Take Two! SAT Retaking and College Enrollment Gaps

By Joshua Goodman, Oded Gurantz and Jonathan Smith*

Only half of SAT-takers retake the exam, with even lower retake rates among low income and underrepresented minority (URM) students. We exploit discontinuous jumps in retake probabilities at multiples of 100, driven by left-digit bias, to estimate retaking's causal effects. Retaking substantially improves SAT scores and increases four-year college enrollment rates, particularly for low income and URM students. Eliminating disparities in retake rates could close up to 10 percent of the income-based gap and up to seven percent of the race-based gap in four-year college enrollment rates of high school graduates.
JEL: I20, I23, D90

U.S. college enrollment varies substantially by income and race. In 2016, 83 percent of students from high-income families enrolled in college immediately after graduating from high school, while only 67 percent of students from low-income families did so (McFarland and Hussar, 2018). Over 70 percent of White high school graduates immediately enroll in college, compared to 56 percent of Black high school graduates. When low income and underrepresented minority (URM) students enroll in college, they are disproportionately likely to do so at colleges with relatively low graduation rates and low expected earnings, such as two-year community colleges.[1] One partial explanation of such disparities comes from informational and procedural complexities in the transition to college, which disproportionately impact students who lack the resources or guidance to navigate college processes as effectively as their more advantaged peers (Page and Scott-Clayton, 2016).

College entrance exams such as the SAT and ACT are one source of complexity. Although much has been written about the decision to take a college entrance exam for the first time (the extensive margin), we know very little about students' decisions to retake such exams (the intensive margin). Only 54 percent of SAT-takers retake the SAT at least once, implying that nearly half of students

* Goodman: Department of Economics, Brandeis University, 415 South Street, Waltham, MA 02453, JoshuaGoodman@Brandeis.edu. Gurantz: Truman School of Public Affairs, University of Missouri, 216A Middlebush Hall, Columba, MO 65201, gurantzo@missouri.edu. Smith: Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302, jsmith500@gsu.edu. These views do not reflect those of the College Board. We are grateful for helpful feedback from John Friedman and three anonymous referees, as well as seminar participants at Brandeis, Columbia, Duke, the Federal Reserve, Harvard, Montana State University, Purdue, University of Connecticut, University of Maryland, University of Memphis, and West Point. We also thank the College Board for sharing data and Melanie Rucinski for providing excellent research assistance.

[1]For purposes of this paper, underrepresented minority (URM) students refers to those who identify as Black, Hispanic, or Native American.

never retake.[2] Those who do not retake appear at a competitive disadvantage, given that retaking is strongly incentivized by current admissions practices of U.S. colleges. Nearly 75 percent of four-year colleges that use SAT scores in the admissions process publicly claim to consider only a student's maximum score.[3] Over 80 percent of those define that maximum as the superscore, the combination of the highest scores a student receives on each section of the exam, even if those scores occurred at different takes. For three-fourths of colleges, retaking can only improve students' chances of being admitted by making their applications more competitive.

In this paper, we assess the consequences of SAT retaking for student achievement, with a focus on the differential effects of retaking across demographic groups. Our analysis has three parts. First, we analyze which students take up the opportunity to retake the SAT. We show that retake rates increase with initial SAT scores across most of the distribution, falling only above the 95th percentile. Retake rates increase with income but are also higher for low income students who use waivers that render retakes free, suggesting that financial costs may deter some students from retaking.[4] Female students are three percentage points more likely to retake than male students and Asian-American students are 12 percentage points more likely to retake than White students, differences robust to conditioning on a rich set of covariates. URM students are nine percentage points less likely to retake than White students, a gap only partly explained by income differences between those two groups. Students who take their first exam at an earlier date are more likely to retake, in part because they have more opportunities to do so.

Second, we estimate the causal impact of retaking on both test scores and college enrollment. Relative to those who do not retake, retakers may believe their first take is not as reflective of their true abilities, may be generally more motivated, or may be better informed about the benefits of retaking, all of which would confound attempts at causal inference. To eliminate endogenous retaking, we borrow from behavioral economics and psychology literature that observes people focusing disproportionately on the leftmost digits of numbers when making decisions. Pope and Simonsohn (2011) first observed such behavior in the test-taking context, using lumpiness in the distribution of maximum scores to infer

[2]Retake rates are even lower for the ACT, the other major American college entrance exam. Only 41 percent of 2009 ACT-takers retook the exam, a proportion that rose to 45 percent by 2015. Many of the associations we document between SAT-retaking, demographics and test scores also appear in the ACT. See the ACT's 2016 Technical Brief entitled "Multiple Testers: What Do We Know About Them?", by Matt Harmston and Jill Crouse.

[3]The remaining 26 percent of colleges that use all SAT scores in the admissions process tend to be more selective than those using maximum scores. See the College Board's 2015 publication "SAT Score-Use Practices by Participating Institution", in which the College Board surveyed all SAT-using colleges about their admissions practices. The cited percentages are conditional on being in the 87 percent of colleges with a valid response.

[4]Students are eligible for such fee waivers if they receive or are eligible for federally subsidized school lunch, receive public assistance, live in federally subsidized public housing, or are homeless. To get the waiver, a student must contact a school counselor or authorized community-based organization that confirms the student's eligibility.

that students scoring just below multiples of 100 must retake the SAT at higher rates than those scoring at or above such round number thresholds.[5]

We confirm this finding directly: students with first SAT scores just below a multiple of 100 are, on average, one percentage point more likely to retake the SAT than students with scores just at or above that number. Discontinuities in retaking rates appear at almost all multiples of 100, ranging in magnitude from 0.5 percentage points at lower thresholds to 12 percentage points for the highest threshold. These thresholds serve as exogenous sources of variation in retake rates, motivating a fuzzy regression discontinuity design to estimate the impact of retaking on SAT scores and college enrollment, with retaking instrumented by an indicator for scoring just below a multiple of 100.

The impact of threshold-induced retaking on SAT scores is substantial. On average, retake scores are 0.15 standard deviations higher than first take scores (46 points on a 2400 point scale). Retaking once improves students' admissions-relevant superscores by nearly 0.3 standard deviations (90 points). For students who initially score in the lower half of the SAT distribution, retaking once boosts superscores by nearly 0.4 standard deviations (120 points). Higher-scoring students see meaningful but smaller test score gains, in part because of ceiling effects. Retaking increases the SAT scores of low income and URM students by more than it does for their high income and non-URM counterparts.

The score increases resulting from retaking are large enough to drive substantial improvements in college enrollment outcomes, particularly for low income and URM students. On average, retaking increases the probability of enrolling in a four-year college by 13 percentage points, driven in large part by substitution away from two-year colleges. Retaking also improves the quality of colleges students attend. These effects are even larger for low income and URM students and substantially smaller for high income and non-URM students.

Retaking also improves college quality, inducing students to attend colleges with historical B.A. completion rates six percentage points higher than they would otherwise. It does so by shifting enrollment from non-selective, low completion rate (often two-year) colleges to somewhat more selective, higher completion rate four-year colleges. Enrollment in the most selective, highest graduation rate colleges does not change. Because it does not alter application patterns to the colleges where enrollment shifts are observed, retaking most likely changes college choices by increasing the odds of admission through increased SAT scores.

The college enrollment effects of threshold-induced retaking are large. This may be partly explained by the fact that taking the SAT itself indicates interest in four-year colleges, so that SAT-takers may generally be less constrained financially or otherwise than non-takers. SAT-takers induced to retake by missing a round number score may also be unusual. We show that compliers do not differ much by income or race from the average retaker but acknowledge that students with target

[5]Their data, a 25 percent sample of SAT scores from 1996-2001, lacked explicit measures of retaking and college enrollment outcomes.

scores may differ along unobservable dimensions. If compliers are particularly motivated to attend four-year colleges, our estimates may provide upper bounds for the enrollment effect of increasing retaking among wider populations.

In the third part of our analysis, we combine calculations from the first two parts to show that retaking behavior can explain a substantial fraction of college enrollment gaps. Among the SAT-taking population, low income students are 21 percentage points less likely to retake, yet doing so increases enrollment at four-year college enrollment by 30 percentage points. This implied six percentage point effect (0.21 * 0.30) represents 25 percent of the four-year college enrollment gap between low and high income SAT-takers. A similar calculation shows that retaking explains 14 percent of the four-year college enrollment gap between URM and non-URM SAT-takers. Because roughly half of high school graduates do not take college entrance exams, the proportion of college enrollment gaps among all high school graduates explained by retaking is roughly half the proportion within the SAT-taking population. Our calculations suggest that disparities in retake rates explain up to 10 percent of the income-based gap and up to seven percent of the race-based gap in four-year college enrollment rates of high school graduates. These calculations ignore any potential general equilibrium effects, driven by actions of colleges or other students, that might result from substantial increases in retake rates of low income and URM students.

Our results suggest that individual students, particularly low-scoring, low income, or URM students, should likely retake college entrance exams more than they currently do. Though our data is too recent to measure retakers' completion rates with precision, evidence from other contexts suggests the shift from two-year to four-year colleges and the general improvement in quality of college chosen will lead to higher degree completion rates (Goodman, Hurwitz and Smith, 2017; Cohodes and Goodman, 2014) and higher labor market earnings (Zimmerman, 2014; Canaan and Mouganie, 2018).

Our paper contributes to the literature in two ways. First, we highlight one part of the college enrollment process that contributes to socioeconomic gaps in educational attaintment but has received relatively little attention. Most recent research has focused on students' decisions about whether to take college entrance exams. For example, states mandating the SAT or ACT as part of accountability systems increased exam-taking rates and meaningfully improved four-year college enrollment rates and the selectivity of colleges chosen, particularly for URM students (Klasik, 2013; Hurwitz et al., 2015; Goodman, 2016; Hyman, 2017). Opening exam testing centers near students' homes also increases test-taking and college enrollment rates (Bulman, 2015). This body of research on first-time exam taking suggests both that some students underestimate their college readiness and that relatively small costs can dissuade such students from taking college entrance exams.

Retaking has received much less attention. Vigdor and Clotfelter (2003) use data on applicants to three selective U.S. colleges to study the predictors of retaking, arguing that retake-induced score increases are too large to be explained by selection and must represent increased familiarity with the test or actual learning. Frisancho et al. (2016) use data from the Turkish college entrance exam to similarly argue that retaking generates familiarity- or learning-based gains that are larger for less advantaged students, thus potentially narrowing socioeconomic gaps. We build on this small literature by using data on the universe of over 10 million SAT takers from the high school classes of 2006-2014 both to document predictors of retaking and, more importantly, to provide the first causal estimates of the impact of retaking on college enrollment.

Our second contribution is to the behavioral economics literature. We add to the research on "left-digit bias", which has been shown to operate in car and housing sales (Busse et al., 2013; Lacetera, Pope and Sydnor, 2012) and athletic performance (Allen et al., 2017; Foellmi, Legge and Schmid, 2016). The example of college entrance exam retaking is particularly striking as it demonstrates large real-life consequences of left-digit bias. This work also adds to a broader literature on behaviorally-based public policy interventions. Our results suggest that students underestimate or undervalue the expected benefits of retaking compared to its small apparent costs. Such costs include a few hours of test-taking time, study effort, and a fee that ranges from $0 for low income students, who qualify for a fee waiver that renders the first two takes free, to about $50-$60 for higher income students. The retaking of college entrance exams may be a margin ripe for the type of behavioral intervention that has generated interest in public policy generally (Chetty, 2015) and in education policy specifically (Lavecchia et al., 2016; Levitt et al., 2016). We discuss a variety of potential interventions in the conclusion but note that encouraging or requiring students to take their first exam earlier might be particularly effective, given that many low income and URM students do so in 12th grade when few retaking opportunities remain.

I. Data and Descriptive Analysis

For descriptive purposes, we begin with student-level data from the College Board on the universe of nearly 14 million students from the high school classes of 2006-14 who had valid scores on all three sections of the SAT: math, critical reading, and writing. For purposes of causal inference, we include in our regression discontinuity sample the 12 million of those students who first took the SAT by November of senior year and thus had at least one opportunity to retake prior to graduating high school.[6]

For each student, we observe SAT math, reading and writing scores from all takes, as well as the timing of those takes. This allows us to construct first total SAT scores across all three sections, as well as the superscores most commonly used by college admissions offices. Each SAT section is scored on a scale of 200

[6]Students who first took the SAT later than that have extremely low retake rates.

to 800 in multiples of 10, so that total SAT scores can range from 600 to 2400.[7] For these cohorts, the SAT was offered seven times a year and students could retake the SAT as many times as their college application timeline allowed. Each SAT take during this time cost roughly \$40-\$60, with low income students who applied for and received fee waivers eligible to take the exam twice at no cost. We observe whether each student used such a fee waiver, as well as self-reported demographic information on gender, race, parental education and family income.

Our main outcome of interest is college enrollment, which we observe via a merge between the College Board data and the National Student Clearinghouse (NSC), which tracks enrollment choices of 94 percent of U.S. college students. We assign students to the first college they enroll in within 180 days of high school graduation and characterize colleges both by sector (four-year vs. two-year) and by expected graduation rate and later income. We use the NSC data to compute each college's expected graduation rate as the fraction of students in our data who initially enroll in a given institution and who earn a B.A. anywhere within six years. Relative to graduation rates available through the Department of Education's Integrated Postsecondary Education Data System, this measure has the advantage of existing for all colleges and being comparable across the two- and four-year sectors.[8] We use data from from Chetty et al. (2017) to characterize the expected income of each college's initial enrollees at ages 32-34.[9] We also observe all colleges to which a student ultimately sends SAT scores, a decent proxy for college application behavior given that many colleges require standardized test scores as part of their admissions process (Pallais, 2015; Bond et al., 2018; Smith, 2018).

Table 1 shows mean characteristics of the full sample, the subset who first took the SAT by November of senior year and thus comprise our regression discontinuity sample, and three subsamples of interest: lower scoring students, defined as those whose first SAT score is closest to or below the median threshold of 1500; low income students, defined as those with self-reported family income below \$50,000; and underrepresented minority (URM) students, defined as those who report their race/ethnicity as Black, Hispanic or Native American. Among the full sample, 27 percent are URMs and 21 percent are low income. Twenty percent of students used a fee waiver on their first SAT take, making the exam free.

