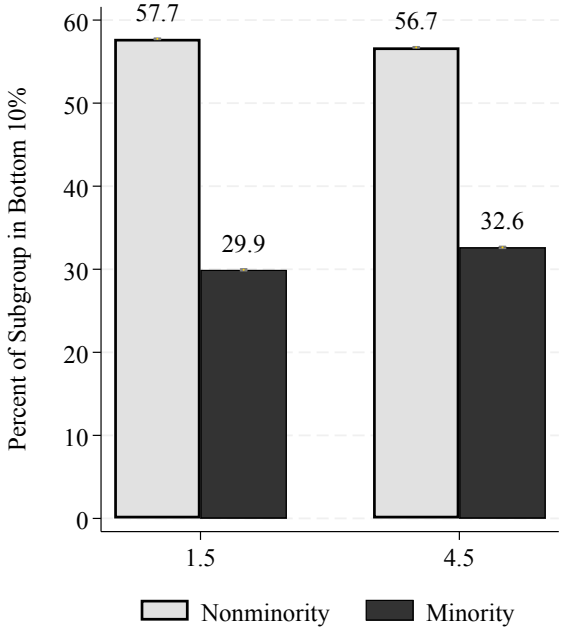
(a) Top 50%: Baseline Levels



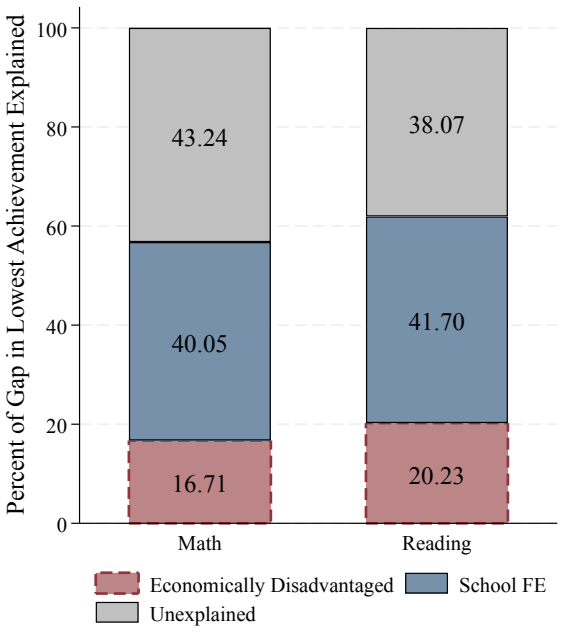
(b) Top 50%: Decomposition



(c) Bottom 10%: Baseline Levels



(d) Bottom 10%: Decomposition



Notes: These figures display the percentage of minority and non-minority groups that reach the (a) top 50 percentile and (b) bottom 10 percent on the MEAP (for all years until 2014) or M-STEP (for all years after 2014), and (c,d) the Gelbach decomposition of each minority and non-minority difference. These measures average over students in grades 3 through 8. "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 households receiving food or cash assistance, is homeless, is migrant, or is in foster care. This indicator is interacted with grade and cohort. "School FE" indicates the inclusion of school fixed effects. Analysis uses the "main sample", as defined in Section I.

D Excellence Gap Across Schools

In the main text, we show empirical measures for school-specific estimates of the excellence gap. We construct this measure by first calculating excellence gaps for each school-year observation in the data, meaning for each school s and year y:

$$\hat{\theta}_{s,y} = \frac{1}{|N_{s,y}|} \sum_{j \in N_{s,y}} \mathbb{1}\{Excellence\}_j - \frac{1}{|M_{s,y}|} \sum_{j \in M_{s,y}} \mathbb{1}\{Excellence\}_j \tag{A1}$$

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 then aggregate these estimates to the school-level:

$$\hat{\theta}_s = \frac{1}{|R_s|} \sum_{y \in R_s} \hat{\theta}_{s,y} \tag{A2}$$

where R_s is the set of years that school s is observed in our data. We estimate the school-specific excellence gaps, $\hat{\theta}_s$, for math and reading. Figure 4 shows the distribution of these gaps. We first test for heterogeneity using a Pearson Chi-squared test: our test statistic is $\sum_s \frac{(\hat{\theta}_s - \bar{\theta})^2}{e_s^2}$ where $\bar{\theta} = \sum_s \hat{\theta}_s$ and e_s is the standard error associated with our school-specific estimate of the excellence gap, $\hat{\theta}_s$. This test is under the assumption that the distribution of each school's excellence gap is normally distributed, centered around the across-school population mean, and has a variance of e_s^2. We reject the null hypothesis that the excellence gap is homogeneous across schools in both math and reading ($p < 0.001$ for math and reading).

Given our use of comprehensive administrative data, these estimates face minimal sampling error for the given schools and years of interest. We find that there is a high correlation in the estimated school-level excellence gap across years, particularly for math (Figure A8).

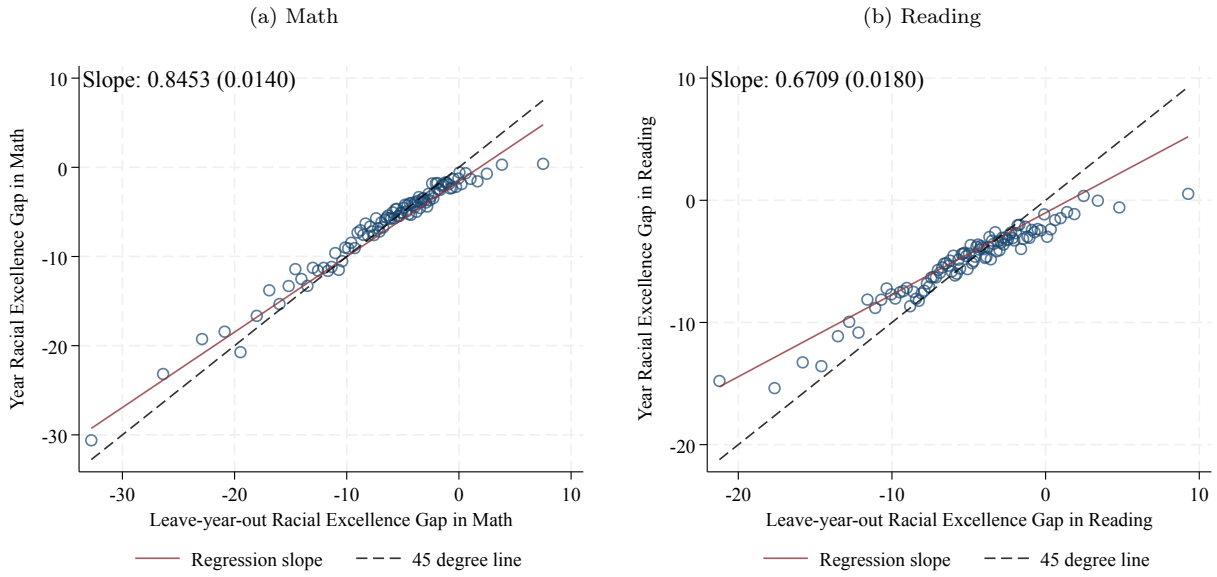
Another set of parameters we can estimate are school-specific latent excellence gaps, where we view a school's yearly excellence gaps as "noisy" estimates of its latent parameter. Let θ_s represent the latent parameter for to school s. Figure 4 reports the standard deviation of $\hat{\theta}_s$, which describes the spread of the realized school-level excellence gaps. We may additionally ask, however, what the variation of the school-specific latent excellence gap, θ_s, is. The standard deviation of $\hat{\theta}_s$ is an over-estimate of the standard deviation of θ_s due to variability in the student sample observed across years. As Kline et al. (2022) show, the variance for this parameter can be estimated using a bias-corrected estimator that adjusts the empirical variance estimate (in our case, the variance of $\hat{\theta}_s$) using the standard error of the estimate (in our case, the standard error of $\hat{\theta}_s$). Following the steps of Kline et al. (2022), we use the estimator:

$$\hat{\sigma}_\theta^2 = \left(\frac{S-1}{S}\right)\left[\frac{1}{S-1}\sum_{s=1}^{S}(\hat{\theta}_f - \bar{\theta})^2 - \frac{1}{S}\sum_{s=1}^{S} e_s^2\right] \tag{A3}$$

where the first component in the bracket corresponds to the variance in $\hat{\theta}_s$ that we report in Figure 4, and the second one represents the bias-correction. Table A5 summarizes the

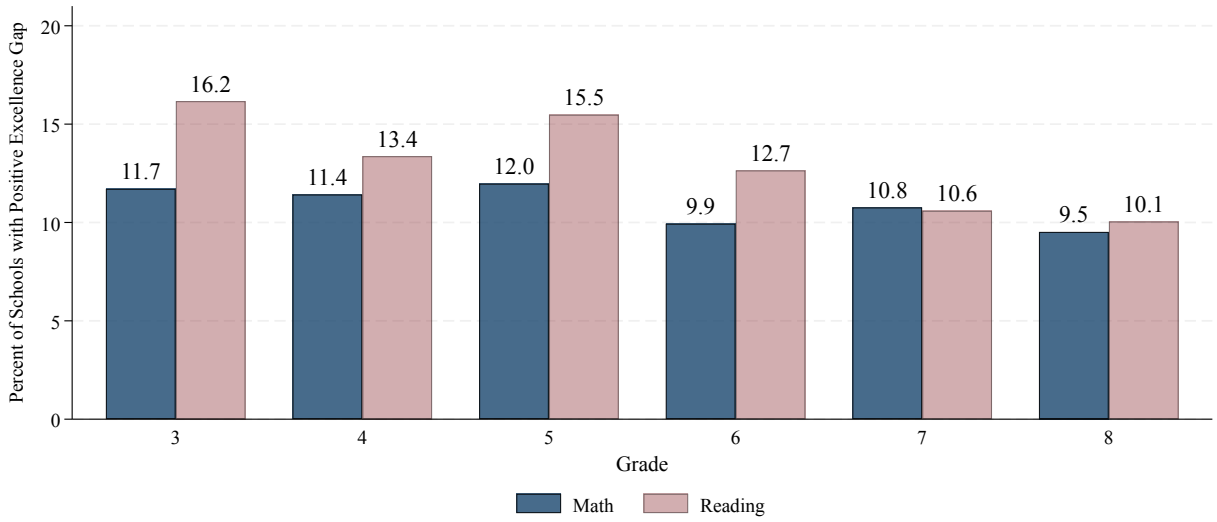
standard deviations (the second row takes the square root of the estimate $\hat{\sigma}^2_\theta$). We find that with respect to the distribution of school-level latent excellence gaps, a school with a one standard deviation larger latent excellence gap corresponds to an increase in around a 5.56 (3.51) percentage points in math (reading).

