College Admissions Testing and Learning about Ability: Evidence from Strategic ACT and SAT Taking*

Hema Shah†
August 28, 2025

Abstract

Using administrative data from North Carolina, I study the decision to take the SAT in the presence of universal ACT testing. I find that low-income students are less likely than their peers to take the SAT in addition to the state-mandated ACT, and that they improve upon their initial ACT scores by less when doing so. Taken together, these disparities decrease low-income students' rankings in the test score distribution when evaluating students on their maximum score rather than their initial ACT score. Income gaps in SAT taking are partially driven by differential responses to the ability signal sent by a student's initial ACT score.

JEL Classification: D83, I23, J24

Keywords: higher education, college admissions testing, college access, ability revelation, learning

*I thank Peter Arcidiacono, Jason Baron, Pat Bayer, Emily Cook, Rob Garlick, Arnaud Maurel, Orgul Ozturk, and Pengpeng Xiao, along with participants of the Duke Labor Lunch, 2023 Association of Policy Analysis and Management annual meeting, and 2023 Southern Economic Association annual meeting, for valuable feedback and suggestions. I thank Kara Bonneau at the North Carolina Education Research Data Center and Tammy Howard, Reenada Perez, and Karl Pond at the North Carolina Department of Public Instruction for data assistance.

†Department of Economics, Duke University. Email: hema.shah@duke.edu

1 Introduction

Students from low-income backgrounds are less likely to apply to, attend, and graduate from college (Bailey and Dynarski, 2011; Belley and Lochner, 2007; Kena et al., 2015). In the past two decades, the gap in college entry rates between high-income and low-income students has narrowed, yet a difference of over ten percentage points remains between students in the highest and lowest income quintiles (McFarland et al., 2018). The divergence between high-income and low-income students is apparent early in the college decision making process: Even the most high-achieving students from low-income families largely fail to apply to *any* selective postsecondary institution (Hoxby and Turner, 2015). As such, there is an extensive body of literature studying policies designed to improve college access (Page and Scott-Clayton, 2016). A central theme of this literature is the complexity of navigating the college admissions process, particularly for low-income students.

One aspect of the college admissions process that has received significant attention in recent debates is the use of admissions tests. Critics argue that admissions tests, namely the SAT and ACT, disadvantage low-income and minority students in the college admissions process and do not reflect students' true potential to perform well in college coursework. Indeed, a long "validity" literature attempts to estimate the predictive power of admissions test scores for collegiate outcomes such as freshman grade point average, finding mixed results (Rothstein, 2004; Westrick et al., 2019).

In this paper, I inform the college admissions testing debate by studying how economically disadvantaged students form their college admissions testing strategies. Using detailed administrative data on North Carolina public school students from 2015-2018, I document selection into taking multiple college admissions tests and the resulting impacts on the distribution of admissions-relevant[1] test scores. During my sample period, the ACT test was mandated for all 11$^{\text{th}}$ grade students in North Carolina public schools. In this setting, I provide the first evidence of strategic multiple test-taking behavior in the presence of universal college admissions testing. Unlike previous research, which studies a selected group of students who choose to take at least one admissions test, I study test-taking behavior among

[1]I define a student's "admissions-relevant" test score as the maximum of his or her ACT superscore and SAT superscore, borrowing this terminology from (Bloem et al., 2021).

the universe of public high school students in North Carolina. This allows me to compare the distributions of admissions test scores with and without selection into test-taking to shed light on the potentially uneven benefits of universal admissions testing policies.

My analysis begins by showing that economically disadvantaged students are 6 percentage points (11%) less likely than their peers to take additional admissions tests beyond the in-school ACT, even within high schools and after conditioning on prior academic performance. Conditional on choosing to take multiple tests, economically disadvantaged students improve upon their initial ACT scores by slightly less than than their peers. I then analyze how selection into multiple test-taking distorts the admissions-relevant test score distribution in favor of non-disadvantaged students. Theory predicts that more signals should give better quality information about ability, but distortions will arise if only relatively privileged students can take multiple admissions tests. I find that taken together, income gaps in multiple test-taking and score changes move low-income students roughly one percentile lower in the admissions-relevant test score distribution than in the initial ACT score distribution.

I next provide a stylized model of multiple test-taking to generate testable predictions of the mechanisms driving disparities in testing strategy. Then, I test whether disparities in testing strategy are partially driven by students' responses to the ability signal sent by their initial ACT score. Using detailed information on students' lagged academic performance, I predict their scores on the initial in-school ACT test to measure over- and under-performance on the exam. I find that non-disadvantaged male students are more likely to take additional tests when they over-perform on the initial ACT test, but the same is not true for their female and economically disadvantaged peers.

Finally, I perform a simple simulation exercise to show that closing the income gap in multiple test-taking and the resulting score improvements would reduce the income gap in admissions-relevant test scores by 5%. The remainder of the gap can be explained by initial gaps in ACT scores, suggesting that policies designed to increase testing access will have limited effects unless coupled with broader test preparation efforts.

This paper contributes to a small body of literature studying the strategic decision to take multiple college admissions tests (Vigdor and Clotfelter, 2003; Frisancho et al., 2016; Krishna et al., 2018; Goodman et al., 2020; Bloem et al., 2021). I provide novel evidence of multiple test-taking behavior in the presence of universal college admissions testing. The presence of universal ACT testing provides all students with an ability signal and creates a strong default option, which may induce different test-taking patterns among various student groups compared to a setting without universal testing.

My results also contribute to the literature on universal college admissions testing policies (Klasik, 2013; Hurwitz et al., 2015; Goodman, 2016; Hyman, 2017; Cook and Turner, 2019). I provide a potential explanation for the relatively small and potentially uneven impacts of universal college admissions testing policies on college enrollment: If many successful college applicants take multiple college admissions tests, then the effect of gaining access to a single test may be diluted by selection into multiple test-taking. This complementary work has important implications for the cost-effectiveness of universal testing policies and the utility of additional interventions to improve college access.

This paper also adds to the literature on educational decision making and beliefs about ability (Arcidiacono, 2004; Altonji et al., 2016). To date, this literature has mainly focused on the impacts of beliefs on college major choice and dropout decisions, with analysis of heterogeneity primarily limited to gender differences. Prior work studying the impacts of ability signals generated by standardized tests on students' beliefs and decision making in the K-12 context includes Papay et al. (2016), studying student responses to proficiency labels on standardized mathematics examinations, Foote et al. (2015), studying the effect of ACT college readiness labels on college enrollment decisions. Most similarly to my study, Bond et al. (2018) estimate the impact of the ability signal sent by a student's SAT score, defining a student's ability signal as the difference between his or her PSAT score and SAT score and finding evidence that positive ability signals lead students to send their SAT scores to more selective portfolios of colleges. I contribute to this literature by providing the first evidence that differential responses to the ability signal sent by a student's initial ACT score contribute to subsequent admissions testing decisions. This result has broad implications for the impacts of student evaluation on educational decision making and later life outcomes. Ultimately, my analysis will contribute to our understanding of how universal college admissions tests reveal ability and impact subsequent decisions among students from various backgrounds.