First SAT scores average 1475 with maximum SAT scores over 50 points higher. Fifty-four percent of students retake the SAT at least once, and some more than once, so that the mean number of takes is over 1.7. On average, students first

[7]Our data cover a roughly ten-year period during which the SAT comprised three sections scored on a 2400 point scale. Beginning in 2016, the SAT has two sections and a 1600 point scale.

[8]To deal with selection into college, students who do not immediately enroll in any college are assigned the mean six-year B.A. completion rate of all students who do not enroll immediately after high school graduation. For more detailed discussion of these measures, see Goodman, Hurwitz and Smith (2017).

[9]We use the baseline income measure favored by Chetty et al. (2017), which measures the 2014 income of each college's initial enrollees from the birth cohorts of 1980-82.

take the SAT about 12 months before high school graduation, in May or June of junior year. Though not shown in the table, 69 percent of students give themselves substantial time to retake by first taking the SAT in 11th grade, while 29 percent first take it in 12th grade.[10] Fifty-seven percent of SAT-takers first enroll in a four-year college within 180 days of high school graduation and another 18 percent enroll in a two-year college. The regression discontinuity sample looks fairly similar to the full sample, though with a higher retake rate (61 percent compared to 54 percent) and a higher four-year college enrollment rate (61 percent compared to 57 percent). The three subsamples come from lower income families, have lower SAT scores, retake less often and are less likely to enroll in four-year colleges.

Table 2 shows predictors of retaking from a linear probability model that regresses a retaking indicator on various covariates. A few clear patterns emerge in the top three panels. Low income students are 21 percentage points less likely to retake than high income students. This could be driven by differential ability to afford the costs of retaking or by differential college preferences or perception of the benefits of retaking. URM students are 9 percentage points less likely to retake than White students, a gap partly but not solely explained by income differences across these groups. Asian students are 12 percentage points more likely to retake than White students and female students are three percentage points more likely to retake than male students, differences robust to inclusion of further controls.

The bottom panel shows that those with higher first SAT scores are actually more likely to retake the SAT, a pattern we explore in more detail in the next section.[11] Though initially higher scoring students may have less need to improve their scores, such students tend to be more advantaged along a number of dimensions that predict higher retake rates. Students who use fee waivers are more likely to retake, suggesting testing fees may affect low income students' propensity to retake. High retake rates among fee waiver users might also be driven by positive selection as the most informed and motivated low income students seek such waivers.

Each month earlier that a student first takes the SAT is associated with a more than four percentage point increase in retaking probability, even controlling for all the aforementioned factors. Early takers are more likely to retake in part because they have more opportunities to do so. The timing of a student's first take is also the single most powerful predictor of the probability of retaking. Controlling for such timing quadruples the predictive power of this regression model, raising its R-squared from five to 20 percent. Though not shown in the table, over 40 percent of URM students first take the SAT in 12th grade, compared to just over 20 percent of white and Asian students, explaining some of the difference in retake

[10]Two percent first take the SAT in 9th or 10th grade.

[11]The relationship between first SAT score and retaking is fairly linear across all but the very top of the score distribution, so that our descriptive conclusions are substantively unchanged when controlling for higher order terms in first SAT scores.

rates by race. All of this suggests that encouraging students to take their first test at an earlier date may be a channel through which to increase retake rates.

II. Methodology

The propensity to retake the SAT is strongly correlated with observed student characteristics but is also likely correlated with unobserved characteristics such as true ability, motivation or information about the college admissions process. As such, even conditional on variables observed in our data, the relationship between retaking and college outcomes is probably biased upwards by the fact that highly motivated or well-informed students are both more likely to retake and more likely to enroll in college. Estimating the causal impact of retaking requires an exogenous source of variation that affects retaking but is not related to students' underlying characteristics. We exploit the fact that the distance of a students' first SAT score from the nearest multiple of 100 provides such a source of variation.

A. *Regression Discontinuity Specification*

We estimate the impact of round-number thresholds on retaking with a regression discontinuity design that fits linear specifications around each such threshold, similar to a linear spline specification but with additional flexibility built into the model. To do so, we generate two identical observations per student containing demographic information, SAT scores from each take, retaking measures, and college enrollment outcomes. The existence of multiple thresholds means that each student's total SAT score from the first take falls in between two multiples of 100 (or on one such multiple). For one of the student's two observations we generate regression discontinuity variables defined relative to the threshold to the left and for the other observation by the threshold to the right.[12]

We generate three such variables. The first, R, is defined in one observation as the nearest round-number threshold at or below a student's first SAT score and in the other observation as the nearest threshold above that score. The second, the running variable D, is defined as the distance between a student's first SAT score and R, which can take on values between -100 and 90. The third, *Below*, indicates whether a student's first SAT score falls below R or, equivalently, whether for that observation D is negative. For example, a student whose first SAT score is 1270 would appear in one observation with $R = 1200$, $Below = 0$ and $D = 70$ and in the other with $R = 1300$, $Below = 1$ and D equal to -30.[13]

[12] All students appear twice in the data except for a small number who appear once because of first SAT scores below the lowest threshold (700) or at or above the highest threshold (2300).

[13] As an example of a student whose first score is a multiple of 100, consider one who initially scored 1200. That student would appear in one observation with $R = 1200$, $Below = 0$ and $D = 0$ and in the other with $R = 1300$, $Below = 1$ and D equal to -100.

Our first stage regression takes the form:

(1)

$$Retook_{ir} = \beta_{FS} Below_{ir} + \sum_{n=7}^{23} [R_{n00} * (\alpha_{n00} + \gamma_{n00} D_{ir} + \delta_{n00} Below_{ir} * D_{ir})] + \mu_{cd} + \epsilon_{ir}$$

The outcome *Retook* indicates that student i near round-number threshold r retook the SAT at least once. The variable R_{n00} indicates an observation being defined relative to threshold 100*n, so that the summed terms in brackets represent 17 local linear regression discontinuity specifications, one at each threshold from 700 to 2300. The three terms within brackets flexibly allow each threshold to have its own intercept, slope below the threshold, and slope above the threshold. This resembles a linear spline specification that would use one observation per student but, unlike that specification, does not impose the restriction of a single slope between each threshold. Instead, each student is allowed to contribute to the estimation of one slope for the threshold below and one slope for the threshold above her first score. Relative to a linear spline, this specification imposes fewer functional form assumptions but has lower precision because nearly twice as many coefficients are being estimated.

The first stage coefficient of interest, β_{FS}, averages across all 17 thresholds the predicted differences in retake rates between those scoring just below and just above a threshold. Because students just on either side of the threshold are nearly identical in terms of academic skill and other characteristics, we interpret this coefficient as the causal impact on retaking of just missing a round number score. High school class by first SAT date fixed effects μ_{cd} control flexibly for class and test date-specific factors such as time trends in retaking behavior, selection into first test dates by different types of students, and varying exam difficulty over time.

Our ultimate goal is to estimate the impact of retaking on these SAT and college outcomes. To do so, we generate instrumental variable estimates of the form:

(2)

$$Y_{ir} = \beta_{IV} Retook_{ir} + \sum_{n=7}^{23} [R_{n00} * (\zeta_{n00} + \eta_{n00} D_{ir} + \theta_{n00} Below_{ir} * D_{ir})] + \nu_{cd} + \xi_{ir}$$

where *Retook* is instrumented by the indicator *Below*, as in equation 1.[14] We focus primarily on two sets of outcomes Y, measures of subsequent SAT performance and measures of college enrollment, including the sector and completion rate of college chosen. The coefficient of interest, β_{IV}, thus estimates the impact on SAT scores and college outcomes of the retaking driven by students' reactions to

[14]We implement this using two-stage least squares but find that, because our instruments are quite strong, limited information maximum likelihood estimation yields indistinguishable results.

round-number thresholds.

This flexible local linear specification allows us to use bandwidths ranging from 20 (below which we have only a single point with which to estimate a line) to 100 (above which our data cross additional thresholds that may affect estimation). We split the difference between these two extremes and choose 60 as our default bandwidth, but show in extensive robustness checks that the linearity of the data means our results are not sensitive to this choice.[15] Because our running variable is discrete, we cluster standard errors in all of these regressions by first SAT score as suggested by Lee and Card (2008). Clustering by first SAT score also accounts for the fact that some students have two observations in the stacked version of the data. We show later that our results remain statistically significant without clustering and remain robust to bandwidths of 50 and below, where each student is observed only once. Consistent with the earlier observed smoothness of covariates, we show that our results are robust to controlling for students' income, race, gender and parental education. Our specification also yields estimates extremely similar to those from a traditional linear spline specification.

B. First Stage

Figure 1 plots the fraction of students retaking the SAT as a function of first SAT scores. Retake rates rise with first SAT scores across most of the distribution, peaking above 65 percent for those with scores in the 1800-2000 range and then dropping for first scores above 2000, the 95th percentile of the distribution. Students whose first SAT score falls just below a multiple of 100 are more likely to retake the SAT than those whose score is that multiple of 100. These difference in retake rates are largest and visually clearest for the higher thresholds but also appear at many of the middle and lower thresholds in denser parts of the distribution.

Figure 2 makes this visually clearer with the stacked version of the data, where the running variable is the distance of each student's first SAT score from the nearest multiple of 100. Averaged across all the thresholds, students just below a round-number threshold appear about one percentage point more likely to retake the SAT than those just above that threshold. The first row of Table 3 provides the first stage estimate of that discontinuity from equation 1, suggesting that students first scoring just below a round-number threshold are 0.9 percentage points more likely to retake the SAT than those just above that threshold.

Figure 3 shows the point estimates and standard errors from running this first stage model at each individual threshold, rather than averaging them into a single estimate. Sixteen of the 17 individual thresholds generate statistically significant variation in retake rates. Consistent with the raw retaking data, we see large retaking impacts of falling below higher thresholds and smaller impacts of falling

[15]The regression specifications above also allow us to choose different bandwidths at different thresholds. The linearity of the data again mean such choices make little difference to our estimates.

below lower thresholds. Columns 2 and 3 of Table 3 average these effects across the two halves of the initial score distribution, showing that retake rates increase by 0.6 percentage points for lower scoring students and 1.4 percentage points for higher scoring students. The impacts of missing thresholds on retake rates varies less by income and race.

Missing round-number thresholds induces more than one retake attempt on average for compliers, perhaps because they have target scores that a second take fails to achieve or because they learn that retaking is less costly than they had previously believed. The second row of Table 3 runs a version of the first stage regression where the outcome is the total number of takes rather than an indicator for retaking. Such coefficients always exceed their counterparts in the first row. The third row provides instrumental variables estimates the impact of threshold-induced retaking on the total number of takes, the ratio of the first and second row coefficients. Compliers retake the SAT about 1.3 times on average, with lower scoring compliers retaking upwards of 1.4 times. The remainder of this paper uses as a treatment variable the indicator for retaking the SAT at least once because we lack independent variation in whether a student retakes once or more than once. Nonetheless, if we assume treatment effects are linearly increasing in the number of takes, subsequent estimates can simply be divided by 1.3 (or the relevant subsample coefficient) to generate a "per-retake" treatment effect.

These thresholds provide extremely strong variation in retake rates. The first stage F-statistic exceeds 180 for the entire sample and exceeds 70 in all of the subsamples we study, far above traditional thresholds for potentially weak instruments (Staiger and Stock, 1997). Though the instrument is strong, the first stage is small in magnitude, raising potential concerns that compliers may be an unusual subset of the population. The bottom panels of Table 3 compute the mean characteristics of compliers, as suggested by Abadie, Angrist and Imbens (2002) and Abadie (2003), and of the whole sample for comparison. Relative to the population of SAT-takers, compliers have higher first SAT scores, look indistinguishable in terms of income, and are somewhat less likely to be URM students. Across the subgroups, compliers generally have higher first SAT scores than the mean student, are equally or more likely to be low income, and appear similar in terms of racial composition. Though this does not exclude the possibility that compliers are unusual with respect to unobservable characteristics such as motivation, these results suggest compliers do not appear to be outliers along observable dimensions.

That students appear to have round number target scores is not surprising. The top two results from a Google search for "Should I retake the SAT?" lead to web pages that advise "you should have an exact target score in mind" and "having a concrete score in mind can be a powerful motivator."[16] Students' focus

[16]These quotes come from Prep Scholar's December 13, 2014 blog post by Allen Cheng entitled "Should You Retake the SAT or ACT? 3-Step Process" and U.S. News & World Report's March 21, 2016 blog post by Brian Witte entitled "When Retaking the SAT Makes Sense", both of which were accessed on

on round number thresholds may arise from irrational left-digit bias. It could also be a rational response to the perceived left-digit bias of admissions or financial aid processes, which may explicitly use round number thresholds to determine eligibility or implicitly overweight scores' first digits due to the mental processes of admissions officers with heavy workloads (Bowman and Bastedo, 2017).

C. Validity Checks

For the regression discontinuity approach to yield valid treatment effect estimates, students must not be able to precisely control the relationship between their first SAT score and round-number thresholds. Details of the test itself suggest no scope for such potential manipulation. The SAT is scored centrally and the scale translating raw scores into scaled scores is unknown to test-takers.

Two pieces of empirical evidence are consistent with this. First, we see no heaping at values just above or below round-number thresholds, suggesting no systematic imbalance in the number of students to the left and right of each threshold.[17] Second, students just above and below the thresholds appear quite similar in terms of the demographic characteristics we observe.[18] All available evidence suggests that students just above and below the thresholds are identical except for their retaking rates.

We provide two further pieces of evidence that students care specifically about their total SAT scores with respect to retake decisions. First, retaking rates were discontinuous at round-number thresholds on the 1600 point math and reading scale that existed prior to the introduction of the third section on writing. When that third section was introduced, retaking discontinuities as measured on the 1600 point scale largely vanished.[19] Retake rates also show no discontinuities at round-number thresholds from scores on individual sections of the SAT.[20] All of this is consistent with students primarily reacting to their total SAT scores, the measure most frequently highlighted by college admissions offices.

Whether round-number thresholds provide a valid instrument for SAT retaking depends also on whether they satisfy the exclusion restriction, affecting outcomes only through retaking and not through other channels. For outcomes such as maximum SAT scores, the exclusion restriction is clearly satisfied because retaking is the only channel through which such scores can change. For college enrollment outcomes, we note potential channels other than retaking through which such thresholds might have an effect.

August 15, 2017.

[17]See Figure 1, which plots the density of observations and appears smooth at all thresholds.