Figure A8: Temporal Stability of Observed School Excellence Gaps



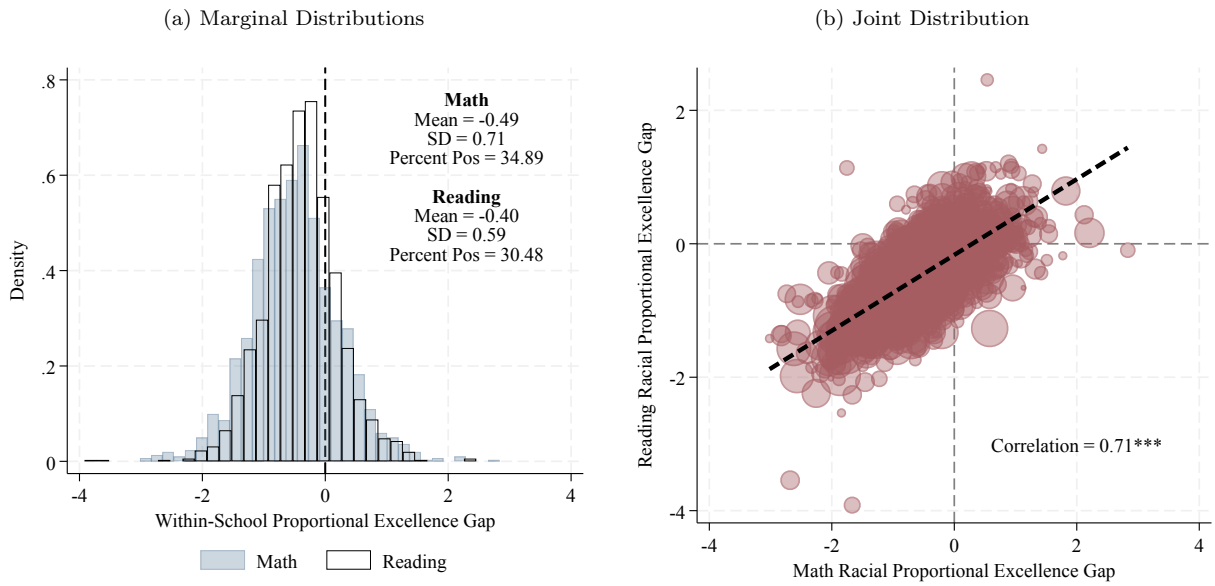
Notes: These figures display the across-year correlation in a school's racial excellence gap in (a) math and (b) reading, in the spirit of Kline et al. (2022). A unit of observation is school-year. Scatter plots correspond to binned averages, with a point for each percentile over the support of the x-axis. Regression slopes are weighted by the number of enrolled students in that year. Standard errors in parentheses, clustered at the school-level. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

Figure A9: Percentage of Schools with Positive Excellence Gap



Notes: This figure displays the percentage of schools at each grade-level that have a positive excellence gap, where positive values mean that minority students are more likely than non-minority students to reach excellence. Analysis uses the "main sample", as defined in Section I.

Figure A10: Proportional Excellence Gaps Across Schools



Notes: These figures display the distribution of proportional excellence gaps. The proportion is calculated by dividing the likelihood that a minority student achieves excellence by the likelihood that a non-minority student does, and then taking the log: $\log\{\mathbb{P}(\text{Top Decile}_i|\text{Minority}_i) - \mathbb{P}(\text{Top Decile}_i|\text{Non-minority}_i)\}$. Figure (a) shows the distribution of school-level proportional excellence gap in math and reading, and (b) the relationship between the two subjects. A unit of observation is a school. These measures pool students across years. The linear fit in panel (b) weights schools by the number of students who take a standardized exam in the school. "Percent Pos" refers to the percentage of schools that have a non-negative excellence gap in the corresponding subject. Analysis uses the "main sample", as defined in Section I.

Table A4: Predicting Positive Excellence Gaps Across Schools

	Positive Math Excellence Gap						Positive Reading Excellence Gap					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Frac. Minority	-0.05					-0.11***	-0.07**					-0.12***
	(0.03)					(0.04)	(0.03)					(0.04)
Frac. Econ Disad.		0.04*				0.02		-0.00				-0.07
		(0.02)				(0.04)		(0.03)				(0.05)
Average Math Score (std)			-0.10***			-0.08***			-0.02			-0.01
			(0.03)			(0.03)			(0.03)			(0.03)
Average Reading Score (std)			0.08**			0.03			-0.00			-0.10**
			(0.04)			(0.04)			(0.04)			(0.05)
N Test-takers (log)				0.00		0.00				0.01		-0.00
				(0.01)		(0.01)				(0.01)		(0.01)
Location: Suburb					0.00	0.00					0.01	0.00
					(0.01)	(0.01)					(0.01)	(0.01)
Location: Rural					0.02*	0.01					0.01	-0.00
					(0.01)	(0.01)					(0.01)	(0.01)
N	1975	1975	1975	1975	1975	1975	1975	1975	1975	1975	1975	1975
Adjusted R^2	0.10	0.10	0.11	0.10	0.10	0.12	0.11	0.11	0.11	0.11	0.11	0.12

Notes: This table correlates whether a school has a positive excellence gap with its observable characteristics. Positive excellence gap corresponds to minorities exhibiting a higher rate of reaching excellence than non-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. Standard errors in parentheses, clustered at school level. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A5: Across-School Variation in Excellence Gaps

	Math	Reading
$\hat{\theta}_s$	6.46	4.93
θ_s	5.56	3.51

Notes: This table shows the estimated standard deviations in the distribution of across-school excellence gaps in math and reading. The first row shows the standard deviations of excellence gaps that are observed in the data, $\hat{\theta}_s$, as illustrated in Figure 4. The second row shows the standard deviation of a school's latent excellence gap, θ_s. The latter standard deviation is estimated by following the bias-correction steps of Kline et al. (2022) and as described in Section D. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

E Racial Excellence Gaps and Gender

Table A6: 8th Grade Excellence by Gender

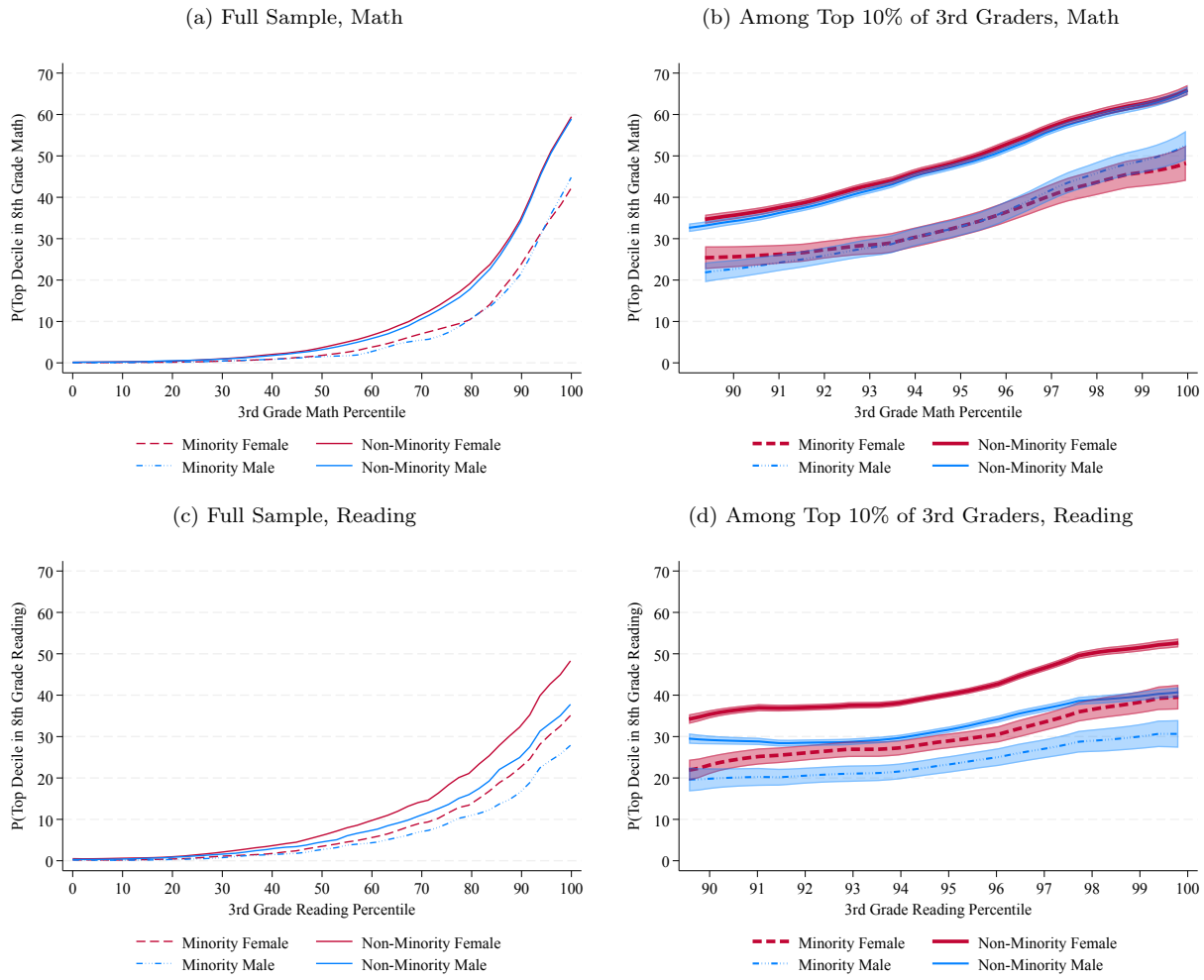
	Math						Reading					
	All		Top in 3		Not Top in 3		All		Top in 3		Not Top in 3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Minority	-9.86***	-1.93***	-14.69***	-13.28***	-5.53***	-1.56***	-6.26***	0.28*	-6.68***	-5.64***	-3.78***	0.04
	(0.42)	(0.19)	(1.24)	(1.19)	(0.24)	(0.14)	(0.23)	(0.14)	(1.02)	(0.99)	(0.14)	(0.11)
Female	-1.71***	-0.15	0.46	1.24**	-0.25**	0.35***	4.41***	5.69***	14.09***	14.67***	4.34***	4.92***
	(0.13)	(0.11)	(0.52)	(0.50)	(0.10)	(0.10)	(0.14)	(0.15)	(0.45)	(0.45)	(0.13)	(0.14)
Minority × Female	1.39***	0.66***	-1.11	-1.36	0.34**	0.08	-2.66***	-3.26***	-3.62***	-3.81***	-2.62***	-2.87***
	(0.16)	(0.15)	(1.44)	(1.41)	(0.14)	(0.14)	(0.18)	(0.17)	(1.34)	(1.34)	(0.17)	(0.16)
3rd grade Percentile		0.42***		3.70***		0.25***		0.34***		2.74***		0.24***
		(0.01)		(0.11)		(0.01)		(0.01)		(0.10)		(0.00)
Adjusted R^2	0.02	0.16	0.01	0.05	0.01	0.08	0.02	0.12	0.02	0.05	0.01	0.07
Equivalence level (Minority × Female)	1.65	0.91	3.47	3.68	0.57	0.30	2.95	3.55	5.82	6.01	2.89	3.14
Equivalence level (Female + Minority × Female)	0.49	0.69	2.93	2.38	0.24	0.58	1.95	2.64	12.53	12.93	1.90	2.23
pval Female + Minority × Female	0.00	0.00	0.64	0.93	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: This table shows the rates of reaching excellence in 8th grade math and reading, by gender, Minority status, and 3rd grade achievement. Dependent variable is whether a student reaches excellence in 8th grade, taking on value 100 if so and 0 otherwise, in math for the first 6 columns and reading for the subsequent columns. Within each subject, the first two columns use the full sample, the third and fourth subset to those who reached excellence in 3rd grade, and the fifth and sixth subset to those who did not reach excellence in 3rd grade. Standard errors in parentheses, clustered at school level. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Equivalence levels correspond to the highest possible equivalence test level (rounded to the nearest 0.01) that rejects the null hypothesis that the coefficient of interest is significantly different from zero at the 95% confidence level with an equivalence test using Dinno (2023). "pval Female + Minority × Female " shows the p-value associated with the test that the sum of the coefficient on Female and Minority × Female are equal to zero: in other words, that the gender difference in excellence among racial minorities is zero. Analysis uses the "main sample", as defined in Section I. * $p < .1$, ** $p < .05$, *** $p < .01$

Diving deeper into gender differences in the dynamics of racial excellence gaps, we find distinct patterns across math and reading. For math, we find that rates of persistence in excellence are similar across genders (Figure A11a and A11b). In other words, a female and male student with the same 3rd grade achievement on average have very similar rates of achieving excellence in 8th grade. This suggests that the notion that females exhibit a greater widening in the racial excellence gap than boys is driven by gender differences in baseline math achievement rather than differential rates of entry and exit across the distribution of 3rd grade achievement.