2 Background and Data

2.1 College Admissions Testing in North Carolina

The SAT and ACT tests are standardized exams used to measure United States students' academic preparation in the college application process. The exams are administered by the College Board and the ACT, two private organizations. During my sample period, the SAT and ACT were widely used in selective college admissions. Among U.S. colleges and universities that are not open enrollment, over 75% of schools either required, recommended, or considered SAT or ACT scores (Bloem et al., 2021).[2]

Every four-year college and university in the United States accepts the SAT and ACT interchangeably (Goodman, 2016), although the tests differ slightly in content and format. The SAT contains two main sections, Math and Reading, and each section is scored on a scale of 200 to 800 by multiples of 10. Students receive a total score between 400 and 1600 that is the sum of their two main section scores.[3] The ACT contains four main sections, English, Math, Reading, and Science, and each section is scored on a scale of 1 to 36. Students receive a composite score between 1 and 36 that is the rounded average of their four main section scores. To compare scores across the two tests, colleges use an SAT-ACT concordance table published by the College Board and the ACT.[4]

Beginning in the 2012-2013 academic year, North Carolina mandated the ACT as part of the state's school accountability program. The test is now administered in high schools during the school day, free of charge, to all 11$^{\text{th}}$ grade students. The test is given during the spring semester of 11$^{\text{th}}$ grade, on or near March 1$^{\text{st}}$. For 87% of students in my sample period, which is entirely post-policy change, the first college admissions test they take is the in-school ACT. This is consistent with test timing patterns in other states and the testing

[2]Due to testing disruptions induced by the Covid-19 pandemic, many colleges have relaxed their admissions testing requirements in the years after my sample period.

[3]Prior to 2016, the SAT contained a third section, Writing, such that total scores ranged from 600 to 2400. In spring 2016, the SAT format was revised to drop Writing and rename Reading "Evidence Based Reading and Writing." Since this change occurred midway through my sample period, I refer to the two sections as Math and Reading and consider only Math and Reading scores throughout my analysis.

[4]A current concordance table can be found here: `https://collegereadiness.collegeboard.org/pdf/guide-2018-act-satconcordance.pdf`. The concordance tables convert SAT scores to ACT scores with the same percentile rank, and vice versa, for a group of students who took both tests.

companies' advice that students take their first test in the spring of their junior year of high school (Goodman et al., 2020; Bloem et al., 2021). Hereafter, I use the phrases "in-school ACT" and "initial ACT" interchangeably.

As in all other U.S. states, North Carolina students are allowed to take additional ACT and SAT tests with no limit on the total number of tests taken. In practice, students are limited by the number of exam dates offered per year, which is typically no more than seven for each test. Students may also be constrained by their ability to afford testing fees, as the state of North Carolina pays only for the in-school ACT test and not for additional testing administrations. The cost of additional tests is $0 for low-income, fee waiver-eligible students and between $40-$60 per test for waiver-ineligible students.[5] Other test-taking costs may include time spent registering for the test, transportation costs, and time taken off from work or extracurricular activities.

While the direct and indirect costs of taking multiple admissions tests are nontrivial, the potential benefits are quite large. Students can improve their scores by maturing or by taking additional coursework between the in-school ACT test and later test administrations. Alternatively, students can improve their scores by gaining familiarity with the standardized admissions testing process. Both the "learning over time" and "learning by doing" explanations have been supported by prior research (Vigdor and Clotfelter, 2003; Frisancho et al., 2016). Furthermore, the most common college admissions testing policy eliminates the risk that a student will decrease his or her test score by taking additional tests. Roughly 75% of colleges and universities follow a policy of "superscoring", that is, summing a student's maximum score on each exam section (College Board, 2015). Under superscoring policies, as opposed to policies that consider most recent test scores or average test scores, a student's admissions-relevant test score can only increase through additional testing attempts. Take, for example, a student who scores a 400 on math and 400 on reading on her first SAT attempt, then on a second attempt scores an 800 on math and 300 on reading. Her SAT superscore would be $800 + 400 = 1200$. Since the most common admissions policy is to superscore within tests but not across the SAT and ACT, I consider a student's

[5] SAT fee waivers cover up to two SAT tests, while ACT fee waivers cover up to four ACT tests. Demonstrating free or reduced price lunch (FRL) eligibility is the primary way to obtain SAT and ACT testing fee waivers, although many FRL-eligible students do not take up fee waivers (Goodman et al., 2020).

admissions-relevant test score to be the maximum of his or her ACT superscore and SAT superscore. This admissions-relevant test score coincides with the maximum composite score from an individual test administration for over 90% of my sample. For brevity, I use the terms "admissions-relevant test score" and "maximum test score" interchangeably.

Admissions test scores are particularly important for North Carolina students planning to attend a public, in-state, 4-year college or university. All public 4-year colleges and universities in North Carolina are part of the University of North Carolina (UNC) system During my sample period, eligibility for admission to any school within the UNC system required a minimum GPA of 2.5 and a minimum ACT score of 17.[6][7] Many public and private universities in North Carolina have admissions standards that are significantly higher than the UNC system-wide threshold; therefore, students stand to benefit from test score increases even above the threshold. For example, the 25$^{\text{th}}$ percentile ACT score of entering students at the state flagship university, UNC Chapel Hill, was 28 in 2019. At UNC Charlotte and UNC Wilmington, the 25$^{\text{th}}$ percentile ACT score was 22 in the same year.

[6]Each college and university in the UNC system is permitted to grant admissions requirement exceptions to up to 1% of admitted students. During my sample period, a pilot program granted additional exceptions to four historically black colleges and universities (HBCUs): Elizabeth City State University, Fayetteville State University, and North Carolina Central University. The pilot program allowed for a sliding scale that offset 10-point SAT score reductions and 1-point ACT score reductions with 0.1-point GPA increases, up to a minimum SAT score of 750 and minimum ACT score of 15. Each of the four universities was permitted to admit up to 100 students per class through the sliding scale.

[7]Students can meet the admissions test score requirement by taking either the SAT or the ACT. The minimum SAT score for admission was 800 until the implementation of the revised SAT test in spring 2016, when the minimum was increased to 880 due to inflation in the revised test scores.

2.2 Data

I use student-level administrative data on the universe of 11^{th} grade students in North Carolina public schools from 2015-2018, provided by the North Carolina Education Research Data Center (NCERDC). The NCERDC data includes each student's initial ACT score from the in-school ACT administration, highest SAT score from each academic year, state standardized test scores, complete course-taking and grade history, demographics, and school characteristics. Notably, the data does not include additional ACT scores from weekend test administrations[8] or administrative data on students' college outcomes. Due to these data restrictions, my study will specifically analyze selection into taking the SAT in the presence of universal ACT testing. Hereafter, I use the phrase "multiple test-taking" to refer to taking the SAT in addition to the in-school ACT test.

I omit from my sample the small percentage of students who are exempted from taking the in-school ACT test.[9] This leaves me with a total sample size of 402,840 students. I convert all ACT scores to the 1600-point SAT scale using the official SAT-ACT concordance tables discussed above.[10] Table 1 presents sample characteristics by economically disadvantaged status (EDS).[11] 43% of the sample is economically disadvantaged (EDS). While 56% of non-EDS students take multiple admissions tests, only 31% of EDS students do so. On average, EDS students have an ACT scores that are 155 points lower and admissions-relevant test scores that are 170 points lower than those of non-EDS students. This suggests that disparities in multiple test-taking widen the income gap in test scores.

[8] Roughly 20% of North Carolina high school students repeated the ACT test during my sample period.

[9] According to NC DPI, exemptions to the mandatory ACT policy include students who have a significant cognitive disability and receive instruction in the Extended Content Standards; students who have a current Individualized Education Program (IEP) documenting participation in the Grade 11 College and Career Readiness Alternate Assessment (CCRAA) as well as a written parental request for participation in the CCRAA; students deemed medically fragile because of a significant medical emergency and/or condition and are unable to participate in testing; students who have been retained in the 11^{th} grade and previously took the ACT; and students who took the SAT or the ACT before January 1, 2016, with scores that meet the ACT college readiness benchmark standards. In practice, many students with qualifying prior SAT or ACT scores still opt in to taking the in-school ACT test.