[18]Table 1 shows instrumental variable estimate from equation 2, using demographic characteristics as outcomes. Across a variety of bandwidths, we see no statistically significant relationship between threshold-driven retaking and income, race, parental education, or gender. Figure 2 shows estimated instrumental variable coefficients at individual thresholds. Differences in such characteristics are very close to zero and statistically insignificant at nearly all thresholds.

[19]See Figure 3.

[20]See Figure 4.

College admissions and financial aid processes may exhibit left-digit bias because of explicit rules or implicit biases of application readers. If so, scoring just below a threshold will disadvantage students by reducing their chance of admission or financial aid. Left-digit bias from individual colleges is unlikely to be empirically important because of the national scope of our sample and our use of so many different thresholds. We also later rule out the possibility that individual states' public postsecondary systems' admissions or financial aid rules might be driving our results. This channel suggests that missing a threshold could harm a student's college outcomes for reasons other than retaking, making our instrumental variable estimates of retaking impacts downward-biased and thus lower bounds on true effects.

Regardless of whether colleges actually exhibit left-digit bias, students may believe they do. If so, missing a threshold could discourage students from applying for admission or financial aid or, conversely, could encourage them to work harder on admission and aid applications on margins other than their SAT score. These reactions have opposing effects on college outcomes, making it ambiguous as to the sign of the bias for instrumental variable estimates of retaking impacts.

III. Results

A. Admissions Relevant SAT Scores

Retaking induced by round-number thresholds substantially increases students' SAT scores. Table 4 shows the impact of retaking on four outcomes: most recent total score after two potential takes; superscore after two potential takes; superscore after all takes; and the probability that a student's final superscore exceeds her first superscore by at least 150 points (0.5 standard deviations).[21] Column 1 shows that, on average, retakers' total scores are 46 points higher on their second take than on their first take, suggesting real gains either through learning or increased familiarity with the test. The magnitude of this round number threshold-induced gain is remarkably similar to the 44 point gain the average retaker achieves between the 1st and 2nd tests. This provides another indication that the compliers driving our estimates are not unusual relative to the average retaker.

Superscoring makes retaking even more effective, with retakers' superscores increasing by 88 points, nearly twice as much as their total scores rise. With additional retakes, superscores ultimately rise by 102 points, with 24 percent of retakers seeing superscore gains of at least 150 points. This superscore increase among compliers is again remarkably similar to the 100 point increase in superscores that the average retaker achieves after completing all takes. Comparing columns 2 and 3, the ratio of superscore gains between the final and second takes is less than the 1.3 retakes induced among compliers. This suggests diminishing

[21] Table 2 shows the reduced form impact of missing a round number threshold on SAT scores.

marginal benefits of retaking, with the magnitude of this implying that the first retake increases superscores by about twice as much as subsequent retakes. Improvements in SAT scores come from all three sections of the test, with about 50-60 percent of retakers raising their scores on any given section and 20 percent raising their score on all three sections.[22]

On average, retaking improves scores for all students but particularly for those with lower initial scores. Figure 4 shows by individual threshold the estimated impact of retaking on ultimate superscores.[23] We see statistically significant superscore gains from retaking at each threshold, with higher gains for initially lower scoring students. Panel B of Table 4 confirms this pattern. Low scoring students' second take total scores are 91 points higher than their first, relative to an only 22 point gain for higher scoring students. On average, retaking ultimately increases low scoring students' superscores by 136 points, relative to 84 points for higher scoring students. The scores of low income and URM students tend to increase more from retaking than the scores of their high income and non-URM counterparts, though this heterogeneity is lower than that by initial SAT score.[24] For all subsamples, the magnitude and statistical significance of retaking's impact on ultimate superscores is remarkably robust to choices of bandwidth, inclusion of covariates, and decision to cluster standard errors by initial SAT score.[25]

B. College Enrollment

Retaking meaningfully improves college enrollment outcomes, including four-year college enrollment rates. As shown in Table 5, retaking the SAT increases the probability that a student will enroll in a four-year college by a statistically significant 13 percentage points.[26] This represents a nearly 20 percent increase over the 67 percent enrollment rate among control compliers. Roughly half of that increase is due to a shift away from two-year colleges, while the other half comes from students who would not otherwise have enrolled in college. This magnitude seems large but aligns quite closely with another recent study of the impact of Boston's charter schools on SAT scores and college enrollment. Angrist et al. (2016) estimate that attending a charter school in Boston increased students' SAT scores by about 100 points and immediate four-year college enrollment by 13 percentage points, much of which represented substitution away from two-year colleges. Our estimated impacts of retaking on both SAT scores and college enrollment are nearly identical to these magnitudes, suggesting the enrollment increase we observe is reasonable given the SAT increase we observe.

[22]See Table 3.

[23]We exclude the lowest three thresholds (700-900) because the small number of students near these thresholds lead to noisier treatment effects that distort the scaling of the graph.

[24]Retaking does not appear to have heterogeneous impacts by gender on SAT scores or subsequent outcomes, so we do not report these results.

[25]See Table 4.

[26]Table 5 shows the reduced form impact of missing a round number threshold on college enrollment.

Retaking causes students to enroll in colleges with completion rates six percentage points higher than they would have otherwise, in part because four-year colleges have higher completion rates than two-year colleges. This improvement in college quality, as measured by graduation rates, comes from a nine percentage point increase in the probability of attending colleges with graduation rates of at least 50 percent, the roughly median graduation rate of colleges attended by SAT takers. We see no change in enrollment at colleges with graduation rates of at least 80 percent, the roughly 80th percentile of the college quality distribution. Characterizing college quality by expected income yields qualitatively similar, though noisier, results. Retaking increases expected income by a statistically insignificant $700 and the probability of attending a college with roughly median expected income ($50,000) by a marginally significant eight percentage points, but has no clear impact on the attending colleges above the 80th percentile of the expected income distribution ($65,000).

Retaking's positive impact on college enrollment outcomes is driven almost entirely by lower scoring students. Figure 5 shows threshold-specific estimates of the impact of retaking on enrollment in four-year colleges. Near the higher thresholds, retaking has little impact on four-year college enrollment. Near the lower thresholds, point estimates are large, positive and sometimes statistically significant. Panel B of Table 5 shows the pooled versions of these estimates. Retaking increases the probability that lower scoring students attend a four-year college by 33 percentage points, a near doubling from a baseline 39 percent attendance rate among compliers. Two-thirds of this increase comes from students who would otherwise have attended two-year colleges. For lower scoring students, retaking increases the graduation rate of their chosen college by 14 percentage points, driven entirely by movement from below to above median quality colleges. Retaking increases lower scoring students' expected income at ages 32-34 by a statistically significant $6,900, a nearly 20 percent increase relative to control compliers.

Conversely, we see little impact of retaking on higher scoring students' college outcomes. For such students, the estimated impacts of retaking on college enrollment and quality are generally close to zero in magnitude and never statistically significant. One possible explanation is that, based on their control complier mean outcomes, higher scoring students have less room for improvement of college outcomes. Over 80 percent of such students enroll in four-year colleges, with at least 40 percent enrolling in colleges at or above the 80th percentile of graduation rate and expected income.

C. Heterogeneity by Income and Race

Retaking improves college outcomes for low income and URM students much more than it does for high income and non-URM students. For low income students, retaking increases four-year college enrollment by 30 percentage points, nearly all of which substitutes for two-year college enrollment. Low income students who retake attend colleges with 16 percentage points higher graduation rates. They are roughly 30 percentage points more likely to attend colleges of above median quality, as measured both by graduation rates and by expected income. We see little evidence of improvement in college outcomes for high income students. Retaking has roughly double the impact on the four-year college enrollment rates of URM students than of non-URM students, raising the former by 20 percentage points and the latter by eight percentage points. Heterogeneity in the impact of retaking on college by race is less clear than by income or initial score.

The heterogeneous impacts of retaking by income are not simply explained by the fact that low income students have lower initial scores on the SAT than do high income students. To show this, Table 6 explores heterogeneity by income and race among only lower scoring students. Panel A suggests that, even among only those with initially lower scores, retaking improves college enrollment and quality much more for low income students than for high income students. Conversely, panel B suggests that heterogeneity in retaking's impacts by race are largely explained by racial differences in initial SAT scores. Among lower scoring students, retaking improves the four-year college enrollment rates of both URM and non-URM students by about 30 percentage points. Point estimates on college quality measures suggest, if anything, greater benefits of retaking for non-URM students in this lower scoring sample.[27]

D. *Robustness Checks and Degree Completion*

The estimated impacts of retaking on college enrollment and quality are robust to changing the empirical choices made in our baseline specification. Figures 5 and 6 show the sensitivity to bandwidth choice of the estimated impacts for the entire sample and the three subsamples of most interest. The magnitude and statistical significance of retaking's impact on four-year college enrollment and chosen college's graduation rate are stable across bandwidths ranging from 30 to 100 SAT points, the maximum bandwidth usable between thresholds. Tables 7 and 8 show that, across the whole sample and all subsamples, estimated impacts of retaking are robust to: choice of bandwidth; flexibly controlling for gender, race, family income and parental education through fully-interacted fixed effects; and choice to cluster standard errors by first SAT score or not.

These enrollment impacts are likely not confounded by the use of round-number thresholds in state public colleges' admissions or financial aid processes. In theory, publicly known admissions thresholds could confound our estimates given that Goodman, Hurwitz and Smith (2017) show, in the context of Georgia's public four-year colleges, that some students retake specifically to gain admission to a particular college or sector. Though some states use SAT score thresholds

[27]Table 6 explores heterogeneity by income and race among only higher scoring students. Point estimates are larger for low income students than high income students among higher scorers but the standard errors are too large to infer much with respect to heterogeneity.

to partly determine admissions or merit aid eligibility, few of these states use round-number thresholds. Furthermore, all such eligibility criteria that we could identify, including Georgia, are based on scores from just the math and reading sections of the SAT, not the three section total score studied here. In case we have overlooked an empirically important state program that might violate our exclusion restriction, we generate estimated impacts excluding each of the eight largest states in our data one at a time. Table 9 shows that no single state is responsible for our primary results, making it unlikely that state-specific policies are confounding our estimates of the impact of retaking on college enrollment.

We cannot infer much about degree completion, as seen in Table 10. We can observe six-year completion rates for only the first three cohorts in the sample, which weakens our instruments for the subgroups whose enrollment rates are positively affected. Though estimated impacts on degree completion are noisy, they are generally quite similar in magnitude to the estimated impacts on expected completion based on college-level graduation rates, particularly for lower scoring and low income students.

E. College Application Behavior

Retaking the SAT boosts four-year college enrollment rates, in large part by shifting enrollment from two-year to four-year institutions. There are two primary explanations for why this might occur. First, by increasing admissions-relevant SAT scores, retaking may strengthen students' college applications and thus increase their chances of admission at four-year institutions that are at least somewhat selective. Second, retaking may change the set of colleges to which students apply. This might occur either because each retake allows students to send their SAT scores to up to four colleges for free or because students' higher scores change their expectations about where they might be admissible.

Though we can not observe admissions decisions, we do observe application behavior as proxied by score sends. Table 7 shows instrumental variables estimates of the impact of retaking on the number of colleges to which students send their SAT scores. We divide colleges into three categories corresponding to important enrollment margins observed earlier, those with graduation rates below 50 percent, between 50 and 80 percent, and above 80 percent. The first two columns show little evidence of changing application behavior to the colleges where we previously observed changed enrollment rates. Point estimates are generally small, negative and statistically insignificant, particularly for the disadvantaged subsamples where we observe increased enrollment.[28] This suggests that retaking shifts

[28]In our data, each additional score send to colleges with graduation rates below 80 percent is associated with a 3.2 percentage point increase in the probability of attending a four-year college (controlling for a cubic in first SAT scores). To understand how much reduced application numbers due to missing round number thresholds might downward bias our enrollment estimates, we multiply this by the 0.1-0.6 decrease in score sends observed in total across those first two columns for various subgroups. This suggests that reduced applications would drive only a one percentage point decrease in overall

students from below average to above average colleges not by changing application behavior but by increasing the strength of applications students would have sent anyway.

We observe reductions in applications to colleges with the highest graduation rates. We suspect this is not the impact of retaking. Instead, scoring below round-number thresholds may discourage students from applying to highly selective colleges where they perceive such small score differences can reduce their admissions chances (Bond et al., 2018). This fact is less relevant to lower scoring students, who rarely apply to or enroll in highly selective colleges and for whom retaking's effects appear largely on the margin between non-selective two-year and less selective four-year colleges. Reduced applications to highly selective colleges suggest, however, that our instrument may cause us to underestimate the impact of retaking for the college choices of higher scoring students.[29]

IV. Gaps in SAT Scores and College Enrollment

A. *SAT Score Gaps*

We use the estimated impacts of retaking on SAT scores to model what portion of income and racial gaps in admissions-relevant SAT scores would be closed by eliminating disparities in retake rates. The main assumption required to model this is that the marginal students induced to retake by a future intervention would have similar test score gains to the average complier. This assumption seems plausible given that compliers have similar test score gains as the average retaker, though newly induced retakers could differ in ability or willingness to study or otherwise prepare for the second take.

Raising the retake rate of low income students to that of high income students would close about nine percent of the admissions-relevant SAT score gap by income. To show this, we begin with the fact from Table 2 that low income students are 20.6 percentage points less likely than high income students to retake. Our estimates in Table 4 imply that retaking once increases the superscores of low income students by 109 points. Inducing 20.6 percent of low income students to retake once would thus boost the average superscores of low income students by 23 points (0.206 * 109). This represents nine percent of the 250 point gap in superscores between low and high income students. This lower 227 point income gap we estimate under equalized retake rates is remarkably close to the actual 226

enrollment, even less for disadvantaged subgroups, and less than a two percentage point decrease for advantaged subgroups. This would not materially affect our conclusions about retaking's impact on college enrollment.

[29]Each additional score send to colleges with graduation rates above 80 percent is associated with a 2.6 percentage point increase in the probability of attending a four-year college (controlling for a cubic in first SAT scores). Multiplying this by the observed decreases in score sends to such colleges suggests that reduced applications would drive a four percentage point decrease in overall enrollment, or a three percentage point decrease for low income and URM students. The latter number is quite small relative to the estimated impact of retaking on those subgroups, again suggesting the bias from this channel does not meaningfully alter our conclusions.

point gap in first SAT scores.[30] This suggests that our estimates of the extent to which retaking contributes to income-based gaps in SAT scores is reasonable.