In contrast, for reading, the levels of persistence differ across the genders. A female and male student with the same 3rd grade achievement have substantially different rates of achieving excellence in 8th grade. Among those who scored in the top 10% of reading in 3rd grade, the racial excellence gap for boys is around 4 percentage points larger than girls, even after controlling for 3rd grade achievement.

Figure A11: 8th grade Excellence by 3rd Grade Rank, by Gender



Notes: These figures show the proportion of each minority-by-gender group that reach the top decile of 8th grade achievement by 3rd grade achievement. Figures (a) and (b) show for math achievement, and figures (c) and (d) for reading. Figure (a) and (c) show for each 3rd grade decile group, while (b) and (d) show for each percentile in the top 10 percent. Analysis uses the "main sample", as defined in Section I.

F Accountability Policy

We construct the Composite Gap Index cutoff following the steps outlined by the Michigan Department of Education in the 2011-2012 "Top-to-Bottom Ranking, Priority, Focus and Rewards Schools Identification Business Rules" document. We get the total number Title I schools in Michigan (2775), a value used when computing the Composite Gap Index cutoff, from publicly available data by the National Center for Education Statistics.

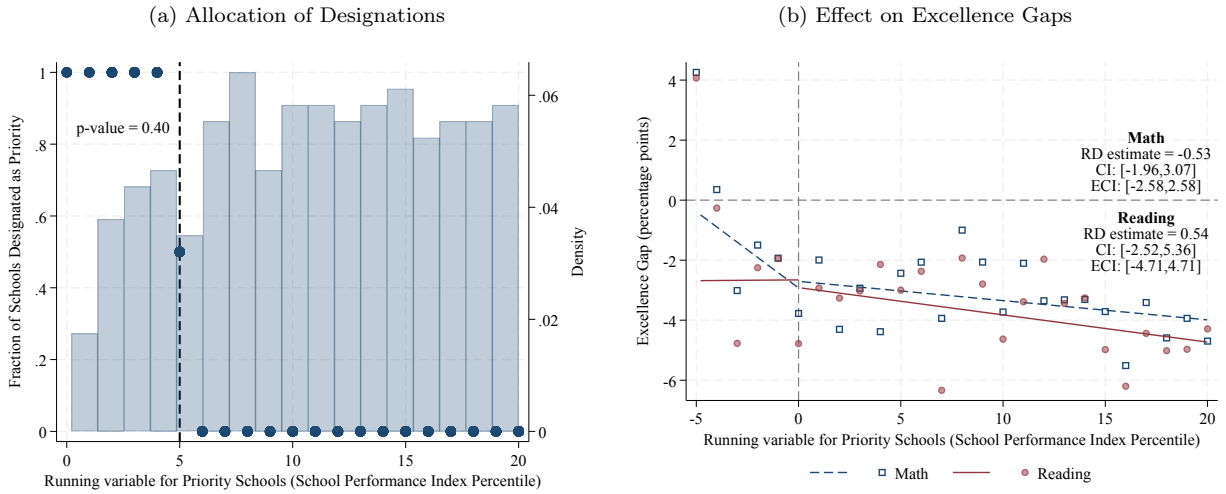
We follow the sample restrictions outlined by Hemelt and Jacob (2020). We exclude schools that serve high school students because the ranking for these schools include graduation rates. Specifically, we remove schools who offer grades higher than 9th grade, and schools whose highest offered grade is below 4th 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" (PLA) or schools that received a School Improvement Grant (SIG) prior to 2012 because these schools did not face the same designation procedures as other schools (Hemelt and Jacob, 2017). Finally, following Hemelt and Jacob (2020), we exclude schools who had fewer than 50 test-takers in the year prior to the policy to remove schools for special populations.

In the regression discontinuity (RD) specification, we pool observations from 1 and 4 years following policy designation. 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. 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 math 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 the RDs that estimate the impacts on the excellence gap using the total number of test-takers in the school in 2011. For the RDs 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.

In the figures below, we use linear fits when considering Priority designations, given the relatively low number of observations that are to the left of the cutoff.

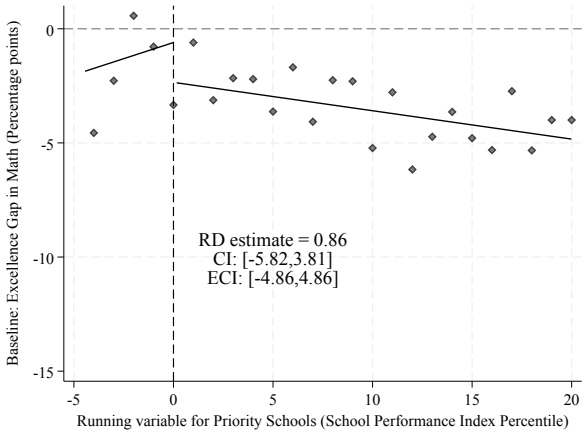
Figure A12: Effect of Priority Designations on Excellence Gaps



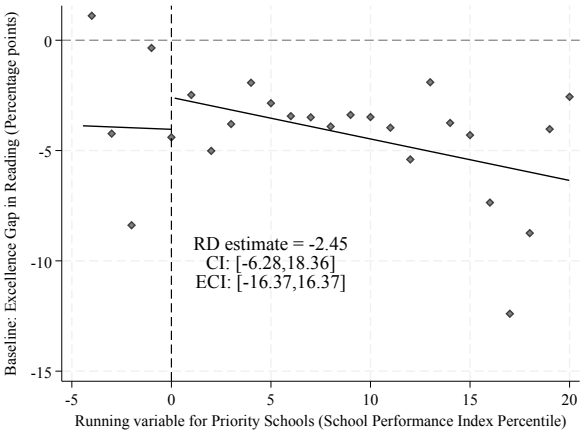
Notes: This figure shows the effects of (a) distribution of the school-level performance index percentile, which is the running variable used to determine Priority designation and (b) the effect of Priority school designations on racial excellence gaps in math and reading. Schools to the right of the Priority cutoff are designated as Priority schools. The scatterplot in panel (a) bins schools by every 1 percentile. The p-value in (a) corresponds to a test of manipulation using local polynomial density estimation by Cattaneo et al. (2020), against the null hypothesis of equality at the cutoff. The lines in (b) represent unweighted linear fits. Positive values of the excellence gap are in favor 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 math MEAP (for all years until 2014) or M-STEP (for all years after 2014). CI are robust bias-corrected 95% confidence intervals (Calonico et al., 2014). ECI are equivalence-test 95% confidence intervals, following Hartman (2021).

Figure A13: Regression Discontinuity Design: Baseline Covariate Balance for Priority

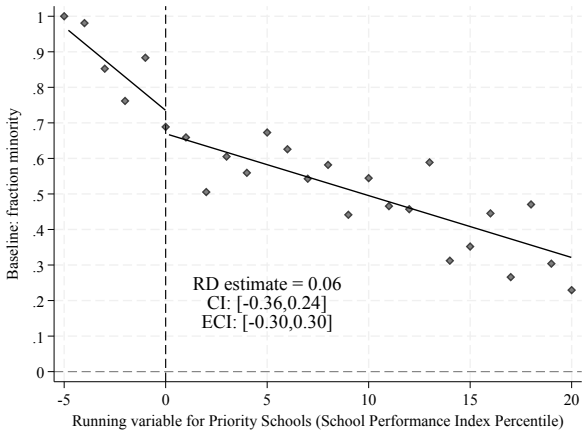
(a) Excellence Gap in Math



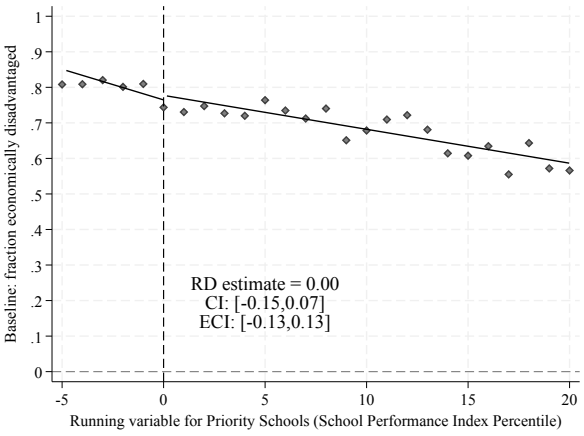
(b) Excellence Gap in Reading



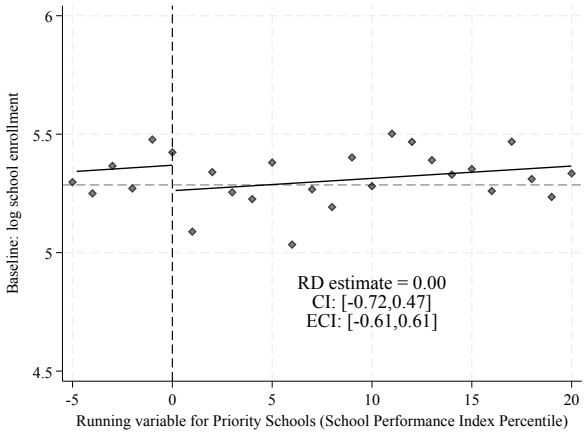
(c) Fraction Minority



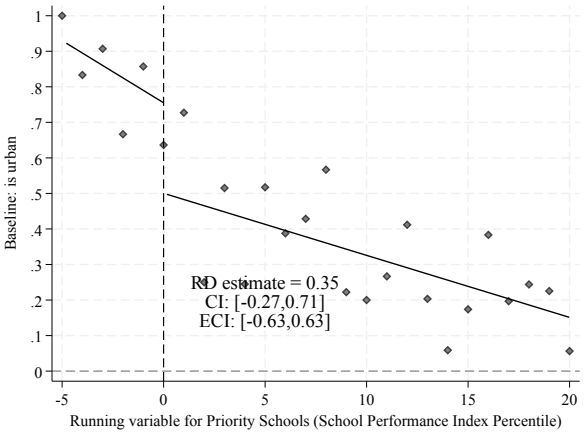
(d) Fraction Economically Disadvantaged



(e) Enrollment



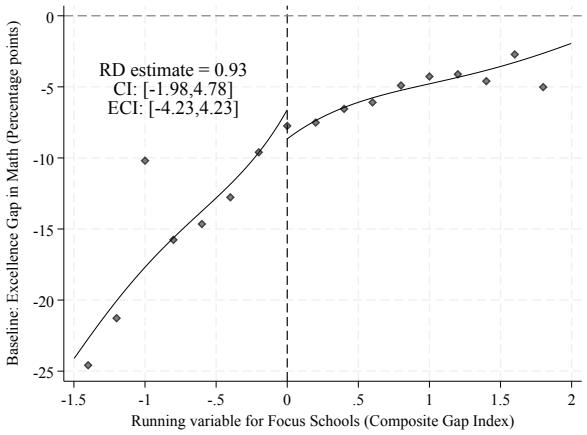
(f) Is Urban



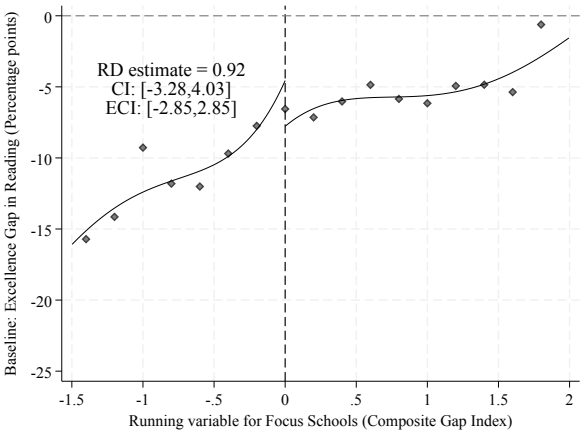
Notes: This figure shows the effects of Priority school policies on pre-determined school-level covariates. The lines represent unweighted linear fits for the presented range. Observations are at the school-year level, and each regression pools observations from the first, second, third, and fourth year after policy designation. Dots bin observations by 1 percentile of the running variable. CI are robust bias-corrected 95% confidence intervals (Calonico et al., 2014). ECI are equivalence-test 95% confidence intervals, following Hartman (2021).