[10] The concordance table was revised in spring 2016 when the format of the SAT changed, midway through my sample period. I apply the old concordance table to convert ACT scores from spring 2015 and the new concordance table to later ACT scores.

[11] The North Carolina Department of Public Instruction (NC DPI) defines economically disadvantaged status using "free or reduced-price lunch (FRL) eligibility", "classification as homeless, runaway, migrant, or foster", and/or "enroll[ment] in Head Start or comparable State-funded Head Start or Pre-kindergarten program".

Table 1: Sample Characteristics by EDS

	Full Sample Mean	Non-Disadvantaged Mean	Disadvantaged Mean
Took SAT	.4525	.5640	.3104
Took Multiple Tests	.4539	.5647	.3125
ACT Score	954.2	1019	866.8
ACT Math+Verbal	944.3	1011	854.4
SAT Superscore	1080	1124	968.5
Admissions-Relevant Test Score	979.8	1050	885.4
Max Score - First ACT Score	25.59	30.96	18.63
Weighted GPA	3.279	3.575	2.851
Observations	402839	226972	161661

Figure 1 plots kernel density estimates of the distributions of ACT scores, SAT superscores, and admissions-relevant test scores. The SAT score distribution is shifted significantly to the right of the ACT score distribution, suggesting that there is positive selection into taking the SAT.[12] Figure 2 plots kernel density estimates of the test score distributions, restricting the sample to multiple test-takers. Among this selected sample, the ACT score distribution is much closer to the SAT distribution. Still, it is clear that that most students who take multiple tests perform better on the SAT, suggesting that students are improving upon their initial ACT scores by learning over time and learning by doing. It is unlikely that the difference between the SAT score distribution and the ACT score distribution is driven by differences in testing difficulty, as the SAT-ACT score concordance tables are designed to allow comparability across test scores. Therefore, the concordance tables are set such that each ACT score is converted to an SAT score with the same percentile rank for a group of students who took both tests.

[12]In my sample, there is no selection into taking the ACT since the test is mandatory for all students.

Figure 1: Test Score Distributions, Full Sample

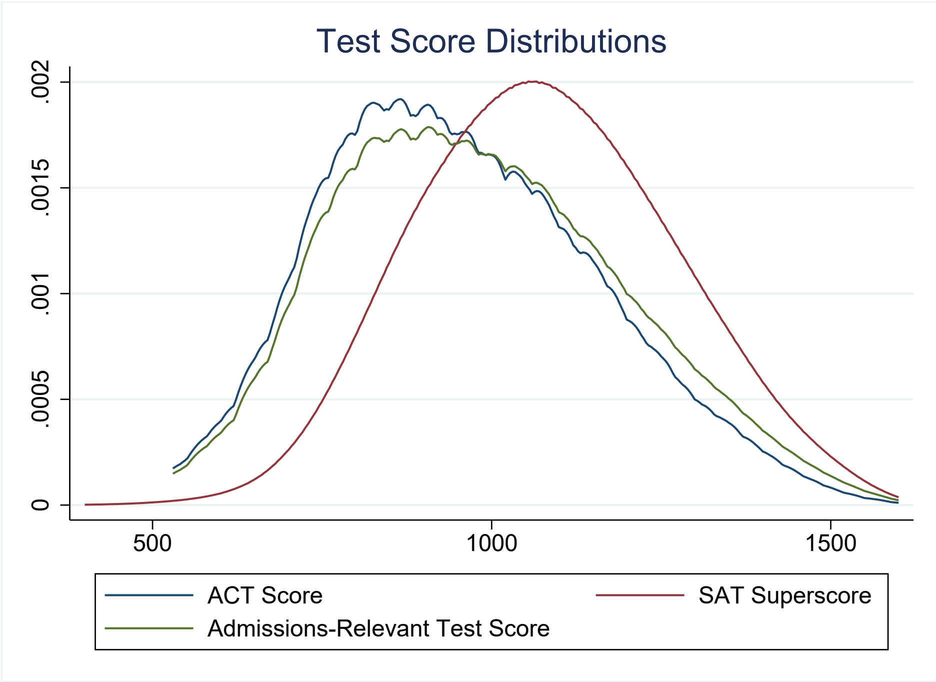
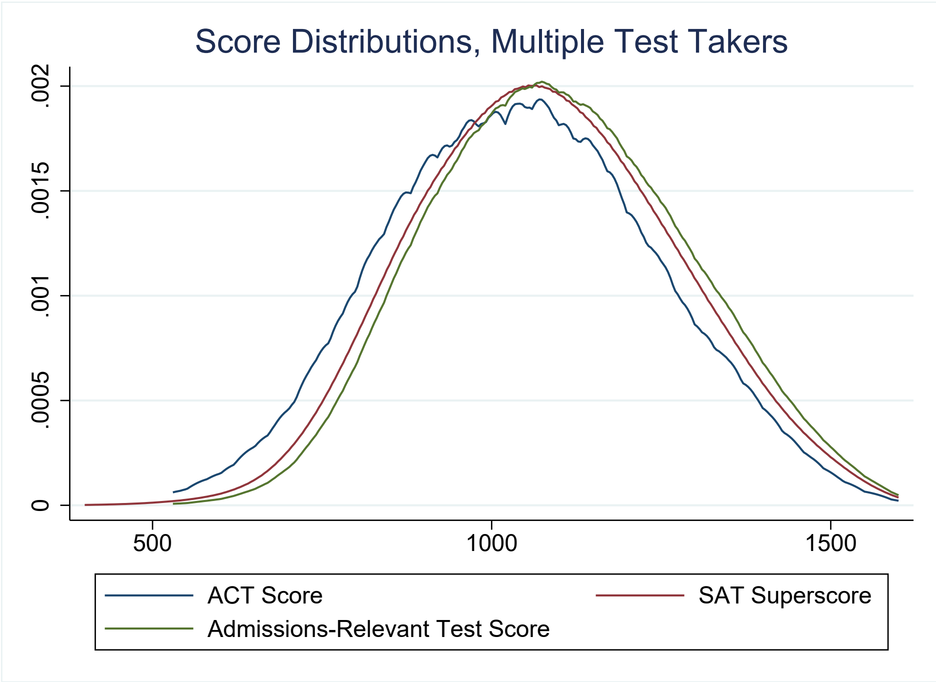
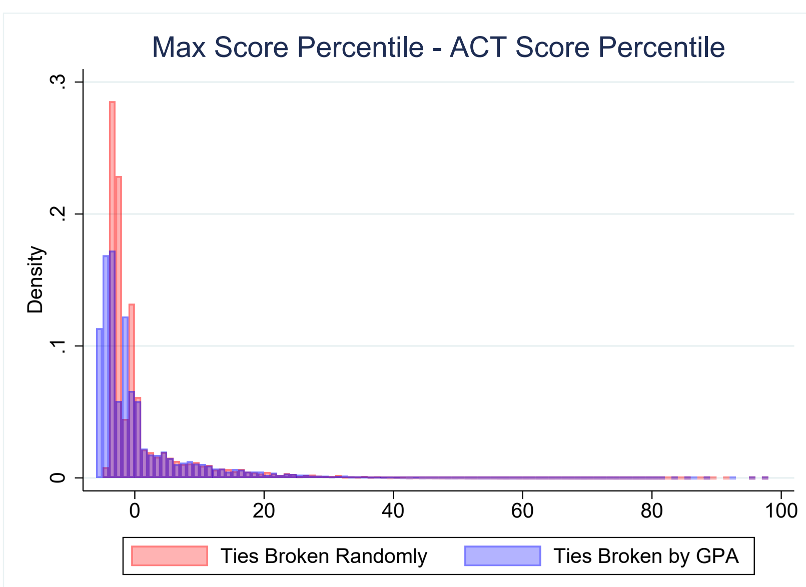