A similar exercise suggests that equalizing retakes rates by race would close about four percent of the gap in admissions-relevant SAT scores between URM and non-URM students. URM students are 11.3 percentage points less likely to retake than non-URM students and retaking once increases the superscores of URM students by 94 points. Inducing 11.3 percent of URM students to retake once would boost the average superscores of URM students by 11 points (0.113 * 94), which represents about four percent of the 269 point gap in superscores between URM and non-URM students. Again, this lower 258 point gap we estimate under equalized retake rates is fairly close to the actual 252 point gap between URM and non-URM students in first SAT scores.

B. College Enrollment Gaps

We also use the estimated impacts of retaking to model what portion of income and racial gaps in four-year college enrollment would be closed by eliminating disparities in retake rates. This requires two assumptions. First, as with our SAT score gap analysis, we assume that the marginal students induced to retake by an intervention would have similar enrollment gains as the average complier here. Second, we ignore general equilibrium effects that might result from population-wide increases in retake rates. Such effects might come from colleges raising admission standards in reaction to higher-scoring applicant pools or from high income and non-URM students increasing their retake rates or numbers of retakes due to increased competition in the admissions process. We begin by calculating the impact of equalizing retake rates on income and race gaps among the SAT-taking population, which requires only the College Board data described previously. We then bring in one additional data set, the National Center for Education Statistics' High School Longitudinal Study of 2009 (HSLS), to estimate the impact on closing gaps among all high school graduates, the full population among which college enrollment gaps are typically estimated.

Raising the retake rates of low income students to that of high income students would close about 18-25 percent of the income gap among SAT takers in four-year college enrollment. To show this, we again begin with the fact from Table 2 that low income SAT-takers are 20.6 percentage points less likely than high income SAT-takers to retake. Our estimates in Table 5 imply that the 1.4 retakes induced by round-number thresholds in low income SAT-takers increase their four-year college enrollment rate by 29.9 percentage points. Inducing 20.6 percent of low income SAT-takers to retake 1.4 times would thus boost the average four-year college enrollment rate of low income SAT-takers by six percentage points (0.206 * 29.9). This represents 25 percent of the 24.9 percentage point enrollment

[30] For detailed income and race gaps in retake rates and SAT scores, see Table 11.

gap between low and high income SAT-takers.[31] If the impact of retaking grows linearly with the number of retakes, then inducing low income SAT-takers to retake once at the same rate as high income SAT-takers would close the enrollment gap by about 18 percent (25/1.4). Given that first retakes have larger impacts on SAT scores than subsequent retakes, the true impact of inducing one additional retake in low income SAT-takers is likely somewhere between these 18 and 25 percent estimates.

We similarly estimate that equalizing retakes rates by race would close 10-14 percent of the race gap among SAT-takers in four-year college enrollment. URM SAT-takers are 11.3 percentage points less likely to retake than non-URM SAT-takers and the 1.4 retakes induced by round-number thresholds increase their four-year college enrollment rate by 20.4 percentage points. Inducing 11.3 percent of URM SAT-takers to retake 1.4 times would boost the average four-year college enrollment rate of URM SAT-takers by two percentage points (0.113 * 20.4). This represents about 14 percent of the 16.3 percentage point enrollment gap between URM and non-URM SAT-takers. Our best per-retake estimate suggests inducing URM SAT-takers to retake once at the same rate as non-URM SAT-takers would close the enrollment gap by about 10 percent (14/1.4). The true impact of inducing one additional retake in URM SAT-takers is likely somewhere between these 10 and 14 percent estimates.

To compute how equalizing retakes rates among SAT-takers affect college enrollment gaps among all high school graduates, we need to supplement our data with information on high school graduates who do not take a college entrance exam. We use HSLS, which allows us to construct a nationally representative sample of the high school class of 2012. This in turn provides us with estimates of the fraction of students (by race and income) who do not take a college entrance exam at all, as well as the four-year college enrollment rates of such students. HSLS does not distinguish between students who take the SAT and the ACT, the other major college entrance exam, so we make the additional assumption that estimates on the impact of SAT retaking would apply to students retaking the ACT.

The extent to which equalizing retake rates closes college enrollment gaps shrinks by half after incorporating students who do not take college entrance exams, because we assume efforts to increase retaking would not affect non-takers. Estimates from HSLS suggest that 58 percent of low income high school graduates never take the SAT or ACT, compared to 39 percent of their high income peers. Similarly, 57 percent of URM high school graduates never take college entrance exams, compared to 49 percent of their non-URM peers. Incorporating these facts, plus information from HSLS on the college enrollment rates of these non-takers, implies that equalizing retake rates by income would shrink the income gap in four-year college enrollment among all high school graduates by 10 percent. Similarly, equalizing retake rates by race would shrink the racial gap in

[31]For detailed income and race gaps in four-year college enrollment, see Table 11.

four-year college enrollment by seven percent.

These estimated enrollment gap closures among all high school graduates are roughly half the size of those estimated within the SAT-taking population. That difference can be explained by two facts documented in HSLS. First, only about half of high school graduates take a college entrance exam. Second, race- and income-based college enrollment gaps among those who do not take college entrance exams are fairly similar to the gaps among exam-takers. As a result, the effect of equalizing retake rates on gaps as measured among all high school graduates is a nearly evenly weighted average between the effect on the gap within the exam-taking population and zero, the effect on the non-taking population of high school graduates.

V. Conclusion

Millions of American students take college entrance exams each year but only half choose to retake such exams. We provide the first causal evidence that retaking can substantially improve the college enrollment outcomes of students, particularly for those who are initially low-scoring or traditionally underrepresented in higher education. Retaking appears to improve college enrollment by increasing admissions-relevant test scores and likely improving the odds of admission. Though we cannot identify the precise reason why retake scores tend to be higher than first scores, our empirical strategy eliminates the possibility that such gains are driven solely by selection into retaking of students who expect to improve. The specific channel through which retaking improves scores does not affect the conclusion that many students who take the SAT only once would benefit from taking it at least one more time. That meaningful portions of college enrollment gaps by income and race could be closed by eliminating disparities in retake rates suggests that efforts to encourage retaking among disadvantaged populations might be worth pursuing.[32]

How might additional retaking be encouraged? Intervening to encourage students to take their first SAT at an earlier date seems particularly promising. Our data suggest that earlier first takes are strongly associated with increased retaking rates. Low income and URM students are substantially more likely to first take the SAT in 12th grade rather than in the 11th grade, at which point they have little opportunity to retake prior to college application deadlines. Encouraging earlier first takes would give students additional time to assess the strength of their college applications and retake the SAT if they thought it helpful.

Intervening to make retakes less costly might also be effective. Our descriptive evidence is consistent with testing fees discouraging retaking, so that efforts to

[32]An alternative way of equalizing retake rates is to prohibit retaking entirely and limit all students to one SAT attempt. We do not focus on this alternative because increasing the stakes associated with any college entrance exam would generate substantial opposition and is thus unlikely to be implemented. It would also likely generate substantial behavioral changes among test-takers, the effects of which would be hard to predict.

lower such fees or improve take-up rates of fee waivers might be fruitful. Many low income students do not use the available fee waiver. In our data, 43 percent of students with self-reported family income under $30,000 do not use a fee waiver. Retake rates might be improved by making the fee waiver application process more transparent or by sending automated reminders that retakes are free for those who used a fee waiver on the first take. Financial barriers cannot, however, fully explain low retaking rates given that 53 percent of students who use a fee waiver on their first take do not then retake, even though that second take would be free.

A lack of clarity about the test-taking and admissions processes suggests that informational campaigns could boost retaking rates. Evidence from web searches suggest that many students do not know whether they can retake the SAT or do not understand the potential benefits of retaking, particularly if colleges' use of multiple scores in admissions is unclear. Typing "retaking the SAT is" into Google yields two auto-complete suggestions: "Is retaking the SAT bad?" and "Is it worth retaking the SAT?", suggesting that these are two common questions students have about retaking. Typing "retake SAT" into Google Trends shows the top related query is "Can you retake the SAT?", implying that some students are unclear about whether retaking is even an option.[33]

Much more research could be done on the phenomenon of retaking and the college admissions processes that incentivize it. We have no evidence on interventions designed to increase retaking rates and retaking is very rarely the primary focus of research papers.[34] We also know little about how or why colleges choose to admit students on the basis of their entire test score history, their single highest score on a given take, or their highest combined score across takes. Such institutional decisions bear directly on students' incentives to retake college entrance exams. Anecdotal evidence from admissions officers suggests that college rankings publications incentivize colleges to admit students on the basis of superscores, mechanically raising that component of the ranking. Research to study individual colleges' incentives to adopt one scheme over another would be helpful, given that Vigdor and Clotfelter (2003) find the most common choice made by colleges is also "the costliest, least accurate, and most biased" among all available options.

Finally, the large retaking benefits we document are partial equilibrium effects comparing current retakers to non-retakers. Whether the general equilibrium effects of policies to broadly increase retaking rates resemble these estimates depends in part on whether the supply of college slots is fixed (Krishna, Lychagin and Robles, 2015). Evidence suggests that the postsecondary sector is often slow to react to changing demand for college education, in part because public funding lags such changes (Bound, Lovenheim and Turner, 2010, 2012; Kelly, 2016).

[33]Both Google queries were conducted on July 16, 2018.

[34]Google Scholar searches for papers with "retaking" or "retake" and "SAT", "ACT", or "college" in the title yield no peer-reviewed academic articles on the topic other than Vigdor and Clotfelter (2003).

If colleges do not expand enrollment, then retaking changes which students gain admission to and enroll in college but not how many. If enrollment expands faster than per-pupil funding, such additional enrollment may not translate into degree completion (Bound and Turner, 2007; Deming and Walters, 2017). The net impact of interventions to increase retaking rates depends heavily on the broader landscape of higher education policy.

REFERENCES

Abadie, Alberto. 2003. "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics*, 113(2): 231–263.

Abadie, Alberto, Joshua Angrist, and Guido Imbens. 2002. "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings." *Econometrica*, 70(1): 91–117.

Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu. 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science*, 63(6): 1657–1672.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters. 2016. "Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice." *Journal of Labor Economics*, 34(2): 275–318.

Bond, Timothy N, George Bulman, Xiaoxiao Li, and Jonathan Smith. 2018. "Updating human capital decisions: Evidence from SAT score shocks and college applications." *Journal of Labor Economics*, 36(3): 807–839.

Bound, John, and Sarah Turner. 2007. "Cohort crowding: How resources affect collegiate attainment." *Journal of Public Economics*, 91(5): 877–899.

Bound, John, Michael F Lovenheim, and Sarah Turner. 2010. "Why have college completion rates declined? An analysis of changing student preparation and collegiate resources." *American Economic Journal: Applied Economics*, 2(3): 129–157.

Bound, John, Michael F Lovenheim, and Sarah Turner. 2012. "Increasing time to baccalaureate degree in the United States." *Education Finance and Policy*, 7(4): 375–424.

Bowman, Nicholas A, and Michael N Bastedo. 2017. "What Role May Admissions Office Diversity and Practices Play in Equitable Decisions?" *Research in Higher Education*, 1–18.

Bulman, George. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics*, 7: 1–36.

Busse, Meghan R, Nicola Lacetera, Devin G Pope, Jorge Silva-Risso, and Justin R Sydnor. 2013. "Estimating the Effect of Salience in Wholesale and Retail Car Markets." *American Economic Review: Papers and Proceedings*, 103: 575–579.

Canaan, Serena, and Pierre Mouganie. 2018. "Returns to education quality for low-skilled students: Evidence from a discontinuity." *Journal of Labor Economics*, 36(2): 395–436.

Chetty, Raj. 2015. "Behavioral economics and public policy: A pragmatic perspective." *The American Economic Review*, 105(5): 1–33.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." National Bureau of Economic Research Working Paper 23618.

Cohodes, Sarah R, and Joshua S Goodman. 2014. "Merit aid, college quality, and college completion: Massachusetts' Adams Scholarship as an in-kind subsidy." *American Economic Journal: Applied Economics*, 6(4): 251–285.

Deming, David, and Chris Walters. 2017. "The impacts of price and spending subsidies on US postsecondary attainment." National Bureau of Economic Research Working Paper 23736.

Foellmi, Reto, Stefan Legge, and Lukas Schmid. 2016. "Do Professionals Get It Right? Limited Attention and Risk-taking Behaviour." *The Economic Journal*, 126(592): 724–755.

Frisancho, Veronica, Kala Krishna, Sergey Lychagin, and Cemile Yavas. 2016. "Better luck next time: Learning through retaking." *Journal of Economic Behavior and Organization*, 125: 120–135.

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith. 2017. "Access to 4-Year Public Colleges and Degree Completion." *Journal of Labor Economics*, 35(3): 829–867.

Goodman, Sarena. 2016. "Learning from the Test: Raising Selective College Enrollment by Providing Information." *Review of Economics and Statistics*, 98: 671–684.

Hurwitz, Michael, Jonathan Smith, Sunny Niu, and Jessica Howell. 2015. "The Maine Question: How Is 4-Year College Enrollment Affected by Mandatory College Entrance Exams?" *Educational Evaluation and Policy Analysis*, 37: 138–159.

Hyman, Joshua. 2017. "ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice." *Education Finance and Policy*, 12: 281–311.

Kelly, Andrew P. 2016. "The Supply Side of College Match." In *Matching Students to Opportunity: Expanding College Choice, Access, and Quality.* , ed. Andrew P Kelly, Jessica S Howell and Carolyn Sattin-Bajaj, Chapter 5, 99–119. Cambridge, MA:Harvard Education Press.

Klasik, Daniel. 2013. "The ACT of Enrollment: The College Enrollment Effects of State-Required College Entrance Exam Testing." *Educational Researcher*, 42: 151–160.

Krishna, Kala, Sergey Lychagin, and Verónica Frisancho Robles. 2015. "Retaking in High Stakes Exams: Is Less More?" National Bureau of Economic Research Working Paper 21640.

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor. 2012. "Heuristic Thinking and Limited Attention in the Car Market." *The American Economic Review*, 102: 2206–2236.

Lavecchia, Adam M, Heidi Liu, Philip Oreopoulos, et al. 2016. "Behavioral Economics of Education." *Handbook of the Economics of Education*, 5: 1–74.