Figure A14: Regression Discontinuity Design: Baseline Covariate Balance for Focus

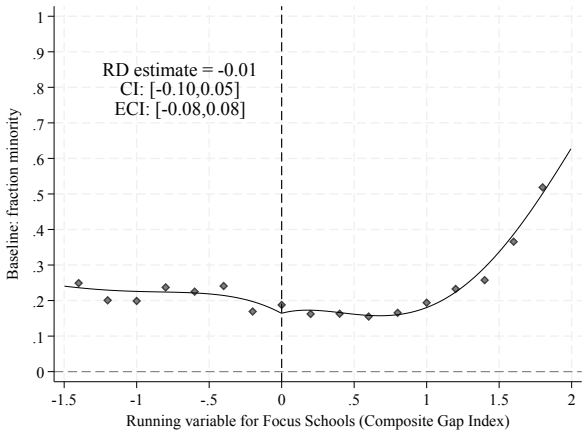
(a) Excellence Gap in Math



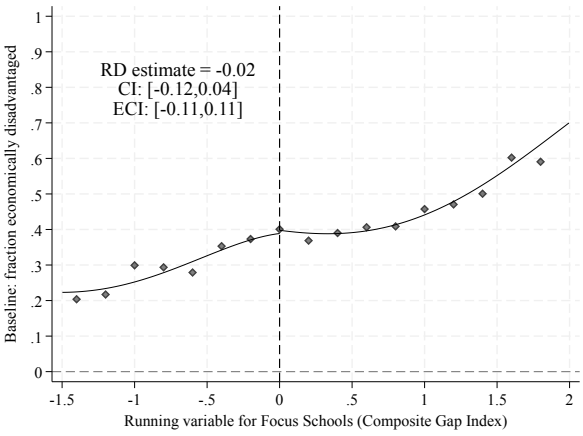
(b) Excellence Gap in Reading



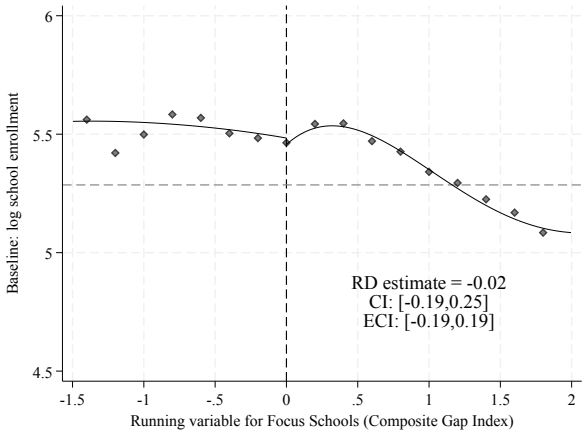
(c) Fraction Minority



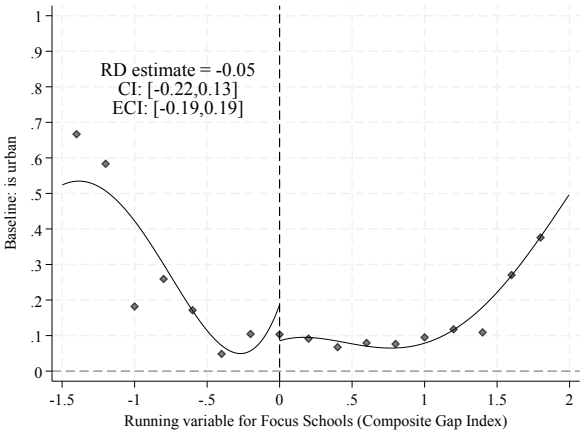
(d) Fraction Economically Disadvantaged



(e) Enrollment



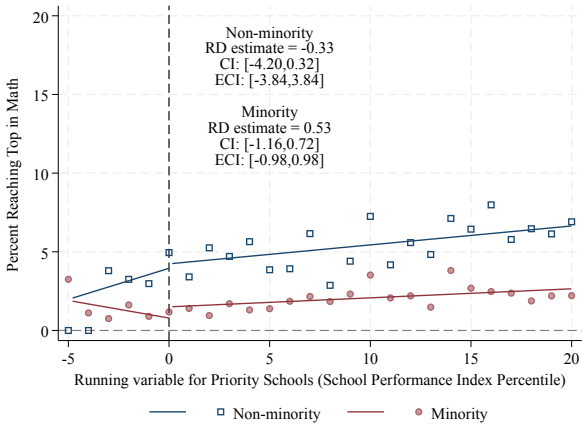
(f) Is Urban



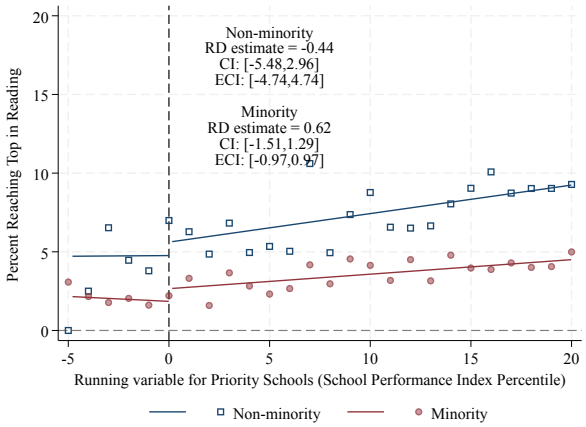
Notes: This figure shows the effects of Focus school policies on pre-determined covariates. Observations are at the school-year level, and each regression pools observations from the first, second, third, and fourth year after policy designation. Dots bin observations by 0.2 units of the running variable. CI are robust bias-corrected 95% confidence intervals (Calonico et al., 2014). ECI are equivalence-test 95% confidence intervals, following Hartman (2021).

Figure A15: Effect of Priority and Focus Designations on Subgroups Reaching Excellence

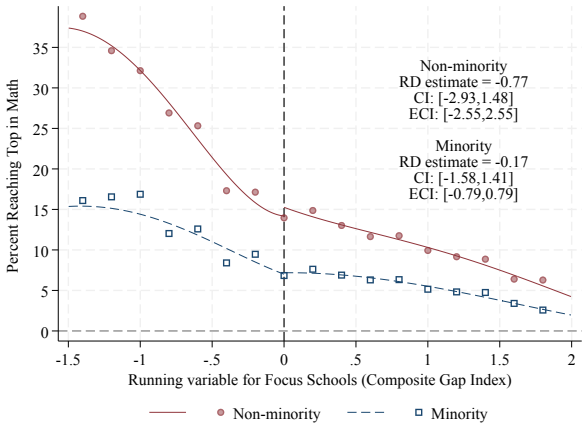
(a) Priority: Top in Math



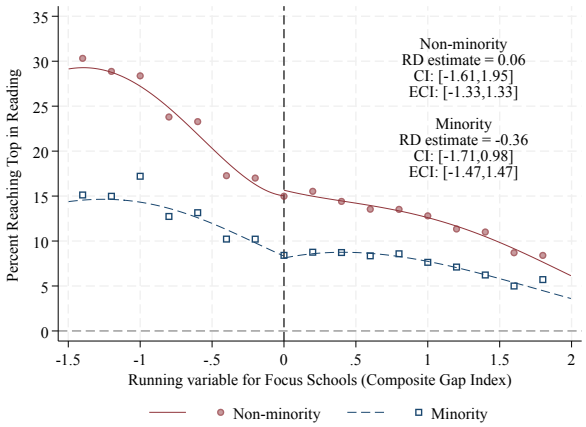
(b) Priority: Top in Reading



(c) Focus: Top in Math



(d) Focus: Top in Reading



Notes: This figure shows the effects of Priority and Focus school policies on rates of reaching excellence by students' minority statuses. For Priority designations (panels a and b), the lines represent unweighted linear fits for the presented range. For Focus, the lines represent local-linear fits following Calonico et al. (2017). Observations are at the school-year level, and each regression pools observations from the first, second, third, and fourth year after policy designation. Dots bin observations by 1 percentile of the running variable. CI are robust bias-corrected 95% confidence intervals (Calonico et al., 2014). ECI are equivalence-test 95% confidence intervals, following Hartman (2021).

Table A7: Specification Details for Accountability Regression Discontinuity

	Panel A: Excellence Gap			
	Priority		Focus	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
RD_Estimate	-0.53	0.54	1.01	0.55
	(1.16)	(1.41)	(0.82)	(0.67)
Robust 90% CI	[-1.56,2.67]	[-1.89,4.73]	[-0.65,2.59]	[-0.87,1.75]
Robust 95% CI	[-1.96,3.07]	[-2.52,5.36]	[-0.96,2.90]	[-1.12,2.00]
Equivalence 90% CI	[-2.06,2.06]	[-3.96,3.96]	[-2.34,2.34]	[-1.51,1.51]
Equivalence 95% CI	[-2.58,2.58]	[-4.71,4.71]	[-2.69,2.69]	[-1.81,1.81]
Kernel Type	Triangular	Triangular	Triangular	Triangular
BW Type	mserd	mserd	mserd	mserd
N	6855	6855	6721	6721
N Clusters (School)	1797	1797	1745	1745
Conventional Std. Error	1.157	1.407	0.817	0.671
Conventional p-value	0.648	0.704	0.216	0.411
Robust p-value	0.666	0.480	0.325	0.582
Order Loc. Poly. (p)	1.000	1.000	1.000	1.000
Order Bias (q)	2.000	2.000	2.000	2.000
BW Loc. Poly. (h)	8.412	7.320	0.502	0.587
BW Bias (b)	9.025	9.694	0.852	1.007

	Panel B: Minority Student's Likelihood of Excellence			
	Priority		Focus	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
RD_Estimate	0.53	0.62	-0.17	-0.36
	(0.35)	(0.60)	(0.64)	(0.60)
Robust 90% CI	[-1.01,0.57]	[-1.29,1.06]	[-1.33,1.17]	[-1.49,0.77]
Robust 95% CI	[-1.16,0.72]	[-1.51,1.29]	[-1.58,1.41]	[-1.71,0.98]
Equivalence 90% CI	[-0.79,0.79]	[-0.47,0.47]	[-0.00,0.00]	[-1.20,1.20]
Equivalence 95% CI	[-0.98,0.98]	[-0.97,0.97]	[-0.73,0.73]	[-1.47,1.47]
Kernel Type	Triangular	Triangular	Triangular	Triangular
BW Type	mserd	mserd	mserd	mserd
N	6880	6880	6734	6734
N Clusters (School)	1797	1797	1745	1745
Conventional Std. Error	0.351	0.603	0.636	0.597
Conventional p-value	0.131	0.306	0.791	0.544
Robust p-value	0.650	0.876	0.915	0.598
Order Loc. Poly. (p)	1.000	1.000	1.000	1.000
Order Bias (q)	2.000	2.000	2.000	2.000
BW Loc. Poly. (h)	7.401	8.046	0.618	0.813
BW Bias (b)	9.454	9.401	1.045	1.481