Figure 2: Test Score Distributions, Multiple Test Takers



In order to examine distortions in the test score distribution induced by multiple test-taking, I examine changes in test score rank between the admissions-relevant test score and initial ACT score distributions. Specifically, I define a student's "rank change" as the difference between his or her percentile rank in the admissions-relevant test score distribution and his or her percentile rank in the initial ACT score distribution. A positive value of this variable indicates that a student moved higher in the test score distribution by taking multiple tests, while a negative value indicates that he or she was surpassed by other students who took multiple tests. Due to the coarse distribution of scores on the 36-point ACT scale, there are inevitably many students tied at the same initial ACT scores. Thus, to create evenly sized ACT score percentile bins, I rank students randomly within each ACT score. In practice, I add a small random noise term to each student's initial ACT score. Alternatively, I could rank students by GPA within each ACT score. Figure 3 demonstrates that the two percentile tie-breaking methods generate very similar distributions of changes in rank. Most students move down in rank when taking multiple tests into account, although negative rank changes are all smaller than 10 percentiles. The long right tail of the rank change distribution reveals that a small number of students greatly increase their rank in the test score distribution by taking multiple tests. The 99^{th} percentile of the rank change distribution translates to an increase in rank of 33 percentiles.

Figure 3: Distribution of Changes in Test Score Rank



3 Multiple Test-Taking and Rank Distortions

I estimate three descriptive regressions to understand how students' positions in the test score distribution change when evaluating students using their initial (in-school) ACT scores versus their admissions-relevant test scores. Each regression specification considers student i in school s and cohort t. First, I estimate a probit regression of multiple test-taking on a vector of student characteristics, along with school and cohort fixed effects:

$$\mathbb{1}(\text{took multiple tests})_i = \Phi(\beta X_i + \alpha_s + \gamma_t) \tag{1}$$

where X_i includes initial ACT score, weighted GPA, gender, race, economically disadvantaged status, and race by disadvantage interactions, α_s are school fixed effects, and γ_t are cohort fixed effects.

Next, using only the subsample of multiple test-takers, I estimate an OLS regression of score improvement, the difference between a student's admissions-relevant test score and initial ACT score, on the same covariates.

$$\text{admissions-relevant score}_i - \text{ACT score}_i = \beta X_i + \alpha_s + \gamma_t + \epsilon_{ist} \tag{2}$$

Taken together, the results of these two regressions inform my analysis of changes in test score rank resulting from multiple test-taking behavior. To evaluate how gaps in multiple test-taking and gaps in score changes combine to distort students' rankings in the admissions-relevant test score distribution, I estimate an OLS regression of change in test score rank on the same vector of student characteristics used in the previous regressions.

$$\text{admissions-relevant score rank}_i - \text{ACT score rank}_i = \beta X_i + \alpha_s + \gamma_t + \epsilon_{ist} \tag{3}$$

Coefficient estimates are presented in Table 2. Standard errors are clustered at the school-cohort level in all regressions. Unsurprisingly, students with higher grade point averages are more likely to take multiple tests. Students with higher grades also improve upon their initial ACT scores by more when taking additional tests. Therefore, these students increase in percentile rank, on average, when ranking students on their admissions-relevant test scores instead of their initial ACT scores.

Conditional on prior academic performance, economically disadvantaged students are significantly less likely to take multiple tests than non-disadvantaged students. Translating coefficient estimates to average marginal effects suggests that disadvantaged students are approximately 6 percentage points less likely to take multiple tests, roughly 11% relative to the non-disadvantaged multiple test-taking rate. When disadvantaged students do take multiple tests, they improve upon their ACT scores by less than non-disadvantaged students, although the average difference of less than 3 SAT points on a 1600-point scale is relatively small. Taken together, these two effects move disadvantaged students roughly 1 percentile lower in the test score distribution, on average, when ranking students on their admissions-relevant test scores instead of their initial ACT scores.

In addition to disparities by disadvantaged status, there are also differences in multiple test-taking behavior by race and gender. Conditional on disadvantaged status and prior academic performance, Black students are significantly more likely than white students to take multiple tests, while Hispanic students are less likely to do so. Both Black and Hispanic students improve upon their initial ACT scores by less than white students when taking multiple tests, although the differences in score changes are relatively small. Ultimately, multiple test-taking increases test score rank by 1.5 percentiles for Black students and decreases test score rank by 0.8 percentiles for Hispanic students, on average. Asian students are more likely than all other racial groups to take multiple tests, and on average they improve upon their initial scores by approximately 19 SAT points (or roughly half of an ACT point) more than their peers. Therefore, multiple test-taking increases test score rank by 1.1 percentiles for Asian students. These patterns do not hold for economically disadvantaged Asian students, who make up less than one third of Asian students and less than one percent of the full sample.

Female students are more likely than male students to take multiple tests, although the gender gap in multiple test-taking is much smaller than disparities by disadvantaged status and race. The gender gap in score improvements, on the other hand, is larger than disparities by disadvantaged status and race. Female students improve upon their initial ACT scores by approximately 23 points (or just over half of an ACT point) less than male students, on average, when taking multiple tests. This disparity decreases test score rank for female students by 1.3 percentiles, on average, when ranking students on their admissions-relevant test score instead of their initial ACT scores. This gender gap in rank change is conditional on prior academic performance, which is higher for female students on average.

Appendix A.1 shows that the results are robust to the inclusion of polynomials in ACT score and GPA and to using a linear probability model instead of the probit model in specification (1).

Table 2: Multiple Test-Taking and Student Characteristics

	(1) Took Multiple	(2) Score Change	(3) Rank Change
ACT Score	-0.000118***	-0.273***	-0.0223***
	(-4.10)	(-108.29)	(-72.03)
Weighted GPA	1.008***	37.85***	4.426***
	(144.19)	(77.05)	(84.29)
Female	0.0313***	-23.55***	-1.308***
	(5.35)	(-60.59)	(-38.91)
Economically disadvantaged (EDS)	-0.301***	-2.860***	-0.671***
	(-31.57)	(-5.07)	(-17.57)
Black	0.481***	-3.307***	1.524***
	(35.45)	(-5.05)	(20.84)
Black × EDS	0.196***	1.007	-0.146
	(13.21)	(1.09)	(-1.74)
Hispanic	-0.168***	-5.755***	-0.850***
	(-10.50)	(-6.16)	(-10.96)
Hispanic × EDS	0.167***	-4.221***	-0.0319
	(8.39)	(-3.48)	(-0.36)
Asian	0.137***	19.15***	1.124***
	(5.20)	(12.24)	(9.23)
Asian × EDS	0.114**	-22.42***	-1.280***
	(2.76)	(-9.50)	(-6.19)
Cohort fixed effects	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes
Probit	Yes	No	No
OLS	No	Yes	Yes
N	325046	155964	325567
R^2/Pseudo-R^2	0.3305	0.255	0.185

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The previous results demonstrate that, on average, disadvantaged students are less likely than their peers to take multiple tests, and that they improve upon their initial scores by less when they do take multiple tests. Taken together, these two effects move the *average* disadvantaged student down in percentile rank when considering multiple test scores. Changes in percentile rank, however, are only relevant on the margins of college admission and college selectivity. Average changes in percentile rank will therefore not be economically significant if the average disadvantaged student has an initial ACT score sufficiently below the margin of college admissions or selectivity. However, the average ACT score percentile of disadvantaged students is 39, which corresponds to an ACT score of approximately 16. This is one point below the threshold for admission into the UNC system, suggesting that disparities in rank change are economically significant.