Lee, David S, and David Card. 2008. "Regression discontinuity inference with specification error." *Journal of Econometrics*, 142(2): 655–674.

Levitt, Steven D, John A List, Susanne Neckermann, and Sally Sadoff. 2016. "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance." *American Economic Journal: Economic Policy*, 8(4): 183–219.

McFarland, Joel, and Bill Hussar. 2018. "The Condition of Education 2018." *National Center for Education Statistics.*

Page, Lindsay C, and Judith Scott-Clayton. 2016. "Improving college access in the United States: Barriers and policy responses." *Economics of Education Review*, 51: 4–22.

Pallais, Amanda. 2015. "Small differences that matter: Mistakes in applying to college." *Journal of Labor Economics*, 33(2): 493–520.

Pope, Devin, and Uri Simonsohn. 2011. "Round numbers as goals: Evidence from baseball, SAT takers, and the lab." *Psychological science*, 22(1): 71–79.

Smith, Jonathan. 2018. "The Sequential College Application Process." *Education Finance and Policy*, 13(4): 545–575.

Staiger, Douglas, and James H Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica: Journal of the Econometric Society*, 557–586.

Vigdor, Jacob L, and Charles T Clotfelter. 2003. "Retaking the SAT." *The Journal of Human Resources*, 38: 1–33.

Zimmerman, Seth D. 2014. "The returns to college admission for academically marginal students." *Journal of Labor Economics*, 32(4): 711–754.

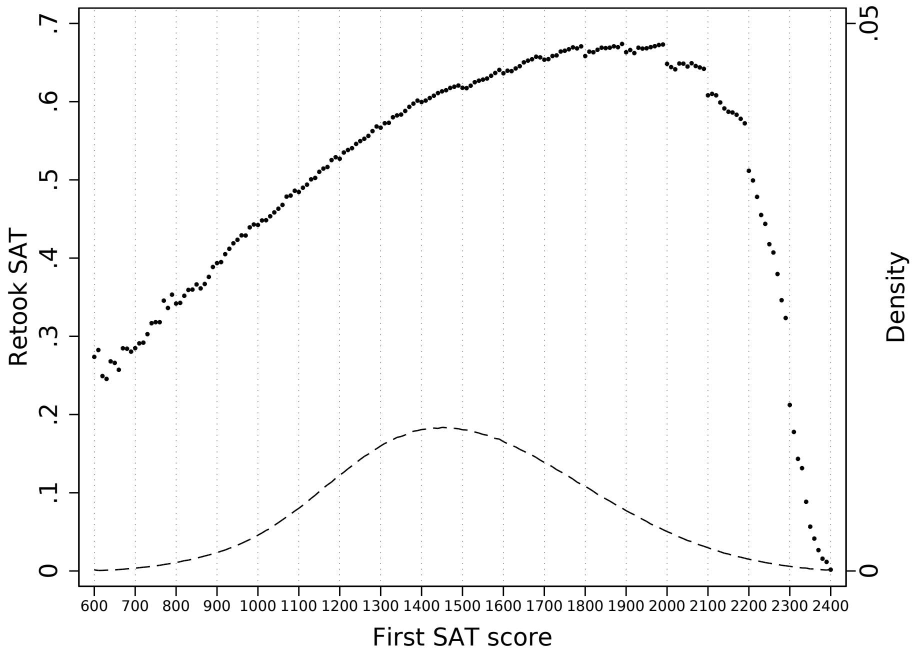

Figure 1. : Round Number Thresholds and SAT Retaking

Note: Shown above are average SAT retaking rates as a function of students' first SAT scores. The sample consists of all SAT-takers from the high school classes of 2006-14 with valid math, reading and writing scores, and who took their first SAT by November of senior year. The dashed line shows the density of observations.

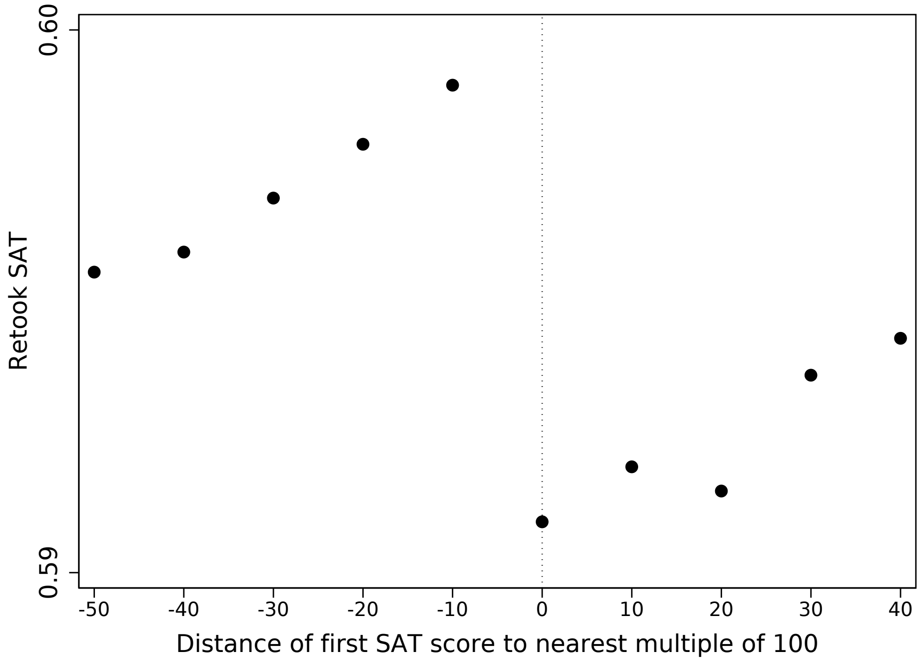

Figure 2. : Retake Rates Averaged across All Thresholds

Note: Shown above are average SAT retaking rates as a function of students' first SAT score distance from the nearest multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 with valid math, reading and writing scores, and who took their first SAT by November of senior year.

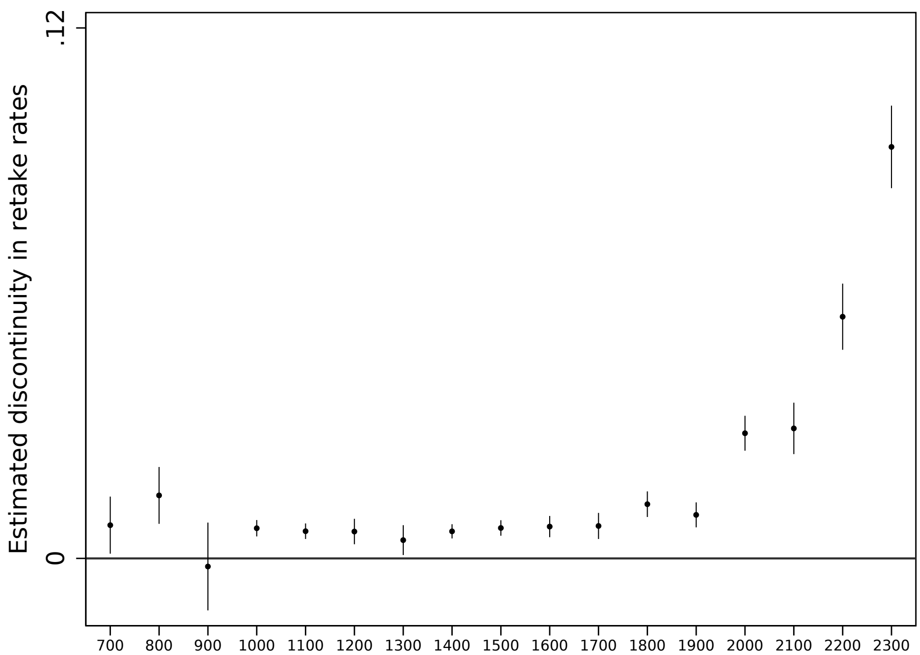

Figure 3. : Retaking Effects at Each Threshold

Note: Shown above are estimated discontinuities in retaking due to one's first SAT score falling below a given multiple of 100. Estimates come from the local linear model described in the first stage Equation 1 but run separately for each threshold.

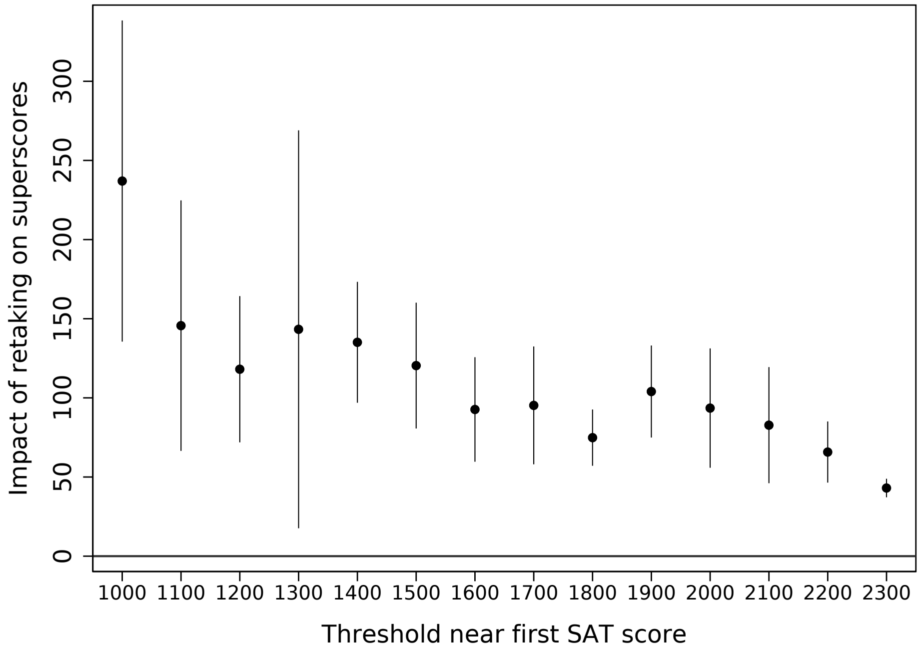

Figure 4. : Retaking and Admissions-Relevant SAT Scores

Note: Shown above are instrumental variables estimates at each threshold of the impact of retaking on SAT superscores. Estimates come from the instrumental variable model described in Equation 2 but run separately for each threshold. The lowest thresholds (700-900) are omitted for scaling purposes.

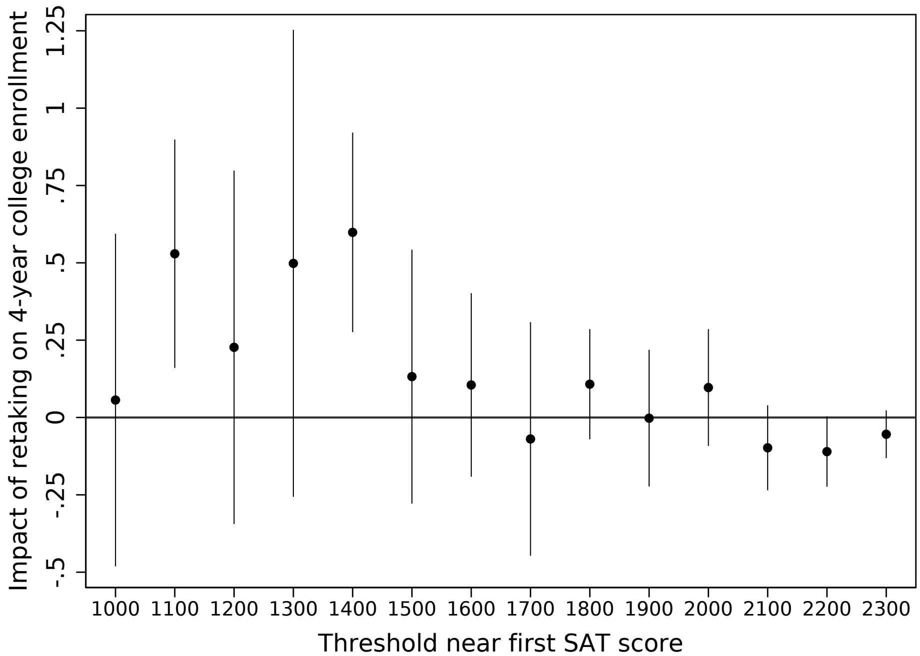

Figure 5. : Retaking and Four-Year College Enrollment

Note: Shown above are instrumental variables estimates at each threshold of the impact of retaking on four-year college enrollment. Estimates come from the instrumental variable model described in Equation 2 but run separately for each threshold. The lowest thresholds (700-900) are omitted for scaling purposes.

Table 1—: Summary Statistics

		Regression discontinuity sample			
	All students (1)	All students (2)	Lower scoring (3)	Low income (4)	URM (5)
(A) Demographics					
Female	0.53	0.54	0.56	0.59	0.57
White	0.56	0.58	0.50	0.37	0.00
URM	0.27	0.24	0.34	0.43	1.00
Asian	0.11	0.11	0.09	0.14	0.00
Low income	0.21	0.19	0.23	1.00	0.35
Fee waiver	0.20	0.18	0.25	0.44	0.46
(B) SAT-taking					
First SAT score	1475	1500	1319	1391	1328
SAT superscore	1531	1562	1376	1442	1378
Retook SAT	0.54	0.61	0.56	0.53	0.56
SAT takes	1.74	1.82	1.76	1.69	1.70
Months available to retake	12.4	13.5	12.6	12.4	12.3
(C) College enrollment					
4-year college	0.57	0.61	0.51	0.51	0.52
2-year college	0.18	0.17	0.23	0.21	0.21
College's graduation rate	0.45	0.48	0.39	0.39	0.38
College's mean income (000s)	47.0	48.8	41.5	43.4	43.7
N (millions)	13.66	11.99	7.58	2.23	3.07

Note: Listed above are mean values of key variables. Column 1 consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores. Columns 2-5 limit that sample to students who took their first SAT by November of senior year. Columns 3-5 respectively include students whose first score was nearest to a threshold below 1600, those with family income below $50,000, and underrepresented minority (Black, Hispanic, or Native American) students.