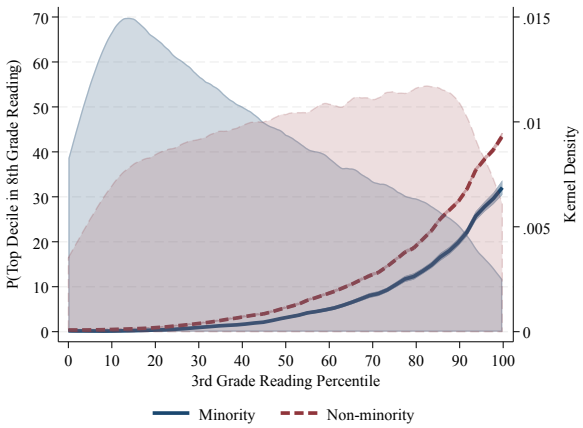
	Panel C: Non-minority Student's Likelihood of Excellence			
	Priority		Focus	
	Math	Reading	Math	Reading
	(1)	(2)	(3)	(4)
RD_Estimate	-0.33	-0.44	-0.77	0.06
	(0.97)	(1.77)	(0.93)	(0.76)
Robust 90% CI	[-3.84,-0.04]	[-4.80,2.28]	[-2.57,1.12]	[-1.33,1.66]
Robust 95% CI	[-4.20,0.32]	[-5.48,2.96]	[-2.93,1.48]	[-1.61,1.95]
Equivalence 90% CI	[-3.42,3.42]	[-3.92,3.92]	[-2.18,2.18]	[-0.64,0.64]
Equivalence 95% CI	[-3.84,3.84]	[-4.74,4.74]	[-2.60,2.60]	[-1.24,1.24]
Kernel Type	Triangular	Triangular	Triangular	Triangular
BW Type	mserd	mserd	mserd	mserd
N	6858	6858	6724	6724
N Clusters (School)	1797	1797	1745	1745
Conventional Std. Error	0.971	1.770	0.934	0.755
Conventional p-value	0.735	0.802	0.413	0.933
Robust p-value	0.092	0.558	0.519	0.855
Order Loc. Poly. (p)	1.000	1.000	1.000	1.000
Order Bias (q)	2.000	2.000	2.000	2.000
BW Loc. Poly. (h)	6.944	7.016	0.570	0.556
BW Bias (b)	9.486	9.571	0.901	0.931

Notes: This table shows the estimated impacts of Michigan's accountability policy on designating schools as Priority and Focus schools. The dependent variables are: (Panel A) a school's excellence gap at a given year, (Panel B) the fraction of minority students at the school who reach excellence, (Panel C) the fraction of non-minority students at the school who reach excellence. Excellence is defined as reaching the top decile on the math or reading (given by the header) section of MEAP (for all years until 2014) or M-STEP (for all years after 2014). Each coefficient is estimated using a fuzzy regression discontinuity following the bias-corrected approach of Calonico et al. (2017). Observations are at the school-year level, and each regression pools observations from either the first and second year after policy designation, or the third and fourth year. Standard errors in parentheses and robust bias-corrected 95% confidence intervals in brackets. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014).

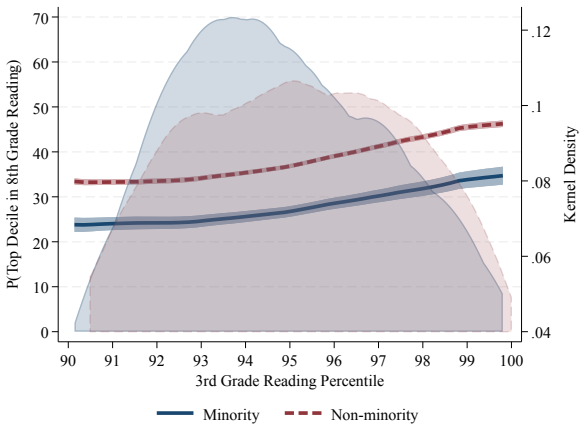
G Persistence in the Excellence Gap

Figure A16: 8th Grade Excellence in Reading by 3rd Grade Rank

(a) Full Sample



(b) Among Top 10% of 3rd Graders



Notes: These figures show the proportion of students who reach the top decile of the 8th grade achievement by 3rd grade achievement, and the distribution of 3rd grade achievement. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Figure (a) shows for each 3rd grade decile group, while (b) shows for each percentile in the top 10 percent. Shaded areas reflect the density of students at each achievement level. Analysis uses the "main sample", as defined in Section I.

G.1 Measurement error

Our results on persistence can be influenced by measurement error in students achievements. Based on the underlying test score distribution by race that we see in the data, one potential partial explanation of differential persistence is regression towards the mean. Some students may fall out of the top decile from early grades to 8th grade because they had a better than average day for themselves in 3rd grade, and worse in 8th. Since the overall distribution of achievement of minority students is to the left of non-minority students, we would expect the average magnitude of the noise among students in the top decile to be larger for minority students.

For instance, suppose μ_i^s denotes the underlying ability of student i in subject s for $s \in \{math, reading\}$, but that test scores are a noisy measure of it:

$$T_{i,t}^s = \alpha_0^s + \alpha^s \mu_i + \epsilon_{i,t}^s \tag{A4}$$

where $T_{i,t}^s$ is student i's test score in grade t in subject s. In our analyses on persistence, we explore the correlation between a student's achievement in two points in time. For example, we consider a linear probability model for the link between fourth grade achievement and whether a student reaches excellence in 8th grade:

$$Y_{i,8}^s = \delta_0 + \delta_1 M_i + \delta_2 T_{i,4}^s + \delta_3 T_{i,4}^s M_i + \nu_{i,t} \tag{A5}$$

where $Y^s_{i,t}$ is an indicator for whether student i reaches excellence in grade t in subject s and M_i is an indicator for whether student i is a racial minority. Given potential measurement error in $T^s_{i,4}$, our estimates may attenuate the true relationships. To address this concern, we explore instrumenting for test scores in two ways, following past work (e.g. Bound et al., 2001; Zabel, 2008; Botelho et al., 2015).

We first consider the case when measurement error is independent across subjects at a given grade:

$$\epsilon^{math}_{i,t} \perp \epsilon^{reading}_{i,t} \ \forall i, t \tag{A6}$$

Under this assumption, and the assumption that underlying skill in math and reading are correlated, we use a student's test score in reading at grade ($T^{reading}_{i,4}$) as an instrument for the student's test score in math ($T^{math}_{i,4}$), and vice versa for reading.

Next, we consider the case when measurement error is independent across grades for each subject:

$$\epsilon^{s}_{i,t} \perp \epsilon^{s}_{i,t-1} \ \forall i, s \tag{A7}$$

Under this assumption, and the assumption that underlying skill in two subsequent years are correlated, we use a student's test score in a given subject (e.g. math) during the 3rd grade ($T^{math}_{i,3}$) as an instrument for the student's test score in the same subject (e.g. math) during 4th grade ($T^{math}_{i,4}$).

Table A8 summarizes, with instruments for interaction terms in the same way. We find that instrumenting makes the minority slope on past achievements more negative.

Table A8: Instrumenting for past achievement

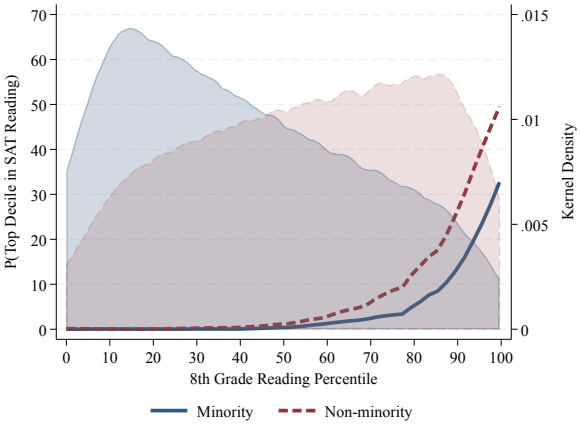
	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Minority	-0.09***	0.11***	0.15***	0.15***	-0.08***	0.07***	0.13***	0.11***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
4th Grade Math Percentile		0.51***	0.64***	0.62***				
		(0.00)	(0.00)	(0.00)				
Minority × 4th Grade Math Percentile		-0.29***	-0.35***	-0.34***				
		(0.00)	(0.01)	(0.00)				
4th Grade Reading Percentile						0.42***	0.61***	0.56***
						(0.00)	(0.00)	(0.00)
Minority × 4th Grade Reading Percentile						-0.19***	-0.27***	-0.25***
						(0.00)	(0.01)	(0.00)
Constant	0.13***	-0.15***	-0.23***	-0.22***	0.12***	-0.11***	-0.22***	-0.19***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Instrument			Read	Lag Math			Math	Lag Read
Adjusted R^2	0.02	0.19	0.18	0.18	0.01	0.14	0.11	0.12

Notes: This table shows the relationship between a student's 4th grade achievement and whether they reach excellence in 8th grade. The dependent variables are whether the student reaches excellence in 8th grade math (columns 1-4) and in 8th grade reading (columns 5-8).

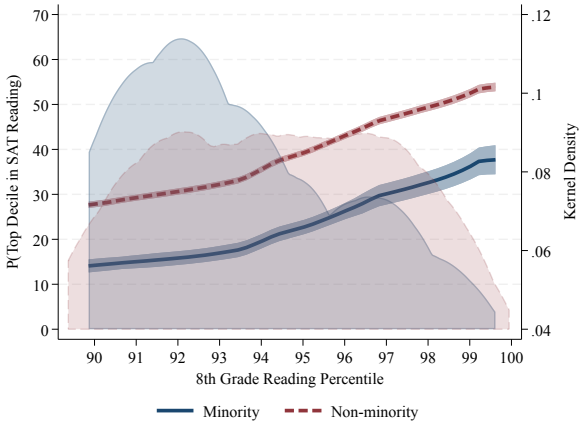
G.2 College readiness

Figure A17: Excellence in SAT Reading by 8th Grade Rank

(a) Full Sample



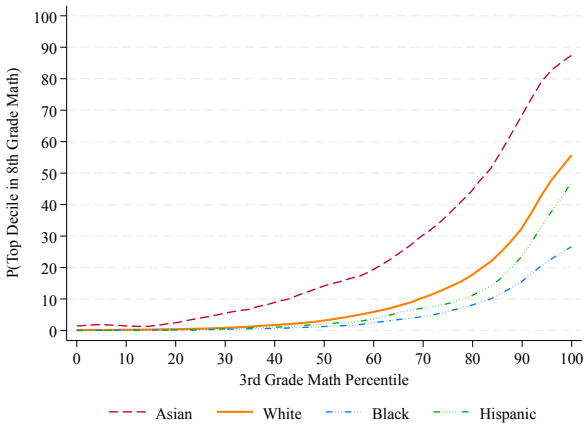
(b) Among Top 10% of 8th Graders



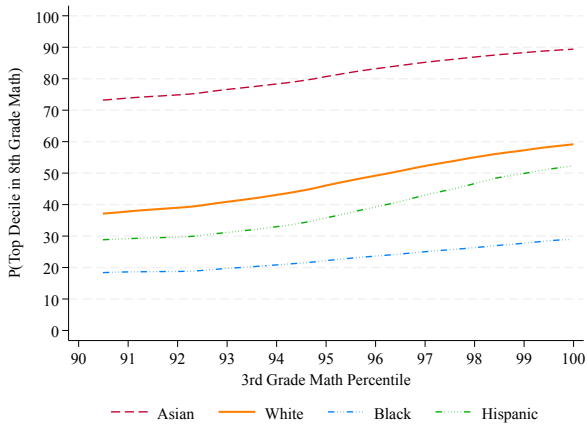
Notes: These figures show the proportion of students who score in the top decile of the SAT math section nationally, and the distribution of 8th grade achievement in reading. Figure (a) shows for each 8th grade decile group, while (b) show for each percentile in the top 10 percent. Analysis uses the "high school sample", as defined in Section I.