To evaluate the extent to which changes in rank due to multiple test-taking are occurring at margins relevant to college enrollment and selectivity, I estimate an OLS regression of rank change on the interactions between disadvantaged status and a set of dummy variables indicating initial ACT score decile. First, I control for year fixed effects; then, I control for cohort and school fixed effects, race, gender, and GPA. Coefficients are relative to the lowest ACT score decile, in which the average rank change is approximately 2.3 percentiles for disadvantaged students and 3.5 for non-disadvantaged students.

$$\begin{aligned}&\text{admissions-relevant score rank}_i - \text{ACT score rank}_i\\ =&\beta_{\text{ACT}}\text{ACT score decile}_i + \beta_{\text{EDS}}\text{EDS}_i + \beta_{\text{interact}}\text{ACT score decile}_i \times \text{EDS}_i\\ &+\beta_X X_i + \alpha_s + \gamma_t + \epsilon_{ist} \end{aligned} \tag{4}$$

Here, X_i includes weighted GPA, gender, and race, α_s are school fixed effects, and γ_t are cohort fixed effects. Coefficient estimates of of β_{ACT}, β_{EDS}, and β_{interact} are presented in Figures 4 and 5, respectively. Figure 4 demonstrates that rank change is decreasing in initial ACT decile. This is unsurprising, given that students who perform very well on the initial ACT test likely have more difficulty improving upon their initial scores. Figure 5 demonstrates that disparities in rank change by disadvantaged status are highly nonlinear in initial ACT decile. Even after controlling for weighted GPA, gender, race, and school and cohort fixed effects, disadvantaged students in deciles 4 through 6 of the initial ACT score distribution decrease in rank relative to their peers due to multiple test-taking.

Figure 4: Coefficients on Initial ACT Decile

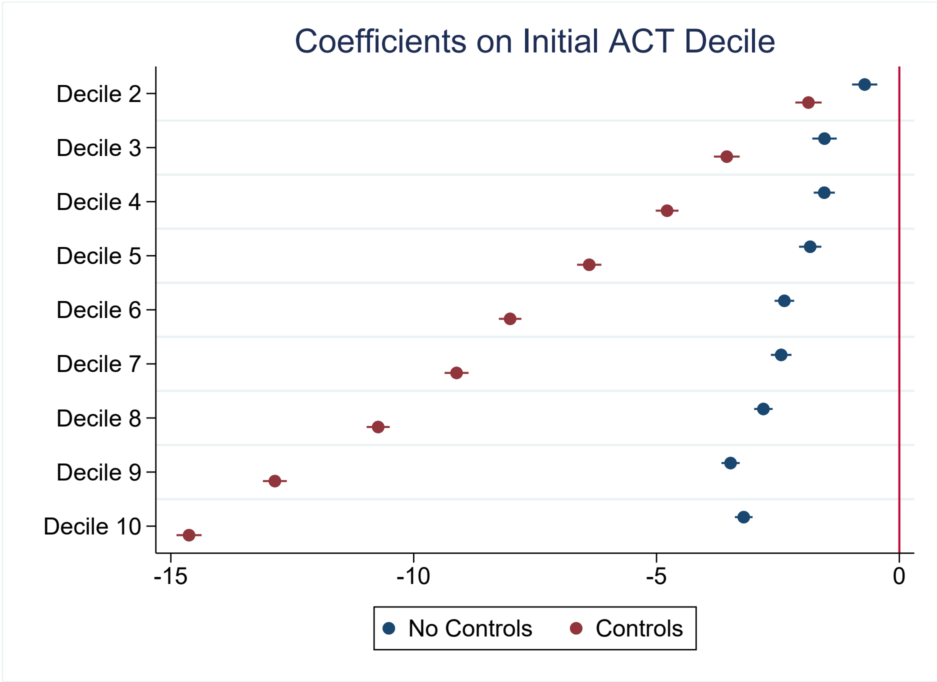
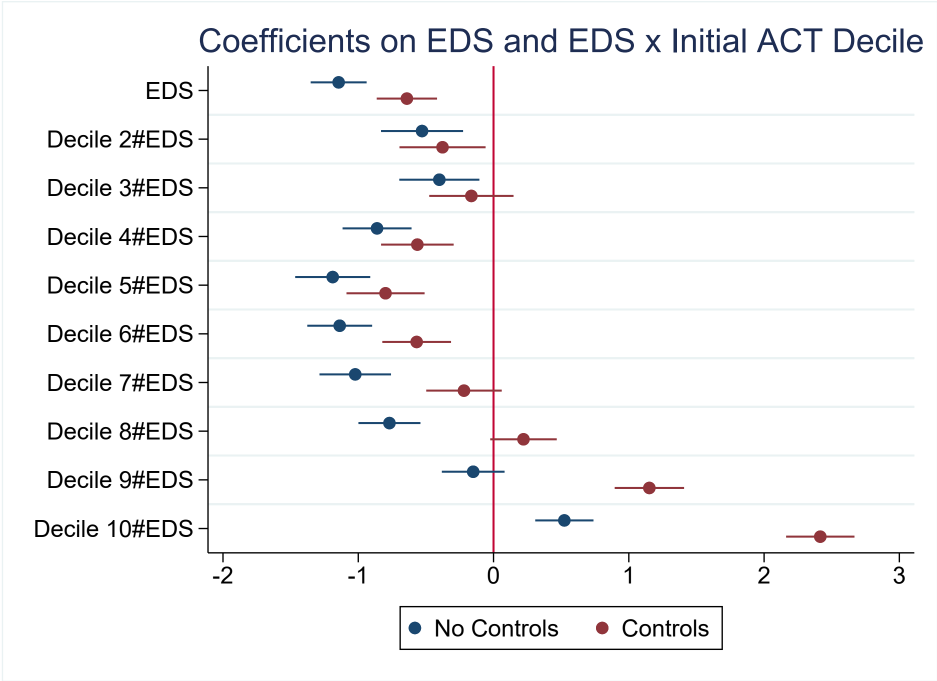


Figure 5: Coefficients on EDS and EDS × Initial ACT Decile



Changes in rank due to multiple test-taking reflect two types of movements: increases in rank due to one's own score improving and decreases in rank due to peers' scores improving while one's own score stays the same. To disentangle these two types of movements, I estimate equation (4) separately among multiple test-takers and single test-takers, with results shown in Figures 6, 7, 8, and 9.

Figure 6 demonstrates that rank change among multiple test-takers is strongly decreasing in initial ACT decile, underscoring the mechanical inability of students with high initial ACT scores to significantly increase their scores on subsequent attempts. Interaction coefficients in Figure 7 reveal that disadvantaged multiple test-takers with high initial ACT scores increase in rank relative to non-disadvantaged multiple-test takers when considering multiple test scores. This suggests that, although disadvantaged students are less likely than their peers to take multiple tests, those who do so benefit from subsequent attempts. This could be driven by positive selection of disadvantaged students into multiple test-taking on unobservables such as familial educational background. This selection explanation suggests that not all disadvantaged students would increase in rank, relative to non-disadvantaged students, from multiple test-taking. Alternatively, relative rank increases could reflect larger gains to disadvantaged students from multiple test attempts, perhaps due to low quality in-school ACT testing conditions in schools with more disadvantaged students. This explanation suggests that disadvantaged students who do not take multiple tests, particularly those with high initial ACT scores, would benefit from doing so. Appendix A.2 demonstrates that average disparities in score changes by disadvantaged status are small within initial ACT scores, providing some support for the selection explanation.