Table 2—: Predictors of SAT Retaking

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Income							
Low income	-0.206***			-0.196***	-0.164***	-0.171***	-0.102***
	(0.002)			(0.002)	(0.002)	(0.003)	(0.002)
Middle income	-0.112***			-0.108***	-0.091***	-0.091***	-0.055***
	(0.002)			(0.002)	(0.001)	(0.001)	(0.001)
(B) Race							
URM		-0.093***		-0.057***	-0.010***	-0.017***	0.013***
		(0.003)		(0.002)	(0.003)	(0.004)	(0.003)
Asian		0.119***		0.141***	0.131***	0.129***	0.108***
		(0.017)		(0.016)	(0.016)	(0.016)	(0.009)
(C) Gender							
Female			0.033***	0.042***	0.045***	0.045***	0.033***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(D) SAT factors							
First score (100s)					0.021***	0.022***	0.006***
					(0.001)	(0.001)	(0.000)
Fee waiver						0.026***	0.048***
						(0.006)	(0.003)
Months to retake							0.044***
							(0.001)
R^2	0.02	0.02	0.00	0.03	0.05	0.05	0.20

Note: Heteroskedasticity robust standard errors clustered by high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each column regresses an indicator for retaking the SAT on the variables shown. Low income students are those with family incomes below $50,000 and middle income students have family incomes between $50,000 and $100,000. Underrepresented minority (URM) students are those who identify as Black, Hispanic, or Native American. Months to retake represents the number of months between a student's first SAT take and June of senior year. All regressions include cohort fixed effects, as well as indicators for missing income or race, so that high income and White students are the reference groups. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores. Each regression uses 13,656,612 observations.

Table 3—: Round-Number Thresholds and SAT Retaking

	All students (1)	Lower scoring (2)	Higher scoring (3)	Low income (4)	High income (5)	URM (6)	Non-URM (7)
Retook (FS)	0.009***	0.006***	0.014***	0.008***	0.010***	0.008***	0.010***
	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Takes (RF)	0.012***	0.008***	0.017***	0.012***	0.014***	0.011***	0.013***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Takes (IV)	1.331***	1.438***	1.275***	1.436***	1.383***	1.394***	1.284***
	(0.053)	(0.131)	(0.040)	(0.139)	(0.107)	(0.079)	(0.047)
First stage F-statistic	180.6	144.1	82.9	71.3	83.4	89.6	149.8
Complier characteristics							
First SAT score	1682	1265	1912	1520	1817	1440	1757
Low income	0.19	0.33	0.12			0.42	0.13
URM	0.22	0.35	0.14	0.44	0.17		
Mean characteristics							
First SAT score	1499	1292	1774	1391	1599	1328	1569
Low income	0.19	0.24	0.12			0.33	0.14
URM	0.26	0.36	0.12	0.45	0.12		
N	14,368,305	8,201,406	6,166,899	2,677,386	2,661,691	3,677,470	9,810,789

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient in the first two rows is an estimate of the impact of scoring below a multiple of 100 on retaking behavior. Each coefficient in the third row is an instrumental variables estimate of the impact of initial retaking on the total number of retakes, SAT retaking, where initial retaking is instrumented with an indicator for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Columns 2 and 3 split the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Columns 4 and 5 split the sample into students with family incomes below $50,000 and above $100,000. Columns 6 and 7 split the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian). The bottom panels show complier and mean characteristics for each subgroup.

Table 4—: Retaking and SAT Scores

	After two potential takes		After all takes	
	Most recent total score (1)	Superscore (2)	Superscore (3)	Superscore increase of 150+ points (4)
(A) All students				
All	46.4***	87.7***	101.9***	0.237***
	(7.9)	(5.0)	(5.9)	(0.035)
(B) By initial score				
Lower scoring	90.7***	114.1***	136.1***	0.438***
	(13.3)	(9.3)	(11.8)	(0.068)
Higher scoring	22.4**	73.3***	83.6***	0.128***
	(9.0)	(5.3)	(5.4)	(0.035)
(C) By income				
Low income	72.1***	108.9***	118.9***	0.397***
	(16.3)	(12.2)	(15.5)	(0.088)
High income	39.4**	81.4***	99.5***	0.168**
	(16.2)	(12.3)	(14.1)	(0.081)
(D) By race/ethnicity				
URM	59.5***	93.8***	106.1***	0.296***
	(15.5)	(11.6)	(13.0)	(0.072)
Non-URM	37.5***	83.8***	97.5***	0.213***
	(9.2)	(6.1)	(6.8)	(0.042)

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 5—: Retaking and College Enrollment

	College type		College's graduation rate			College's mean income		
	Four-year	Two-year	Overall	>50%	>80%	Overall	> $50,000	> $65,000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) All students								
All	0.125***	-0.060	0.062**	0.088**	0.001	0.724	0.081*	-0.036
	(0.042)	(0.037)	(0.024)	(0.038)	(0.030)	(2.104)	(0.044)	(0.035)
Control complier mean	0.67	0.15	0.53	0.61	0.27	55.11	0.51	0.29
(B) By initial score								
Lower scoring	0.329***	-0.209**	0.138***	0.182**	0.003	6.877***	0.150*	0.053
	(0.088)	(0.081)	(0.042)	(0.070)	(0.050)	(2.561)	(0.077)	(0.051)
Control complier mean	0.39	0.33	0.32	0.34	0.04	37.25	0.21	0.04
Higher scoring	0.014	0.020	0.021	0.038	0.001	-2.509	0.045	-0.083*
	(0.040)	(0.029)	(0.029)	(0.043)	(0.038)	(2.941)	(0.053)	(0.047)
Control complier mean	0.82	0.04	0.65	0.76	0.40	64.89	0.68	0.42
(C) By income								
Low income	0.299***	-0.238***	0.155***	0.308***	0.076	3.242	0.275***	0.027
	(0.109)	(0.088)	(0.058)	(0.091)	(0.073)	(4.243)	(0.102)	(0.078)
Control complier mean	0.51	0.25	0.40	0.43	0.15	47.66	0.36	0.21
High income	0.025	-0.093	0.004	0.044	0.015	-1.821	0.131	-0.066
	(0.084)	(0.071)	(0.050)	(0.078)	(0.079)	(4.311)	(0.083)	(0.083)
Control complier mean	0.76	0.13	0.62	0.72	0.38	61.02	0.59	0.35
(D) By race/ethnicity								
URM	0.204**	-0.192**	0.074*	0.097	0.019	-0.456	0.038	-0.037
	(0.089)	(0.077)	(0.044)	(0.069)	(0.050)	(3.186)	(0.072)	(0.050)
Control complier mean	0.55	0.29	0.43	0.47	0.15	48.71	0.42	0.16
Non-URM	0.084*	-0.032	0.046	0.066	-0.014	0.543	0.083	-0.028
	(0.046)	(0.035)	(0.030)	(0.045)	(0.037)	(2.369)	(0.053)	(0.041)
Control complier mean	0.71	0.11	0.57	0.66	0.31	57.20	0.55	0.33

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 6—: Retaking and College Enrollment among Lower Scoring Students

	College type		College's graduation rate			College's mean earnings		
	Four-year	Two-year	Overall	>50%	>80%	Overall	> $50,000	> $65,000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) By income								
Low income	0.407**	-0.396***	0.171**	0.418***	0.009	7.292	0.302**	0.054
	(0.169)	(0.130)	(0.081)	(0.114)	(0.110)	(4.994)	(0.145)	(0.095)
Control complier mean	0.27	0.43	0.26	0.20	0.03	35.97	0.17	0.04
High income	0.023	-0.319	-0.079	-0.001	-0.019	5.968	0.217	0.278**
	(0.355)	(0.321)	(0.171)	(0.325)	(0.151)	(9.187)	(0.273)	(0.138)
Control complier mean	0.50	0.49	0.44	0.45	0.05	40.64	0.36	0.02
(B) By race/ethnicity								
URM	0.320**	-0.287**	0.094	0.116	0.001	3.410	0.050	-0.020
	(0.137)	(0.113)	(0.066)	(0.099)	(0.056)	(3.885)	(0.103)	(0.051)
Control complier mean	0.34	0.40	0.28	0.28	0.02	35.75	0.21	0.03
Non-URM	0.296***	-0.202**	0.150**	0.200*	0.016	9.228**	0.204	0.124
	(0.110)	(0.099)	(0.067)	(0.103)	(0.085)	(4.434)	(0.126)	(0.084)
Control complier mean	0.44	0.32	0.36	0.39	0.06	38.83	0.23	0.05

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who took their first SAT by November of senior year and whose first score was nearest to a threshold of at most 1500. Panel A splits the sample into students with family incomes below $50,000 and above $100,000. Panel B splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 7—: Retaking and College Applications

	Score sends, by college graduation rate		
	0-50% (1)	50-80% (2)	80-100% (3)
(A) All students			
All	-0.161*	-0.213	-1.654***
	(0.094)	(0.183)	(0.263)
Control complier mean	0.67	2.03	3.15
(B) By initial score			
Low scoring	-0.093	-0.059	-0.566
	(0.222)	(0.366)	(0.375)
Control complier mean	0.76	2.03	0.76
High scoring	-0.198***	-0.296	-2.241***
	(0.075)	(0.198)	(0.310)
Control complier mean	0.63	2.02	4.47
(C) By income			
Low income	-0.077	-0.242	-1.193**
	(0.325)	(0.397)	(0.578)
Control complier mean	0.90	2.13	1.94
High income	-0.242	-0.289	-1.777***
	(0.199)	(0.345)	(0.664)
Control complier mean	0.58	2.26	4.15
(D) By race/ethnicity			
URM	-0.099	0.144	-1.237**
	(0.263)	(0.384)	(0.521)
Control complier mean	0.86	2.11	2.01
Non-URM	-0.183*	-0.387*	-1.916***
	(0.094)	(0.201)	(0.293)
Control complier mean	0.60	2.05	3.48

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

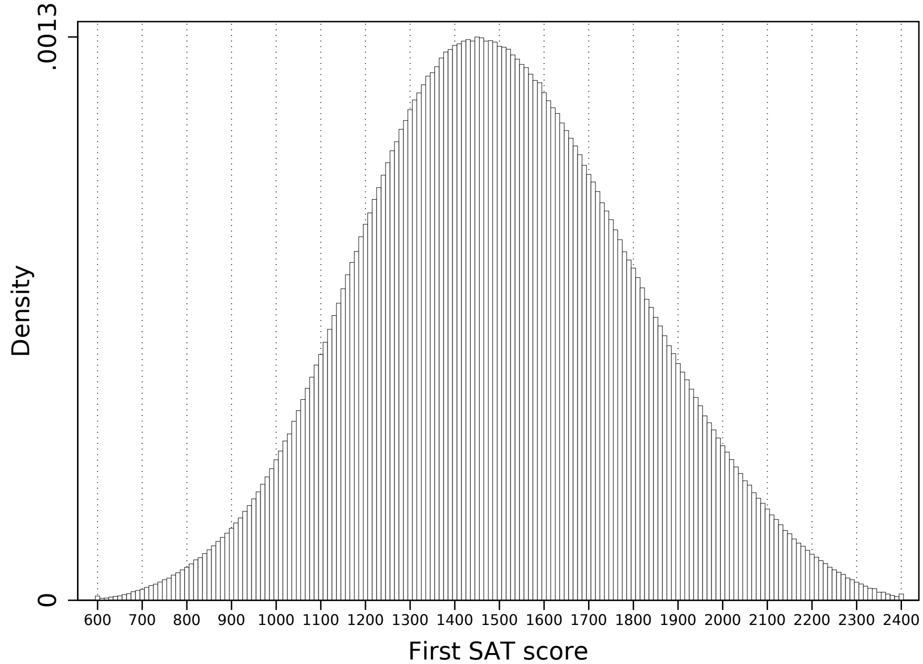

Figure 1. : Density of First SAT Scores

Note: Shown above is the density of students' first SAT scores. The sample consists of all SAT-takers from the high school classes of 2006-14 with valid math, reading and writing scores, and who took their first SAT by November of senior year.

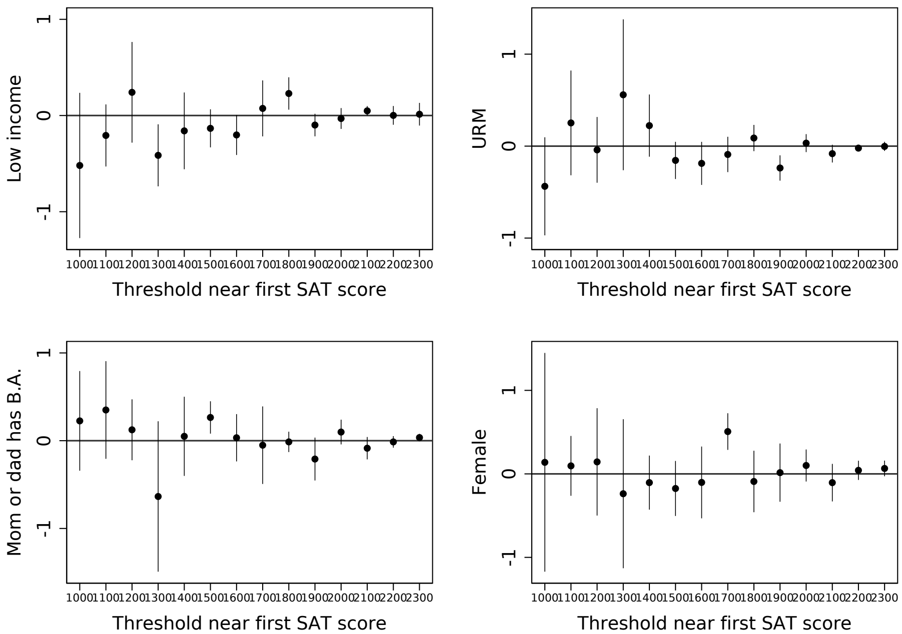

Figure 2. : Covariate Balance

Note: Shown above are estimated discontinuities in various demographic characteristics due to one's first SAT score falling below a given multiple of 100. Vertical lines represent 95 percent confidence intervals. The lowest thresholds (700-900) are omitted for scaling purposes. The sample consists of all SAT-takers from the high school classes of 2006-14 with valid math, reading and writing scores, and who took their first SAT by November of senior year.