Figure A18: 8th grade Excellence by 3rd Grade Rank, by Finer Race Groups

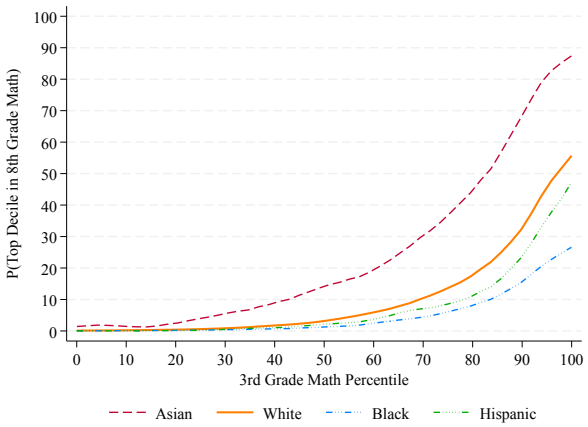
(a) Full Sample, Math



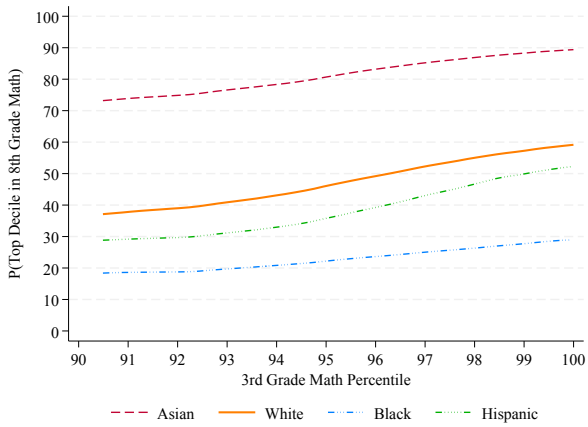
(b) Among Top 10% of 3rd Graders, Math



(c) Full Sample, Reading



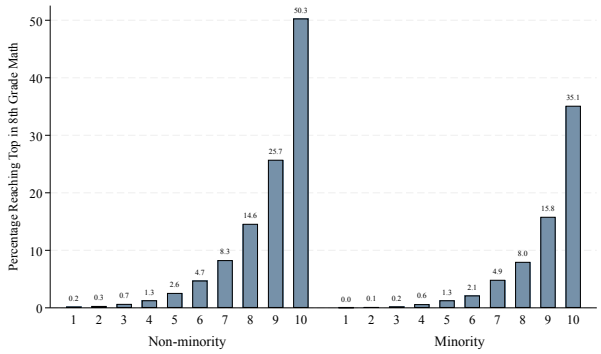
(d) Among Top 10% of 3rd Graders, Reading



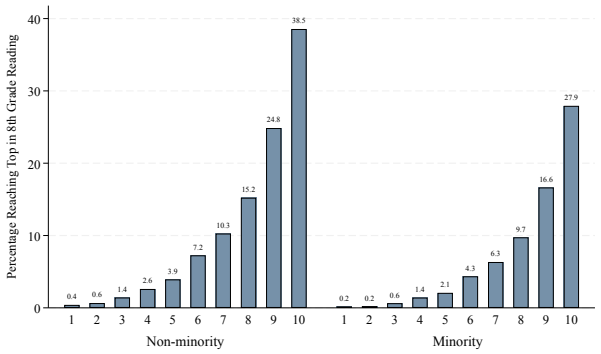
Notes: These figures show the proportion of each race group that reach the top decile of 8th grade achievement by 3rd grade achievement. Figures (a) and (b) show for math achievement, and figures (c) and (d) for reading. Figure (a) and (c) show for each 3rd grade decile group, while (b) and (d) show for each percentile in the top 10 percent. Analysis uses the "main sample", as defined in Section I.

Figure A19: Racial Differences in Reaching 8th Grade Excellence

(a) Math

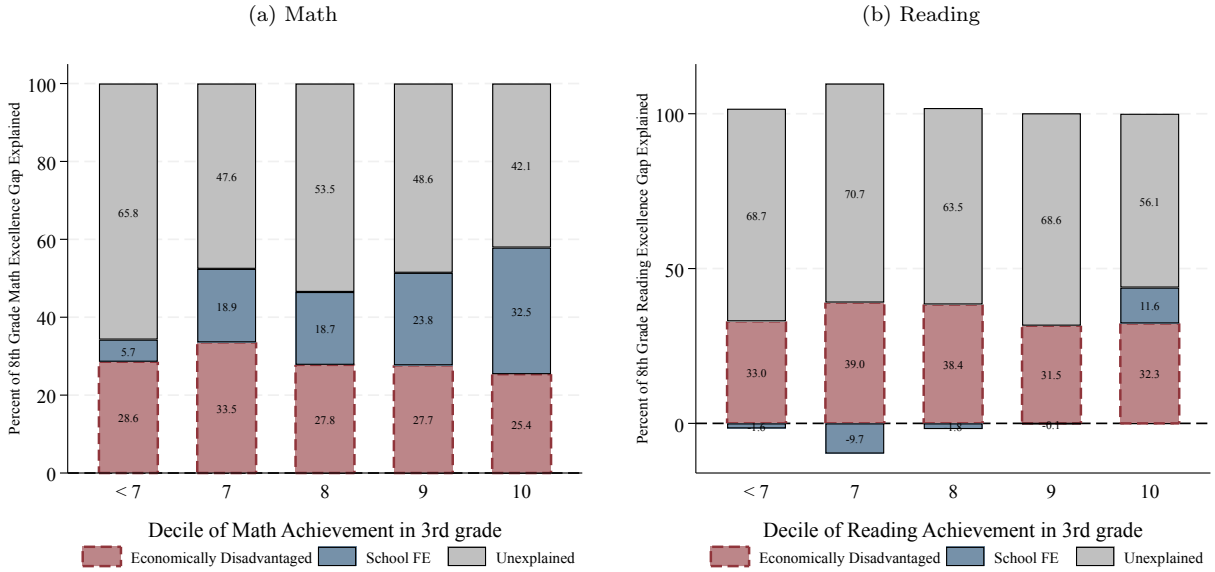


(b) Reading



Notes: These figures show the rates of reaching excellence in 8th grade (a) math and (b) reading, conditional on the decile of achievement in 3rd grade for that subject. Excellence is defined as reaching the top decile on the math MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

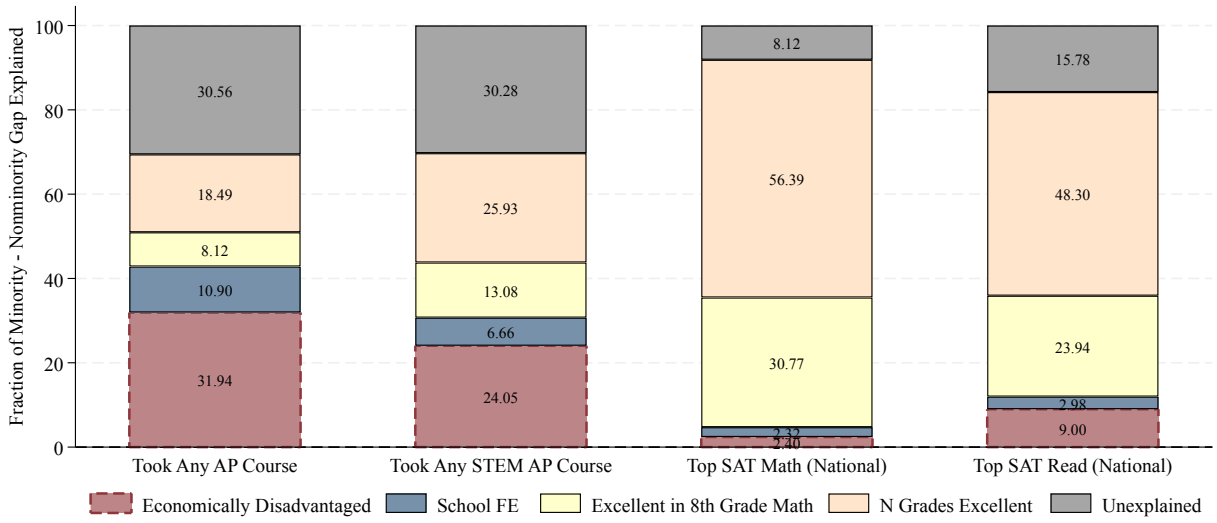
Figure A20: Explaining Racial Differences in Persistence and Entry into 8th Grade Excellence



Notes: These figures display the Gelbach decomposition in the 8th grade excellence gap in (a) math test scores (b) reading test scores. Both figures split the sample by the decile in achievement that a student achieved in that subject in 3rd grade. Each bar reflects results from its own Gelbach decomposition. "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 households receiving food or cash assistance, is homeless, is migrant, or is in foster care. "School FE" indicates the inclusion of school fixed effects. Analysis uses the "main sample", as defined in Section I.

H College Readiness

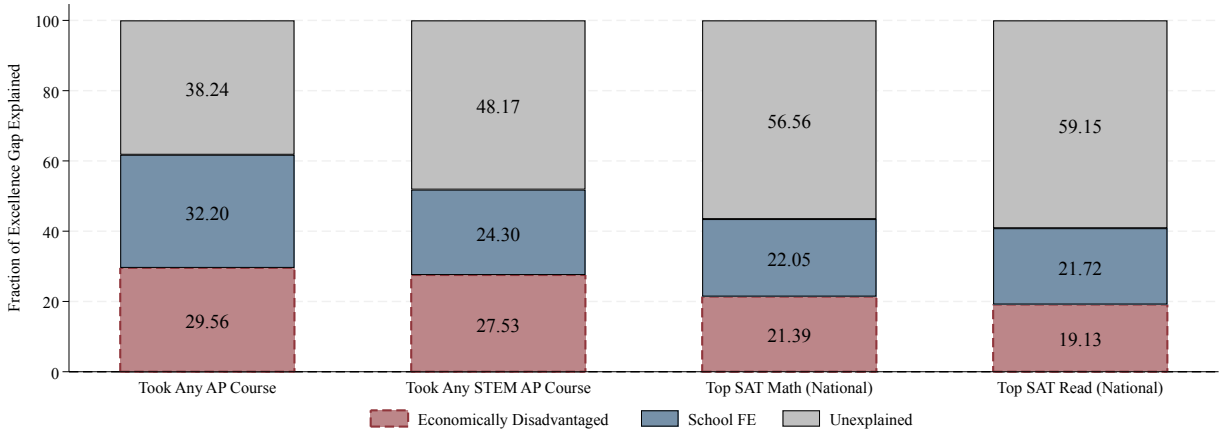
Figure A21: Decomposing Gaps in College Readiness