Figure 8 demonstrates that, while all single test-takers mechanically decrease in rank when considering multiple test scores, students at the middle of the initial ACT score distribution see the largest decreases. Interaction coefficients in Figure 9 are small, suggesting that disadvantaged and non-disadvantaged single test-takers decrease in rank similarly when considering multiple test scores. These results support the previous evidence that disparities in rank change by disadvantaged status are driven by disparities in multiple test-taking rather than disparities within the multiple test-taking and single test-taking groups.

Figure 6: Coefficients on Initial ACT Decile, Multiple Test-Takers

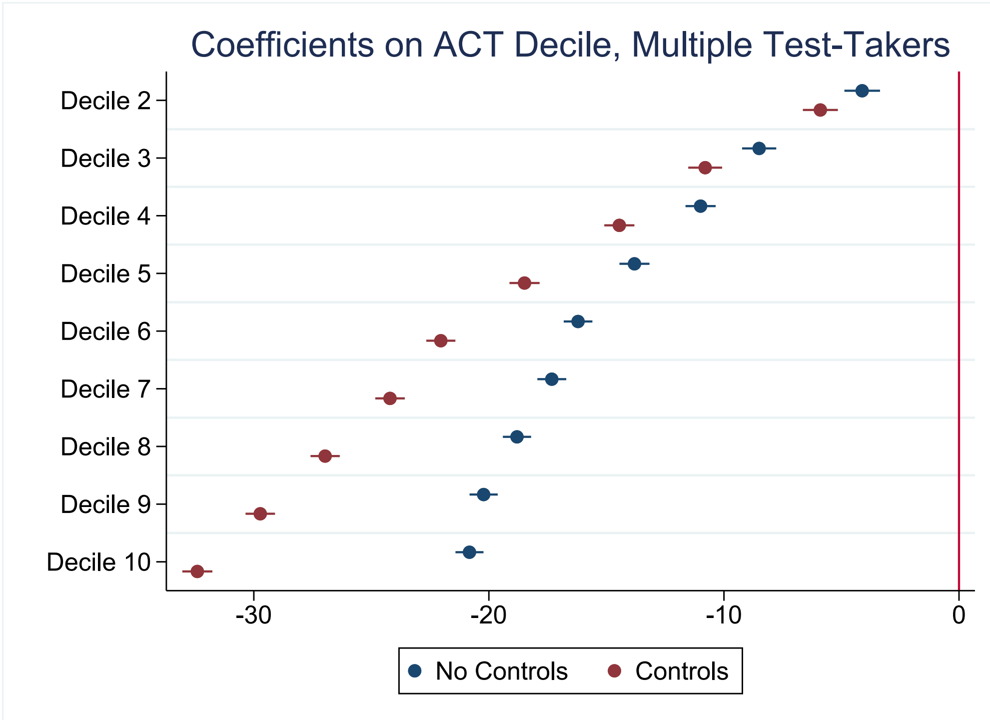


Figure 7: Coefficients on EDS and EDS × Initial ACT Decile, Multiple Test-Takers

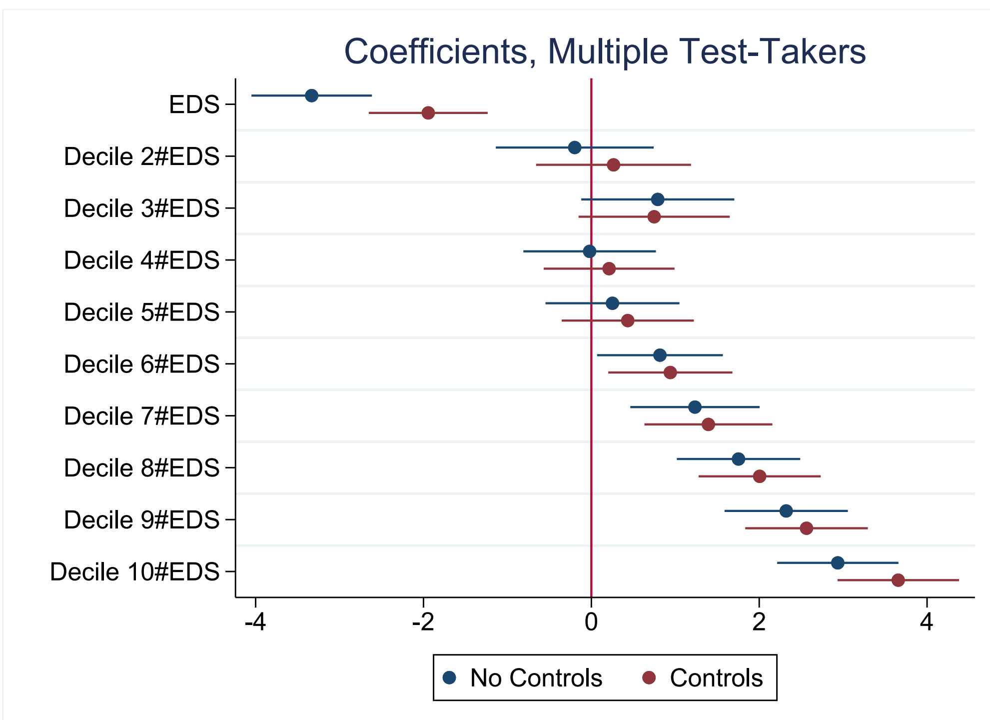


Figure 8: Coefficients on Initial ACT Decile, Single Test-Takers

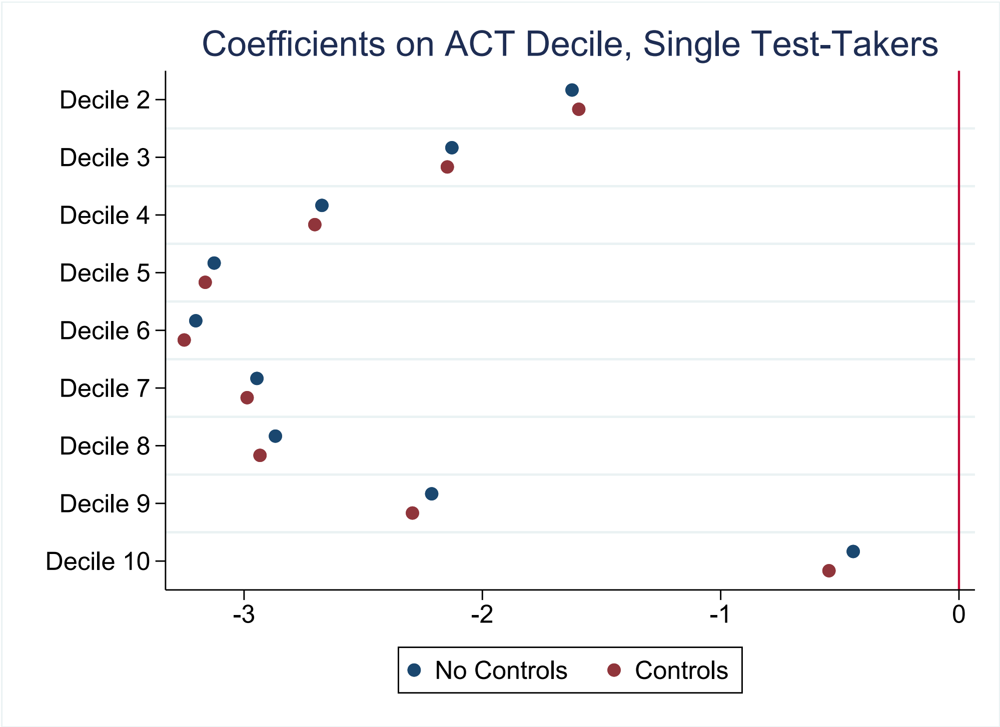
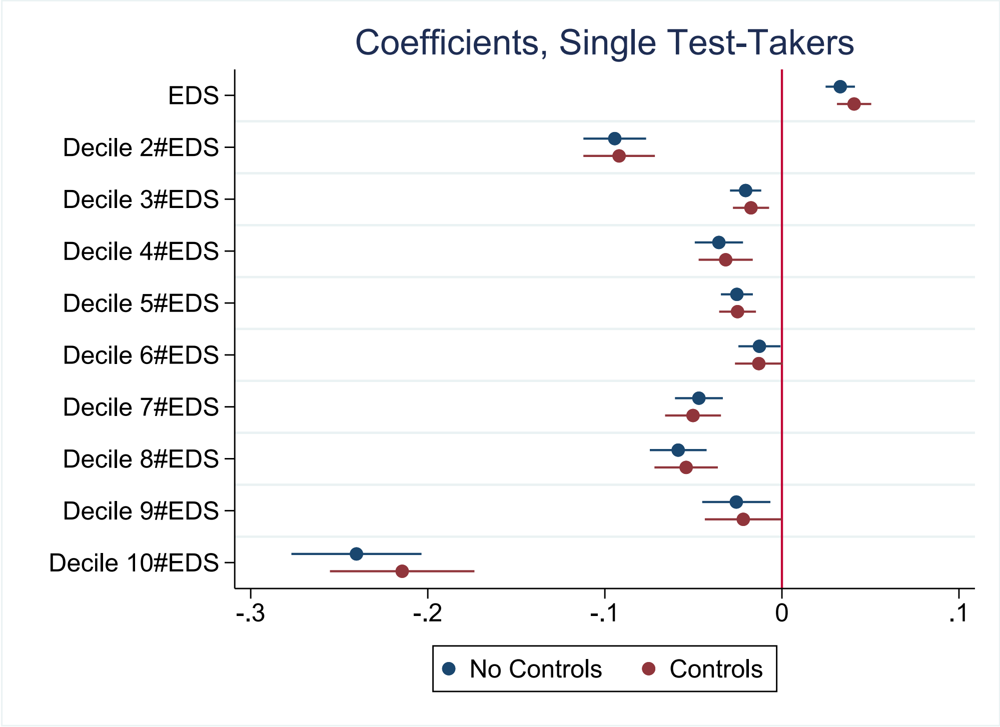


Figure 9: Coefficients on EDS and EDS × Initial ACT Decile, Single Test-Takers



4 A Stylized Model of Multiple Test-Taking

I provide a stylized model of multiple-test taking behavior building on Vigdor and Clotfelter (2003). Suppose student i has true academic ability ρ_i, which I assume to be one-dimensional for simplicity. College admissions test scores are noisy estimates of ρ_i, which is unknown to the student. Each time student i takes a college admissions test, she receives a score p_i drawn from a distribution $f_i(p)$ with mean ρ_i. After taking the mandatory in-school ACT test and receiving score p_i^*, student i decides to take an additional admissions test at time t if and only if the expected benefit exceeds the test-taking cost c_i; that is, if

$$\tilde{\mathbb{E}}_{it}[V_i(a(\max\{p, p_i^*\}) - a(p_i^*)) \mid \mathcal{F}_{it}] > c_i \tag{5}$$

The expected benefit of taking an additional admissions test is equal to student i's valuation of college admission, V_i, times student i's expected change in admissions probability from taking an additional admissions test. The student's expected change in admissions probability depends on two unknown objects: $a(\cdot)$, admissions probability as a function of admissions test score, and p, the score received on the additional admissions test. Student i forms subjective expectation $\tilde{\mathbb{E}}_{it}$ over $a(\cdot)$ and p using time t information $\mathcal{F}_{it}$. $\mathcal{F}_{it}$ contains information about the college admissions process as well as information about student i's academic ability. Information about the college admissions process is unobserved in my data and can come from family, friends, teachers, counselors, and other sources. Information about academic ability consists of observable pre-ACT signals from standardized test scores and course grades as well as an observable signal from the in-school ACT test itself. Information about academic ability may also include signals, such as SAT tutoring participation, that are unobserved in my data. This model suggests four potential mechanisms driving income gaps in multiple test-taking.