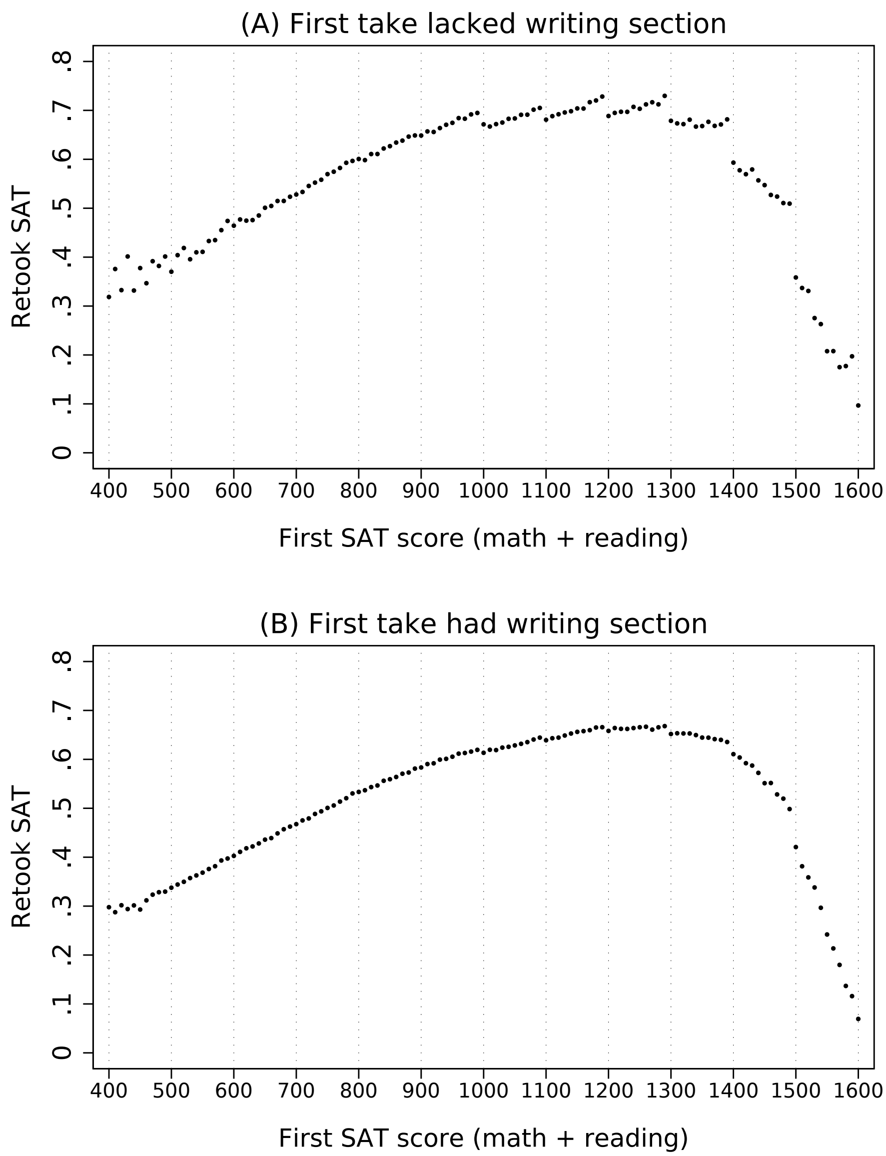

Figure 3. : Using the Prior SAT Scale

Note: Shown above are average SAT retaking rates as a function of students' first math plus reading scores. The sample in panel A consists of all SAT-takers from the high school classes of 2004-6 with valid math and reading scores but not writing scores, and who took their first SAT by November of senior year. The sample in panel B consists of all SAT-takers from the high school classes of 2006-14 with valid math, reading and writing scores, and who took their first SAT by November of senior year.

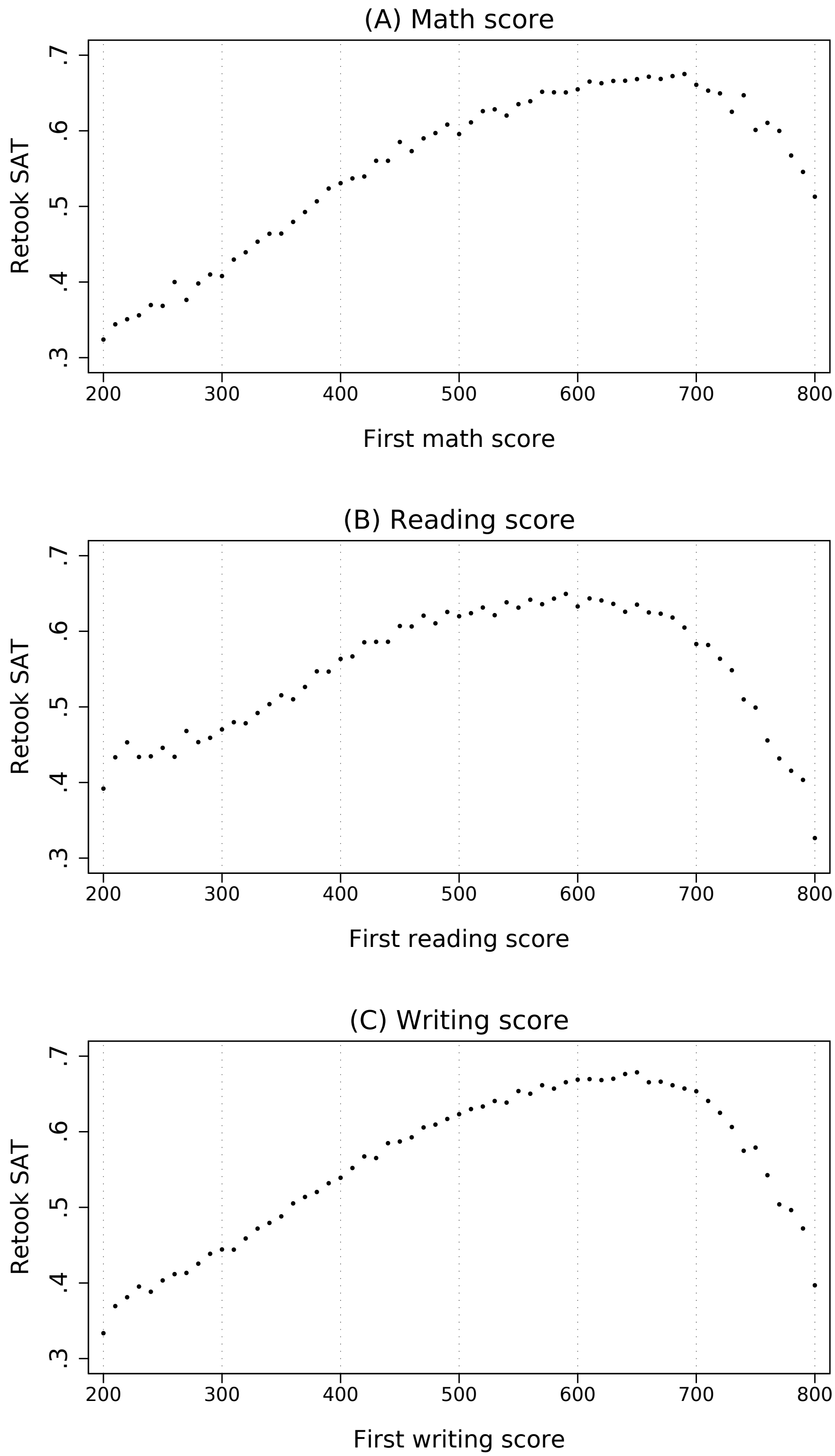

Figure 4. : Individual Subject Scores

Note: Shown above are average SAT retaking rates as a function of students' first scores on each individual subject. The sample in all panels consists of all SAT-takers from the high school classes of 2006-14 with valid math, reading and writing scores, and who took their first SAT by November of senior year.

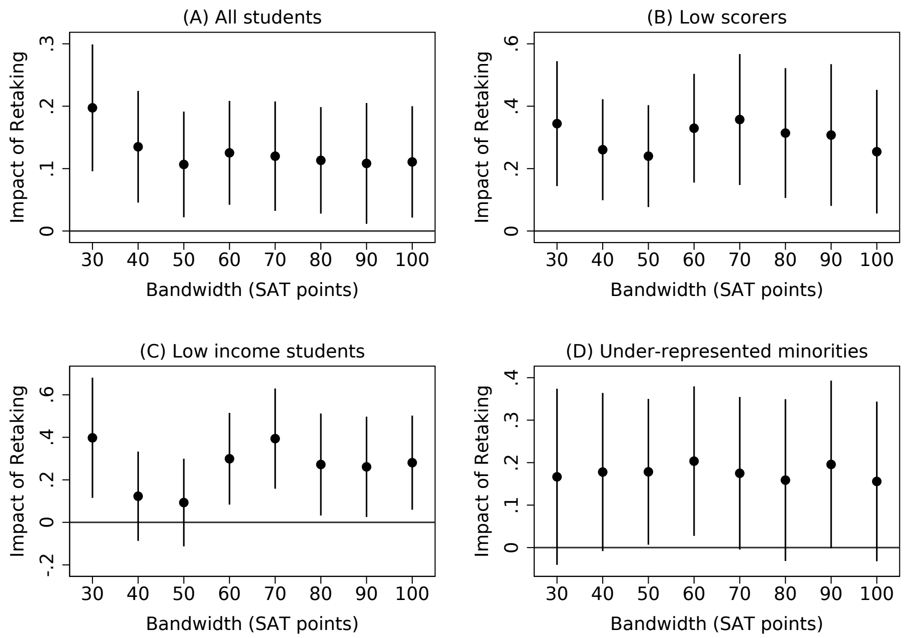

Figure 5. : Bandwidth Robustness - Four-Year College Enrollment

Note: The above figure shows instrumental variables estimates from Equation 2 of the impact of retaking on four-year college enrollment across a variety of bandwidths.

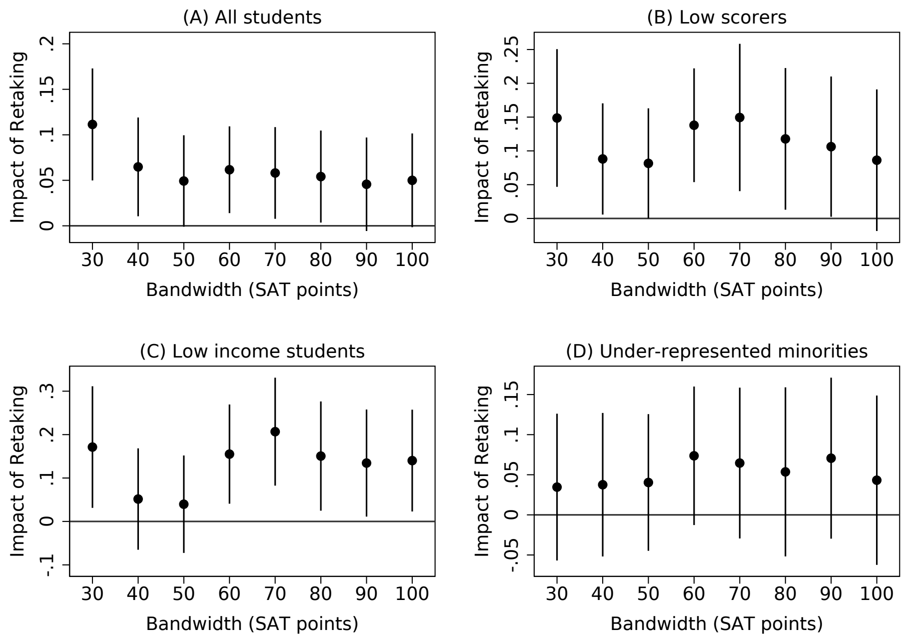

Figure 6. : Bandwidth Robustness - College's Graduation Rate

Note: The above figure shows instrumental variables estimates from Equation 2 of the impact of retaking on the graduation rate of college chosen across a variety of bandwidths.

Table 1—: Covariate Balance

	Low income (1)	URM (2)	Mom or dad has B.A. (3)	Female (4)
Bandwidth = 40	-0.075*	-0.012	0.058	-0.000
	(0.043)	(0.036)	(0.041)	(0.047)
Bandwidth = 60	-0.052	-0.002	0.022	0.023
	(0.036)	(0.034)	(0.038)	(0.045)
Bandwidth = 80	-0.051	0.005	0.030	0.023
	(0.038)	(0.035)	(0.036)	(0.048)
Bandwidth = 100	-0.058	0.015	0.026	0.031
	(0.039)	(0.038)	(0.039)	(0.046)

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year.