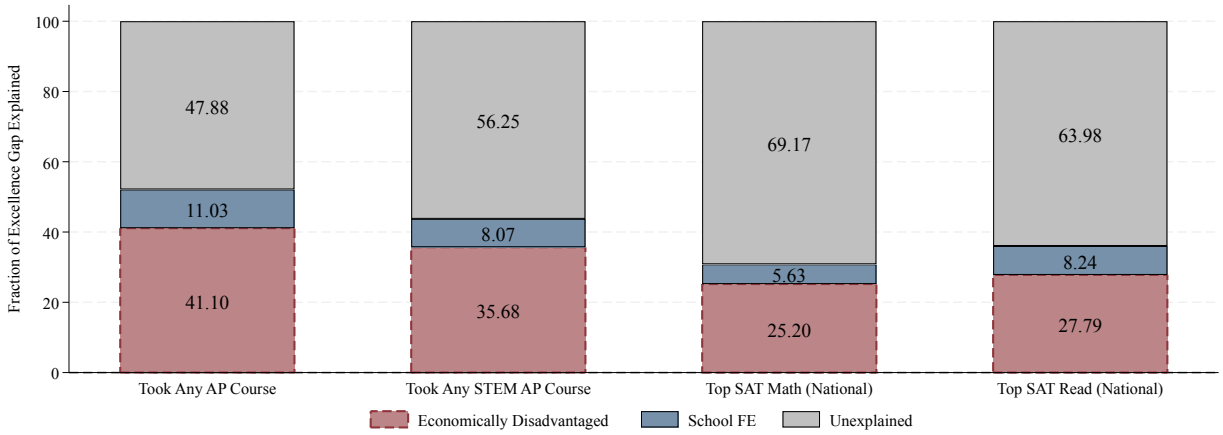
Notes: This figure shows the Gelbach decomposition of the minority - non-minority gaps in college readiness: takes at least one AP course, takes at least one STEM AP Course, reaches the top decile of SAT Math (i.e. above the 90th percentile reported by College Board in the year the exam was taken), and reaches the top decile of SAT Evidence-Based Reading & Writing (i.e. above the 90th percentile reported by College Board in the year the exam was taken). "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 households receiving food or cash assistance, is homeless, is migrant, or is in foster care. Analysis uses the "high school sample", as defined in Section I.

Figure A22: Explaining Racial Differences Persistence and Entry in College Readiness (Reading)

(a) Among Those Top in 8th Grade Reading

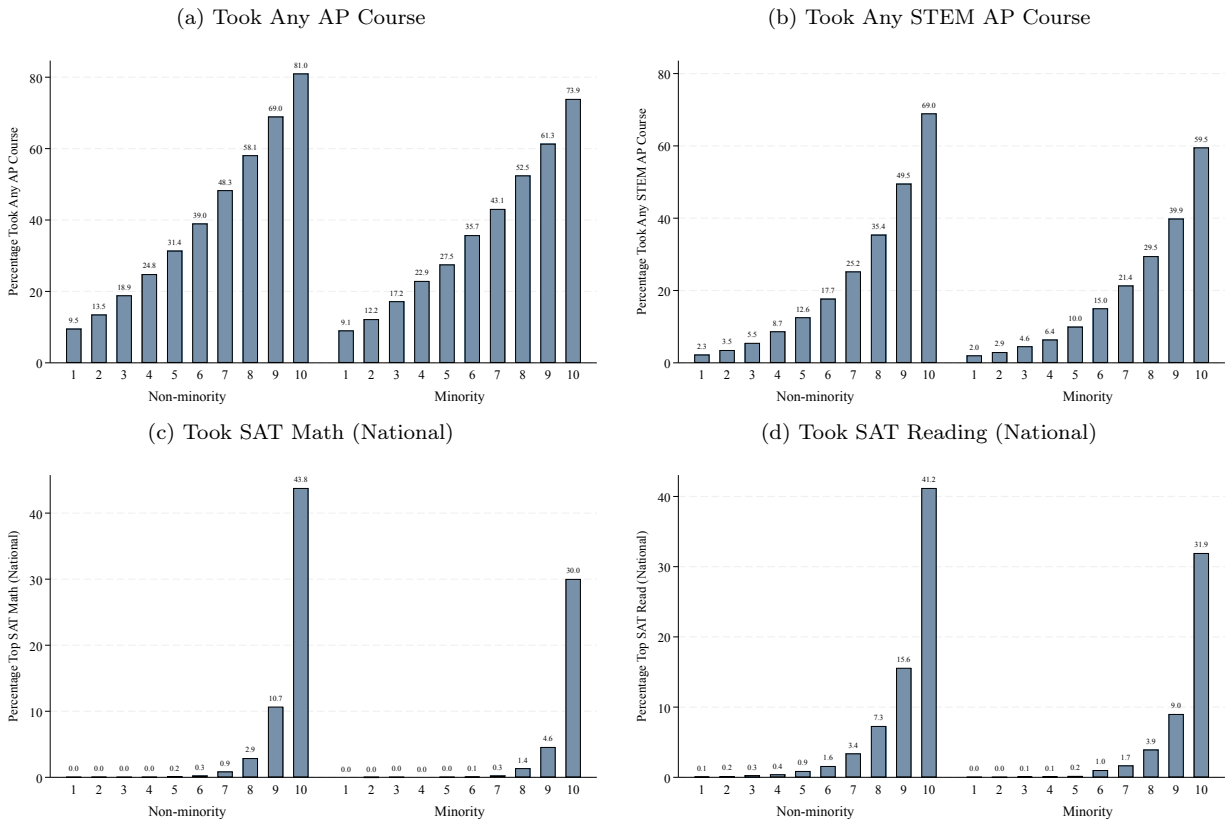


(b) Among Those Not Top in 8th Grade Reading



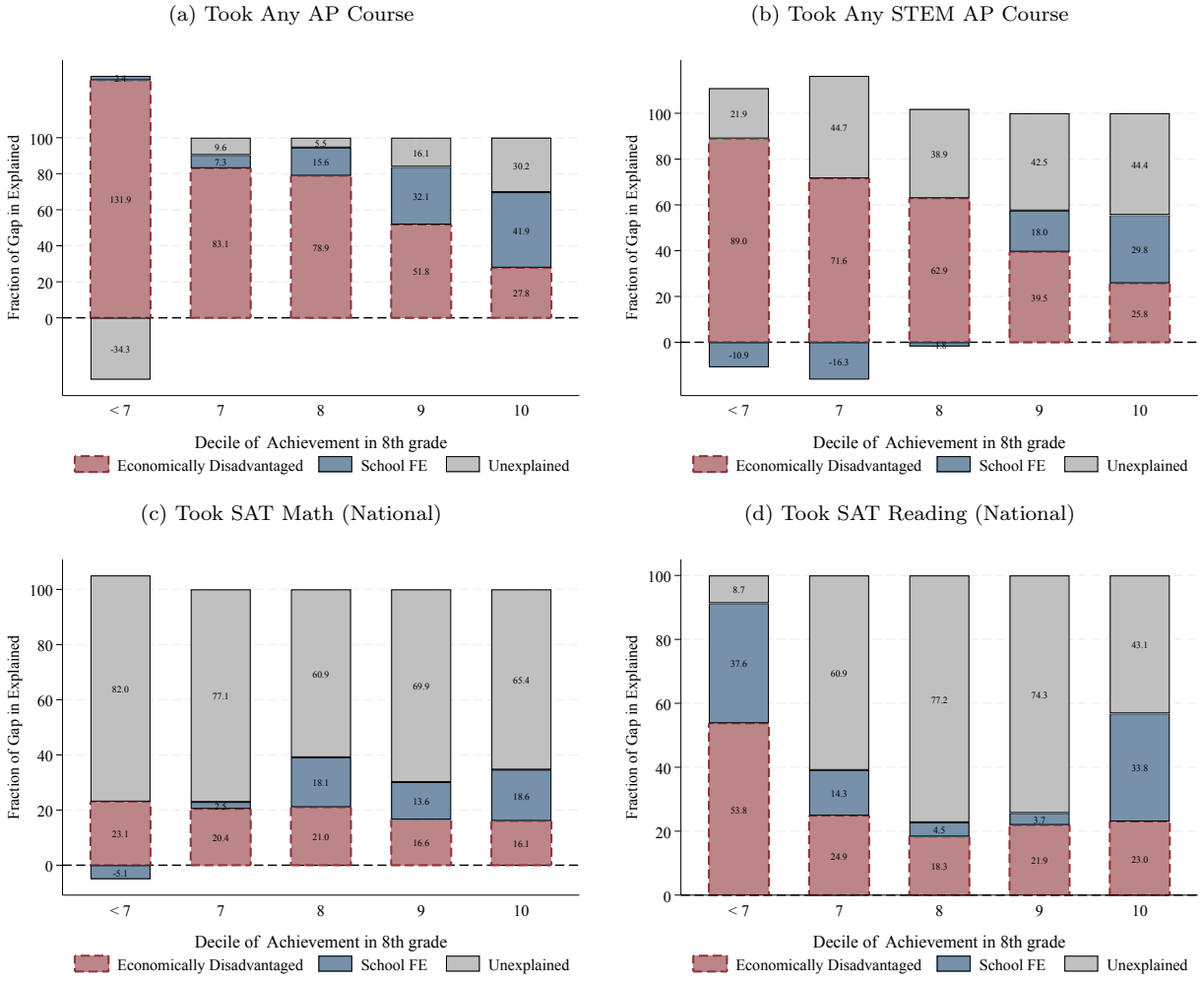
Notes: This figure displays the Gelbach decomposition of each minority non-minority difference in whether a student takes at least one AP course, takes at least one STEM AP Course, reaches the top decile of SAT Math, and reaches the top decile of SAT Evidence-Based Reading & Writing, among (a) students who achieved excellence in 8th grade reading and (b) students who did not achieve excellence in 8th grade reading. For SAT standards, we use the top decile of nationwide test-takes, using the 90th percentile score reported by College Board for the year that the exam was taken. "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 the "high school sample", as defined in Section I.

Figure A23: Racial Differences in Persistence and Entry into College Readiness



Notes: These figures show the rates of reaching college-readiness, conditional on the decile of achievement in 3rd grade for that subject. Each panel shows for a different outcome: (a) whether a student takes at least one AP course, (b) takes at least one STEM AP Course, (c) reaches the top decile of SAT Math, and (d) reaches the top decile of SAT Evidence-Based Reading & Writing. For SAT standards, we use the top decile of nationwide test-takes, using the 90th percentile score reported by College Board for the year that the exam was taken. Analysis uses the "high school sample", as defined in Section I.