First, direct test-taking costs c_i may vary with disadvantaged status for multiple reasons. Economically disadvantaged students who do not register for testing fee waivers may face budget constraints when paying the testing fee. Economically disadvantaged students may also have difficulties securing transportation to testing centers or face higher opportunity costs of test-taking due to part-time jobs or family responsibilities. Second, economically disadvantaged students may differ in their valuation of college V_i.

Third, beliefs about admissions probability function $a(\cdot)$ may vary with disadvantaged status if economically disadvantaged students have access to less information about the college admissions process. This could be the case, for example, if college counseling quality is lower in schools with more disadvantaged students.

The fourth mechanism through which disparities in multiple test-taking may arise is through the belief updating channel. Specifically, economically disadvantaged students may update their subjective expected test scores p differently in response to their initial ACT scores. Since admissions probability $a(\cdot)$ is weakly increasing over the support of test scores p, the expected benefit of taking an additional admissions test is increasing in student i's subjective expected test score at the time of the test-taking decision, $\tilde{\mathbb{E}}_{it}[p \mid \mathcal{F}_{it}]$. I assume students process information in a Bayesian manner, so that $\tilde{\mathbb{E}}_{it}[p \mid \mathcal{F}_{it}]$ is a weighted average of pre-ACT and post-ACT beliefs, as follows:

$$\tilde{\mathbb{E}}_{it}[p \mid \mathcal{F}_{it}] = \alpha_{\text{i,pre-ACT}}\tilde{\mathbb{E}}_i[p \mid \mathcal{F}_{i,\text{pre-ACT}}] + \alpha_{i,\text{ACT}}\tilde{\mathbb{E}}_i[p \mid \mathcal{F}_{i,\text{ACT}}] \tag{6}$$

$$\alpha_{i,\text{pre-ACT}} = \frac{var(p \mid \mathcal{F}_{i,\text{ACT}})}{var(p \mid \mathcal{F}_{i,\text{ACT}}) + var(p \mid \mathcal{F}_{i,\text{pre-ACT}})} \tag{7}$$

$$\alpha_{i,\text{ACT}} = \frac{var(p \mid \mathcal{F}_{\text{i,pre-ACT}})}{var(p \mid \mathcal{F}_{i,\text{ACT}}) + var(p \mid \mathcal{F}_{i,\text{pre-ACT}})} \tag{8}$$

Disparities in belief updating by disadvantaged status could arise for two reasons. First, disadvantaged and non-disadvantaged students may have pre-ACT beliefs $\tilde{\mathbb{E}}_i[p \mid \mathcal{F}_{i,\text{pre-ACT}}]$ that are biased differently. For example, disadvantaged students may be overly confident in their academic abilities if schools with more disadvantaged have less rigorous coursework.