Table 2—: Missing Round Number Thresholds and SAT Scores (Reduced Form)

	Mean gain, 1st retake (1)	Superscore gain by 2nd take (2)	Superscore by last take (3)	150+ point increase (4)
(A) All students				
All	0.429***	0.810***	0.941***	0.002***
	(0.064)	(0.055)	(0.063)	(0.000)
(B) By initial score				
Lower scoring	0.528***	0.664***	0.792***	0.003***
	(0.078)	(0.064)	(0.077)	(0.000)
Higher scoring	0.307***	1.004***	1.144***	0.002***
	(0.110)	(0.098)	(0.112)	(0.000)
(C) By income				
Low income	0.606***	0.916***	1.000***	0.003***
	(0.152)	(0.139)	(0.157)	(0.001)
High income	0.409**	0.844***	1.032***	0.002**
	(0.163)	(0.139)	(0.155)	(0.001)
(D) By race/ethnicity				
URM	0.470***	0.741***	0.838***	0.002***
	(0.132)	(0.116)	(0.130)	(0.001)
Non-URM	0.372***	0.831***	0.967***	0.002***
	(0.080)	(0.065)	(0.074)	(0.000)

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is a reduced form estimate of the impact of one's first SAT score missing a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 3—: Retaking and SAT Scores

	Raised score in			Raised score in at least		
	Math	Reading	Writing	1 section	2 sections	3 sections
	(1)	(2)	(3)	(4)	(5)	(6)
(A) All students						
All	0.526***	0.625***	0.573***	0.920***	0.602***	0.203***
	(0.044)	(0.047)	(0.045)	(0.021)	(0.049)	(0.034)
(B) By initial score						
Lower scoring	0.531***	0.815***	0.757***	1.017***	0.758***	0.327***
	(0.091)	(0.090)	(0.070)	(0.039)	(0.091)	(0.066)
Higher scoring	0.525***	0.521***	0.474***	0.867***	0.517***	0.135***
	(0.047)	(0.051)	(0.052)	(0.022)	(0.057)	(0.039)
(C) By income						
Low income	0.567***	0.557***	0.538***	0.927***	0.529***	0.207***
	(0.084)	(0.112)	(0.094)	(0.052)	(0.112)	(0.073)
High income	0.568***	0.611***	0.622***	1.013***	0.602***	0.185**
	(0.082)	(0.093)	(0.103)	(0.046)	(0.091)	(0.077)
(D) By race/ethnicity						
URM	0.510***	0.670***	0.561***	0.962***	0.577***	0.203***
	(0.087)	(0.086)	(0.084)	(0.045)	(0.102)	(0.068)
Non-URM	0.519***	0.607***	0.559***	0.899***	0.593***	0.194***
	(0.053)	(0.052)	(0.053)	(0.025)	(0.054)	(0.039)

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of retaking the SAT once, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 4—: Robustness Checks: SAT Superscore

	(1)	(2)	(3)	(4)	(5)	(6)
(A) All students						
All	102.9***	101.9***	98.9***	96.7***	101.7***	101.9***
	(6.7)	(5.9)	(6.2)	(6.5)	(5.9)	(6.8)
(B) By initial score						
Low scoring	130.0***	136.1***	132.7***	120.3***	137.4***	136.1***
	(10.7)	(11.8)	(16.3)	(15.2)	(11.7)	(15.0)
High scoring	85.2***	83.6***	83.1***	84.4***	82.2***	83.6***
	(7.0)	(5.4)	(4.6)	(5.4)	(5.4)	(6.9)
(C) By income						
Low income	102.6***	118.9***	124.4***	111.7***	117.2***	118.9***
	(17.1)	(15.5)	(15.5)	(15.7)	(15.9)	(16.4)
High income	87.6***	99.5***	93.9***	97.2***	99.8***	99.5***
	(16.1)	(14.1)	(13.9)	(14.0)	(14.4)	(14.7)
(D) By race/ethnicity						
URM	101.3***	106.1***	103.8***	97.6***	106.0***	106.1***
	(13.4)	(13.0)	(12.9)	(13.4)	(12.7)	(14.1)
Non-URM	98.6***	97.5***	95.2***	94.8***	96.9***	97.5***
	(8.0)	(6.8)	(6.9)	(7.0)	(7.0)	(7.7)
Bandwidth	40	60	80	100	60	60
Covariates	N	N	N	N	Y	N
Clustered standard errors	Y	Y	Y	Y	Y	N

Note: Heteroskedasticity robust standard errors (clustered by first SAT score in all but the final column) are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking on SAT superscore, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 5—: Missing Round Number Thresholds and College Enrollment (Reduced Form)

	College type		College's graduation rate			College's mean earnings		
	Four-year	Two-year	Overall	>50%	>80%	Overall	> $50,000	> $65,000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) All students								
All	0.0012***	-0.0006	0.0006***	0.0008**	0.0000	0.0067	0.0007*	-0.0003
	(0.0004)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0194)	(0.0004)	(0.0003)
(B) By initial score								
Lower scoring	0.0019***	-0.0012**	0.0008***	0.0011***	0.0000	0.0400***	0.0009*	0.0003
	(0.0005)	(0.0005)	(0.0002)	(0.0004)	(0.0003)	(0.0144)	(0.0004)	(0.0003)
Higher scoring	0.0002	0.0003	0.0003	0.0005	0.0000	-0.0343	0.0006	-0.0011*
	(0.0005)	(0.0004)	(0.0004)	(0.0006)	(0.0005)	(0.0400)	(0.0007)	(0.0006)
(C) By income								
Low income	0.0025***	-0.0020***	0.0013**	0.0026***	0.0006	0.0273	0.0023***	0.0002
	(0.0009)	(0.0007)	(0.0005)	(0.0008)	(0.0006)	(0.0363)	(0.0009)	(0.0007)
High income	0.0003	-0.0010	0.0000	0.0005	0.0002	-0.0189	0.0014	-0.0007
	(0.0009)	(0.0007)	(0.0005)	(0.0008)	(0.0008)	(0.0445)	(0.0009)	(0.0009)
(D) By race/ethnicity								
URM	0.0016**	-0.0015**	0.0006	0.0008	0.0001	-0.0036	0.0003	-0.0003
	(0.0007)	(0.0006)	(0.0004)	(0.0006)	(0.0004)	(0.0251)	(0.0006)	(0.0004)
Non-URM	0.0008*	-0.0003	0.0005	0.0007	-0.0001	0.0054	0.0008	-0.0003
	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0235)	(0.0005)	(0.0004)

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is a reduced form estimate of the impact of one's first SAT score missing a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 6—: Retaking and College Enrollment among High Scorers

	College type		College's graduation rate			College's mean earnings		
	Four-year	Two-year	Overall	>50%	>80%	Overall	> $50,000	> $65,000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) By income								
Low income	0.123	-0.026	0.107	0.130	0.138	-3.675	0.213	-0.032
	(0.120)	(0.098)	(0.077)	(0.130)	(0.093)	(7.273)	(0.131)	(0.133)
Control complier mean	0.81	0.03	0.60	0.71	0.32	63.03	0.59	0.44
High income	0.026	-0.041	0.027	0.056	0.032	-3.616	0.116	-0.145
	(0.058)	(0.043)	(0.047)	(0.056)	(0.093)	(5.044)	(0.079)	(0.100)
Control complier mean	0.83	0.04	0.66	0.78	0.45	66.04	0.65	0.43
(B) By race/ethnicity								
URM	0.026	-0.034	0.052	0.080	0.063	-5.219	0.043	-0.044
	(0.079)	(0.078)	(0.043)	(0.080)	(0.094)	(5.161)	(0.087)	(0.102)
Control complier mean	0.88	0.10	0.66	0.76	0.36	69.30	0.76	0.37
Non-URM	0.006	0.029	0.008	0.017	-0.023	-2.573	0.039	-0.082*
	(0.043)	(0.027)	(0.031)	(0.046)	(0.040)	(2.724)	(0.055)	(0.046)
Control complier mean	0.81	0.03	0.64	0.76	0.40	64.02	0.66	0.43

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who took their first SAT by November of senior year and whose first score was nearest to a threshold of at least 1600. Panel A splits the sample into students with family incomes below $50,000 and above $100,000. Panel B splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 7—: Robustness Checks: Four-Year College Enrollment

	(1)	(2)	(3)	(4)	(5)	(6)
(A) All students						
All	0.135***	0.125***	0.113***	0.111**	0.115***	0.125**
	(0.045)	(0.042)	(0.043)	(0.045)	(0.042)	(0.051)
(B) By initial score						
Low scoring	0.261***	0.329***	0.314***	0.254**	0.314***	0.329***
	(0.081)	(0.088)	(0.105)	(0.1)	(0.089)	(0.113)
High scoring	0.049	0.014	0.020	0.036	0.016	0.014
	(0.052)	(0.04)	(0.038)	(0.038)	(0.041)	(0.049)
(C) By income						
Low income	0.123	0.299***	0.272**	0.281**	0.314***	0.299**
	(0.106)	(0.109)	(0.122)	(0.112)	(0.112)	(0.137)
High income	-0.119	0.025	0.164*	0.116	0.014	0.025
	(0.101)	(0.084)	(0.09)	(0.089)	(0.083)	(0.1)
(D) By race/ethnicity						
URM	0.178*	0.204**	0.159	0.156	0.209**	0.204*
	(0.094)	(0.089)	(0.097)	(0.095)	(0.089)	(0.121)
Non-URM	0.102*	0.084*	0.084*	0.086*	0.072	0.084
	(0.052)	(0.046)	(0.047)	(0.045)	(0.045)	(0.057)
Bandwidth	40	60	80	100	60	60
Covariates	N	N	N	N	Y	N
Clustered standard errors	Y	Y	Y	Y	Y	N

Note: Heteroskedasticity robust standard errors (clustered by first SAT score in all but the final column) are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking on four-year college enrollment, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 8—: Robustness Checks: College's Graduation Rate

	(1)	(2)	(3)	(4)	(5)	(6)
(A) All students						
All	0.065**	0.062**	0.054**	0.050*	0.052**	0.062**
	(0.027)	(0.024)	(0.026)	(0.026)	(0.024)	(0.031)
(B) By initial score						
Low scoring	0.088**	0.138***	0.118**	0.086	0.126***	0.138**
	(0.041)	(0.042)	(0.053)	(0.053)	(0.043)	(0.061)
High scoring	0.049	0.021	0.026	0.031	0.018	0.021
	(0.037)	(0.029)	(0.027)	(0.027)	(0.029)	(0.035)
(C) By income						
Low income	0.052	0.155***	0.151**	0.14**	0.157***	0.155**
	(0.059)	(0.058)	(0.064)	(0.059)	(0.059)	(0.078)
High income	-0.072	0.004	0.089*	0.060	-0.005	0.004
	(0.061)	(0.05)	(0.053)	(0.051)	(0.048)	(0.065)
(D) By race/ethnicity						
URM	0.038	0.074*	0.054	0.043	0.076*	0.074
	(0.045)	(0.044)	(0.053)	(0.053)	(0.044)	(0.066)
Non-URM	0.062*	0.046	0.046	0.045	0.036	0.046
	(0.035)	(0.030)	(0.031)	(0.029)	(0.030)	(0.036)
Bandwidth	40	60	80	100	60	60
Covariates	N	N	N	N	Y	N
Clustered standard errors	Y	Y	Y	Y	Y	N

Note: Heteroskedasticity robust standard errors (clustered by first SAT score in all but the final column) are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking on the chosen college's graduation rate, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 9—: Robustness to Exclusion of Individual States

State excluded:	CA (1)	NY (2)	TX (3)	PA (4)	FL (5)	NJ (6)	MA (7)	GA (8)
(A) Four-year college								
All	0.139***	0.110**	0.122***	0.129***	0.142***	0.132***	0.117***	0.135***
	(0.049)	(0.044)	(0.045)	(0.045)	(0.042)	(0.043)	(0.043)	(0.044)
Low scoring	0.266***	0.351***	0.339***	0.366***	0.395***	0.321***	0.299***	0.347***
	(0.097)	(0.100)	(0.094)	(0.098)	(0.082)	(0.093)	(0.087)	(0.092)
Low income	0.221	0.288**	0.281**	0.350***	0.263**	0.305***	0.321***	0.323***
	(0.134)	(0.126)	(0.108)	(0.117)	(0.106)	(0.110)	(0.109)	(0.113)
URM	0.185*	0.262***	0.209**	0.200**	0.246**	0.236**	0.198**	0.228**
	(0.101)	(0.090)	(0.103)	(0.094)	(0.096)	(0.098)	(0.082)	(0.092)
(B) College's grad. rate								
All	0.086***	0.044*	0.066**	0.060**	0.068***	0.064**	0.058**	0.066***
	(0.030)	(0.025)	(0.027)	(0.026)	(0.024)	(0.025)	(0.024)	(0.025)
Low scoring	0.132***	0.127***	0.157***	0.143***	0.172***	0.131***	0.127***	0.147***
	(0.048)	(0.047)	(0.045)	(0.049)	(0.043)	(0.045)	(0.041)	(0.044)
Low income	0.148**	0.134*	0.162**	0.168***	0.132**	0.167***	0.173***	0.166***
	(0.068)	(0.068)	(0.064)	(0.063)	(0.061)	(0.059)	(0.056)	(0.060)
URM	0.110*	0.087**	0.075	0.064	0.084*	0.086*	0.068*	0.080*
	(0.059)	(0.044)	(0.052)	(0.044)	(0.048)	(0.044)	(0.041)	(0.045)
N excluded (all)	2,091,030	1,544,628	1,305,635	983,854	903,840	818,201	609,131	559,704

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Each column excludes a single state from the sample.

Table 10—: Retaking and College Completion

	Maximum SAT score (1)	Four-year college (2)	College's grad. rate (3)	Earned B.A. in 6 years (4)
(A) All students				
All	102.6***	0.160	0.058	0.016
	(12.4)	(0.118)	(0.066)	(0.092)
(B) By initial score				
Lower scoring	139.9***	0.281	0.078	0.078
	(28.4)	(0.282)	(0.129)	(0.209)
Higher scoring	85.8***	0.096	0.043	-0.017
	(12.5)	(0.115)	(0.076)	(0.095)
(C) By income				
Low income	167.1***	0.351*	0.216*	0.267
	(29.8)	(0.193)	(0.120)	(0.208)
High income	70.2***	-0.002	-0.109	-0.107
	(18.4)	(0.189)	(0.120)	(0.156)
(D) By race/ethnicity				
URM	93.3***	-0.040	-0.053	-0.165
	(33.4)	(0.257)	(0.163)	(0.264)
Non-URM	96.6***	0.164	0.066	0.033
	(13.2)	(0.127)	(0.082)	(0.090)

Note: Heteroskedasticity robust standard errors clustered by first SAT score are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient is an instrumental variables estimate of the impact of SAT retaking, where retaking is instrumented with indicators for scoring below a multiple of 100. The sample consists of all SAT-takers from the high school classes of 2006-08 who had valid math, reading and writing scores and who took their first SAT by November of senior year. Panel B splits the sample into students with first scores nearest to the 700-1500 thresholds and those nearest to the 1600-2300 thresholds. Panel C splits the sample into students with family incomes below $50,000 and above $100,000. Panel D splits the sample into students who are underrepresented minorities (Black, Hispanic or Native American) and those who are not (White or Asian).

Table 11—: Income and Race Gaps

	Retook (1)	First SAT score (2)	Final superscore (3)	Four-year college (4)
(A) Income				
Low income	-0.206***	-226***	-250***	-0.249***
	(0.002)	(6)	(7)	(0.004)
Middle income	-0.112***	-103***	-117***	-0.115***
	(0.002)	(2)	(2)	(0.002)
High income mean	0.638	1589	1654	0.704
(B) Race				
URM	-0.113***	-252***	-269***	-0.163***
	(0.003)	(2)	(2)	(0.008)
Non-URM mean	0.550	1550	1603	0.640

Note: Heteroskedasticity robust standard errors clustered by high school are in parentheses (* p<.10 ** p<.05 *** p<.01). Each column regresses the listed outcome on the demographic group indicators. All regressions include cohort fixed effects, as well as indicators for missing income or race, so that high income and non-URM students are the reference groups. Below each column is the mean outcome for the reference group in each panel. The sample consists of all SAT-takers from the high school classes of 2006-14 who had valid math, reading and writing scores. Each regression uses 13,656,612 observations.