Figure A24: Explaining Racial Differences in College Readiness

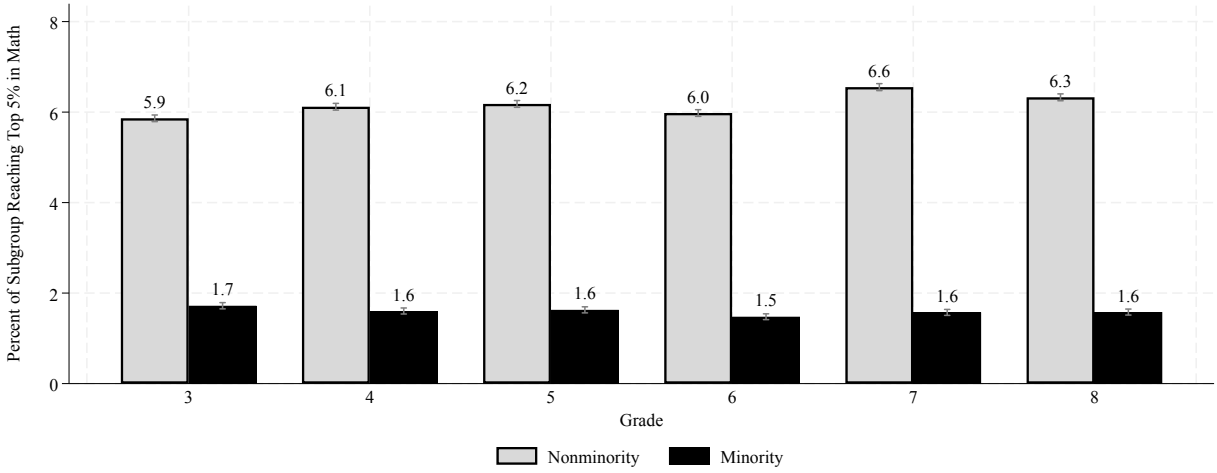


Notes: This figure displays the Gelbach decomposition of each minority non-minority difference in college-readiness, conditional on the decile of achievement in 3rd grade for that subject. Each panel shows for a different outcome: (a) whether a student takes at least one AP course, (b) takes at least one STEM AP Course, (c) reaches the top decile of SAT Math, and (d) reaches the top decile of SAT Evidence-Based Reading & Writing. For SAT standards, we use the top decile of nationwide test-takes, using the 90th percentile score reported by College Board for the year that the exam was taken. "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 the "high school sample", as defined in Section I.

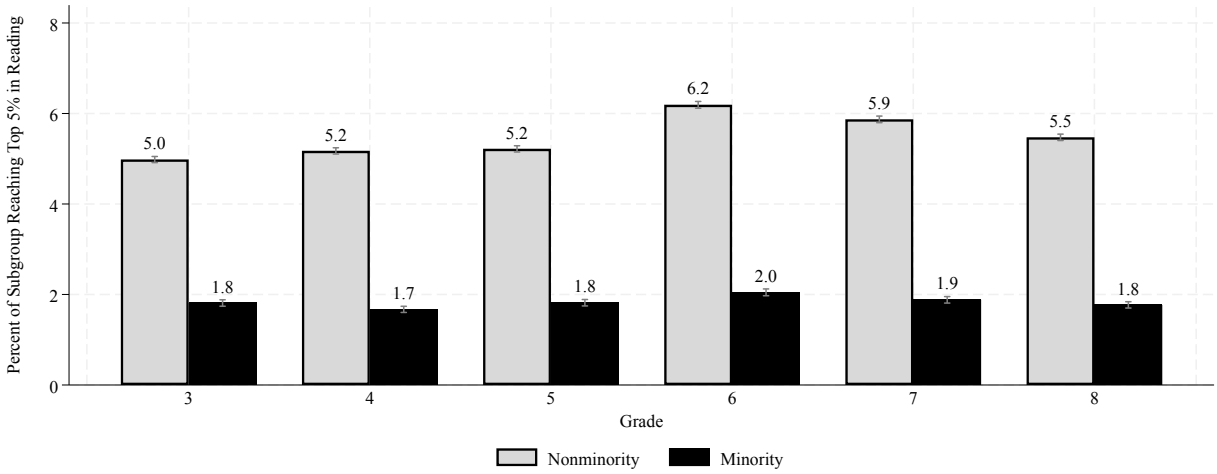
I Alternative Definitions of the Excellence Gap

Figure A25: The Excellence Gap as Top 5%

(a) Math



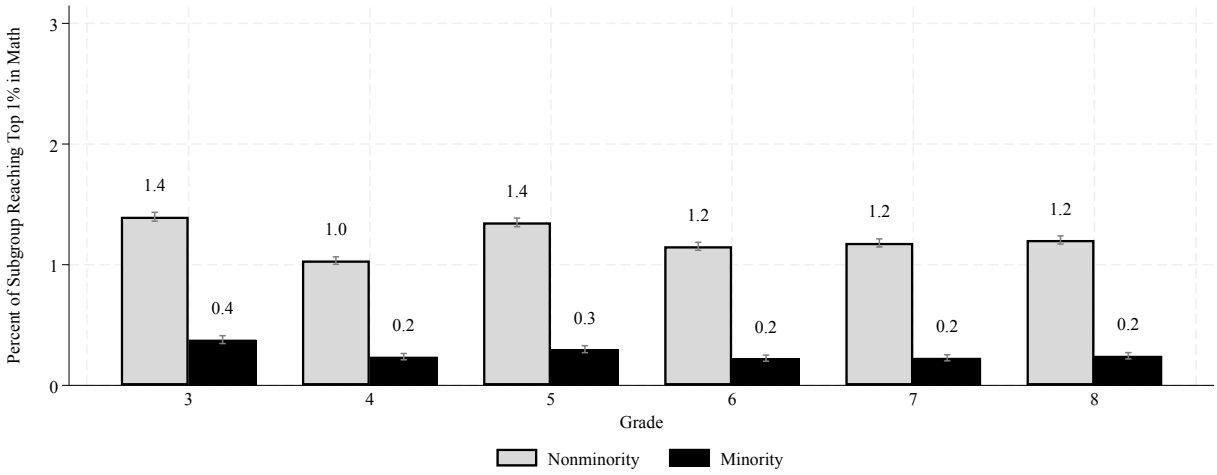
(b) Reading



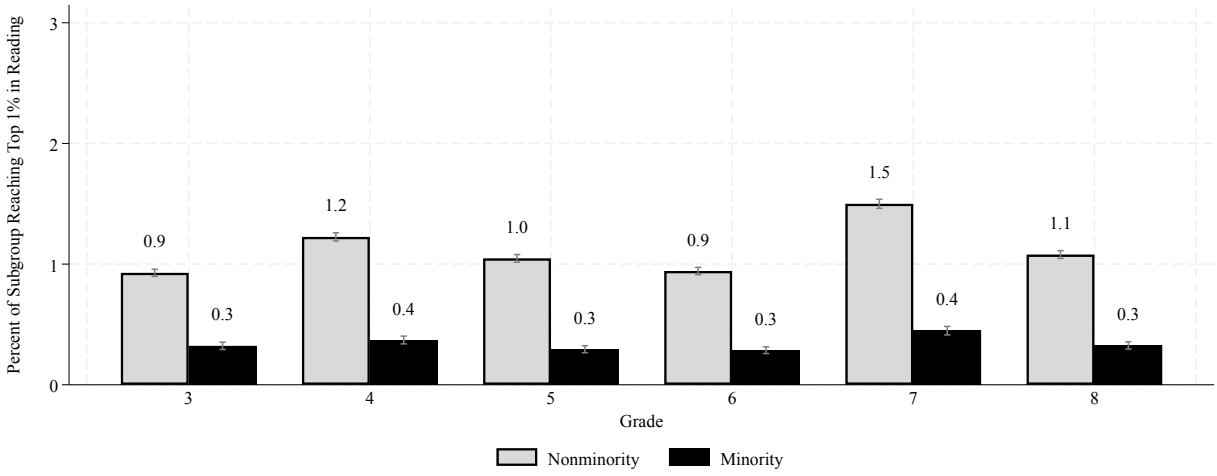
Notes: These figures display the percentage of minority and non-minority groups that reach the top 5% in (a) math and (b) reading of MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

Figure A26: The Excellence Gap as Top 1%

(a) Math

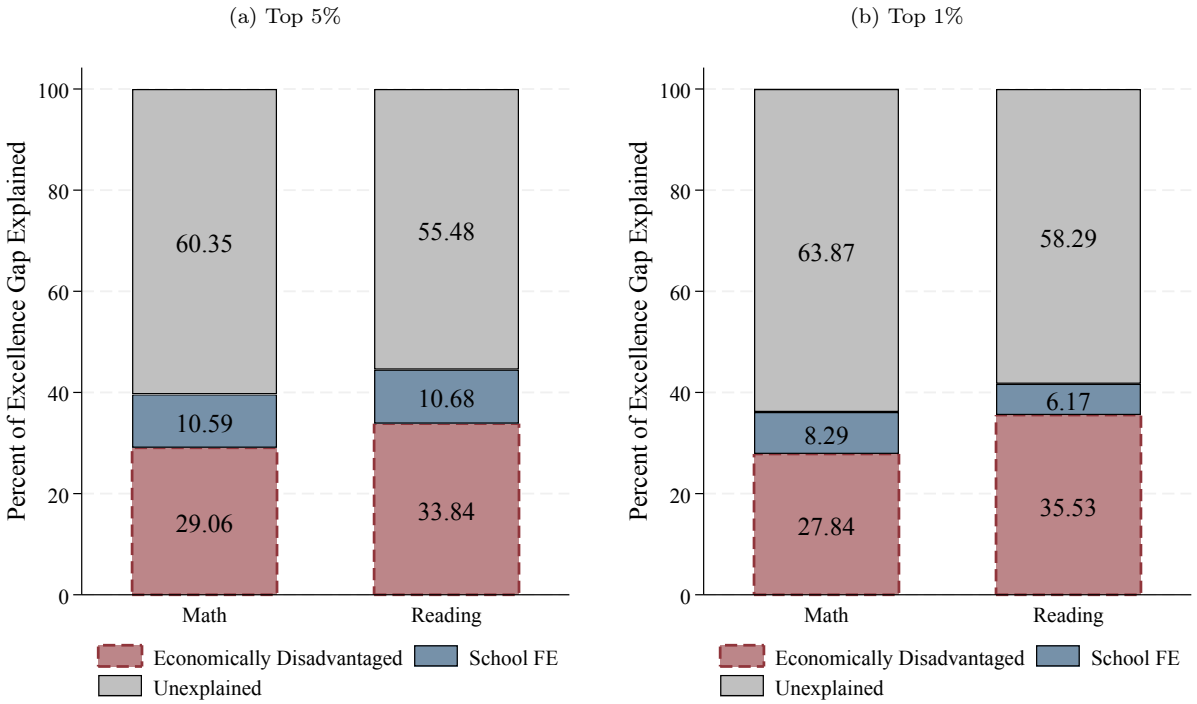


(b) Reading



Notes: These figures display the percentage of minority and non-minority groups that reach the top 1% in (a) math and (b) reading of MEAP (for all years until 2014) or M-STEP (for all years after 2014). Analysis uses the "main sample", as defined in Section I.

Figure A27: Explaining the Excellence Gap



Notes: This figure displays the Gelbach decomposition of the excellence gap in math test scores among 3rd through 8th graders, where excellence is defined as reaching the the top (a) 5% and (b) 1% on the MEAP (for all years until 2014) or M-STEP (for all years after 2014) math section. "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 the "main sample", as defined in Section I.

References

Botelho, F., R. A. Madeira, and M. A. Rangel (2015). Racial discrimination in grading: Evidence from brazil. *American Economic Journal: Applied Economics 7*(4), 37–52.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In *Handbook of econometrics*, Volume 5, pp. 3705–3843. Elsevier.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

Hartman, E. (2021). Equivalence testing for regression discontinuity designs. *Political Analysis 29*(4), 505–521.

Hemelt, S. W. and B. Jacob (2017). Differentiated accountability and education production: Evidence from nclb waivers. Technical report, National Bureau of Economic Research.

Hemelt, S. W. and B. A. Jacob (2020). How does an accountability program that targets achievement gaps affect student performance? *Education Finance and Policy 15*(1), 45–74.

Kline, P., E. K. Rose, and C. R. Walters (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics 137*(4), 1963–2036.

Zabel, J. E. (2008). The impact of peer effects on student outcomes in new york city public schools. *Education Finance and Policy 3*(2), 197–249.