Second, the variance of pre-ACT ability information may vary with disadvantaged status. Since students are Bayesian updaters, $\tilde{\mathbb{E}}_{it}[p \mid \mathcal{F}_{it}]$, and consequently the decision to take an additional college admissions test, will place more weight on the signal sent by the ACT if pre-ACT information is more diffuse. If standardized test scores and course grades are less predictive of college readiness for disadvantaged students, then disparities in belief updating may arise. Although I am unable to disentangle biased beliefs from diffuse priors, I test whether disadvantaged students respond differently to initial ACT performance.

5 Belief Updating

Restricting my analysis to the subsample of students whose first admissions test taken is the in-school ACT, I examine multiple test-taking behavior in response to the ability signal sent by students' initial ACT scores. Specifically, I test whether disadvantaged and non-disadvantaged students respond differently to underperformance or overperformance on the initial ACT test. To measure the ability signal sent by a student's initial ACT score, I construct a predicted ACT score measure based on past academic performance. I predict ACT performance using an OLS regression of ACT score on cubic polynomials in weighted GPA and lagged test scores, with school and cohort fixed effects.

$$\begin{aligned}
\widehat{ACT}_i = {} & \hat{\beta}_{GPA}GPA_i + \hat{\beta}_{GPA2}GPA_i^2 + \hat{\beta}_{GPA3}GPA_i^3 + \hat{\beta}_{MA}MA_i + \hat{\beta}_{MA2}MA_i^2 + \hat{\beta}_{MA3}MA_i^3 \\
& + \hat{\beta}_{RD}RD_i + \hat{\beta}_{RD2}RD_i^2 + \hat{\beta}_{RD3}RD_i^3 + \hat{\beta}_{BIOL}BIOL_i + \hat{\beta}_{BIOL2}BIOL_i^2 + \hat{\beta}_{BIOL3}BIOL_i^3 \\
& + \hat{\beta}_{ALG}ALG_i + \hat{\beta}_{ALG2}ALG_i^2 + \hat{\beta}_{ALG3}ALG_i^3 + \hat{\beta}_{ENG}ENG_i + \hat{\beta}_{ENG2}ENG_i^2 + \hat{\beta}_{ENG3}ENG_i^3 \\
& + \hat{\beta}_X X_i + \hat{\alpha}_s + \hat{\gamma}_t + \hat{\epsilon}_{ist} \qquad (9)
\end{aligned}$$

Here, MA_i refers to 8$^{\text{th}}$ grade math test score, RD_i refers to 8$^{\text{th}}$ grade math test score, $BIOL_i$ refers to 9$^{\text{th}}$ grade biology test score, ALG_i refers to 9$^{\text{th}}$ grade algebra test score,[13] and ENG_i refers to 10$^{\text{th}}$ grade English test score. X_i includes race, gender, and disadvantaged status.[14] Appendix A.4 provides more detail on the fit of the prediction model, which predicts the ACT score distribution well with an R^2 of 0.795 and has similar fit between disadvantaged and non-disadvantaged students. A positive value of the residual indicates that a student overperformed on the ACT test relative to her past academic performance.

To understand the relationship between multiple test-taking and ACT performance relative to expected performance, I estimate a probit regression of multiple test-taking on predicted ACT score, ACT score residual, and the interaction between the residual and dis-

[13]Prior to the 2012-2013 school year, Algebra 1 was the standard mathematics course taken in 9$^{\text{th}}$ grade, or as early as 7$^{\text{th}}$ grade for students on advanced math tracks. In the 2012-2013 school year, North Carolina implemented a revised math curriculum and renamed the 9$^{\text{th}}$ grade math course "Math 1." The first cohort in my sample to take the revised 9$^{\text{th}}$ grade math curriculum is the 2016 cohort. I account for cross-year differences in course curriculum by standardizing test scores within years.

[14]I normalize all standardized test scores to have mean zero and standard deviation one within each cohort.

advantaged status, shown in equation (10). I estimate the regression on the subsample of students with predicted ACT scores in the middle 50% of the predicted score distribution, as students with low academic achievement are unlikely to attend college and therefore unlikely to take multiple admissions tests and students with very high academic achievement may plan to take multiple admissions tests regardless of their initial performance to be competitive for selective college admissions. Results are presented in Table 3. I estimate the regression separately on the female and male samples. Coefficient estimates suggest that the response to the ACT score residual is insignificant for females and positive for males. Among non-disadvantaged males, converting coefficient estimates to average marginal effects suggests that a one standard deviation increase in ACT score residual increases multiple test-taking by roughly 2 percentage points. Among disadvantaged males, the effect size is halved.[15]

$$\mathbb{1}(\text{took multiple tests})_i = \beta_{resid}\hat{\epsilon}_{ist} + \beta_{EDS}EDS_i\beta_{interact}\hat{\epsilon}_{ist} \times EDS_i + \alpha_s + e_{ist}$$

Table 3: Multiple Test-Taking and ACT Score Residual

	(1) Female	(2) Male
Predicted ACT Score	0.150***	0.154***
	(49.63)	(49.36)
ACT Score Residual	0.00422	0.0294***
	(1.25)	(10.34)
Economically disadvantaged	-0.203***	-0.226***
	(-18.58)	(-19.60)
Economically disadvantaged × Residual	-0.00789	-0.0123**
	(-1.54)	(-2.66)
School Fixed Effects	Yes	Yes
N	67526	64655
Pseudo-R^2	0.1269	0.1255

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[15] Appendix A.3 demonstrates that the results are robust to using a linear probability model and to allowing each predictor of initial ACT score to enter separately instead of summarizing them in the predicted ACT score measure. Results are qualitatively similar but smaller in magnitude among the full sample.

To further understand how multiple test-taking behavior varies across the distribution of ACT score residuals, I estimate a probit regression of multiple test-taking on predicted ACT score and the interactions between disadvantaged status and a set of dummy variables indicating initial ACT score decile, shown in equation (10). First, I control for year fixed effects; then, I control for cohort and school fixed effects. Coefficients are relative to the lowest decile of the ACT score residual distribution, in which the multiple test-taking rate is approximately 36% for disadvantaged students and 54% for non-disadvantaged students.

$$\begin{aligned}\mathbb{1}(\text{took multiple tests})_i = \Phi(&\beta_{predict}\hat{ACT}_i + \beta_{resid}\text{residual decile}_i + \beta_{EDS}EDS_i \\ &+ \beta_{interact}\text{residual decile}_i \times EDS_i + \alpha_s + \gamma_t + e_{ist})\end{aligned} \tag{10}$$

Female coefficient estimates of β_{resid}, β_{EDS}, and $\beta_{interact}$ are presented in Figures 10 and 11, respectively. Male coefficient estimates are presented in Figures 12 and 13. Estimates of β_{resid} suggest that female students' multiple test-taking behavior varies nonlinearly with ACT score residual. Female students in the highest and lowest residual deciles have the lowest multiple test-taking rates conditional on predicted ACT score; that is, female students who perform very well or very poorly, respectively, on the initial ACT test relative to expectations are less likely than their peers to take additional tests. Estimates of $\beta_{interact}$ suggest that, among female students, responses to initial ACT performance do not differ significantly by disadvantaged status. Among male students, the multiple test-taking rate conditional on predicted ACT score is increasing in ACT score residual; that is, male students who perform well on the initial ACT test relative to expectations are more likely than their peers to take additional tests, even among those who overperform significantly. Further, the increasing response to ACT residual is stronger among non-disadvantaged male students.

Figure 10: Coefficients on Residual Decile, Females

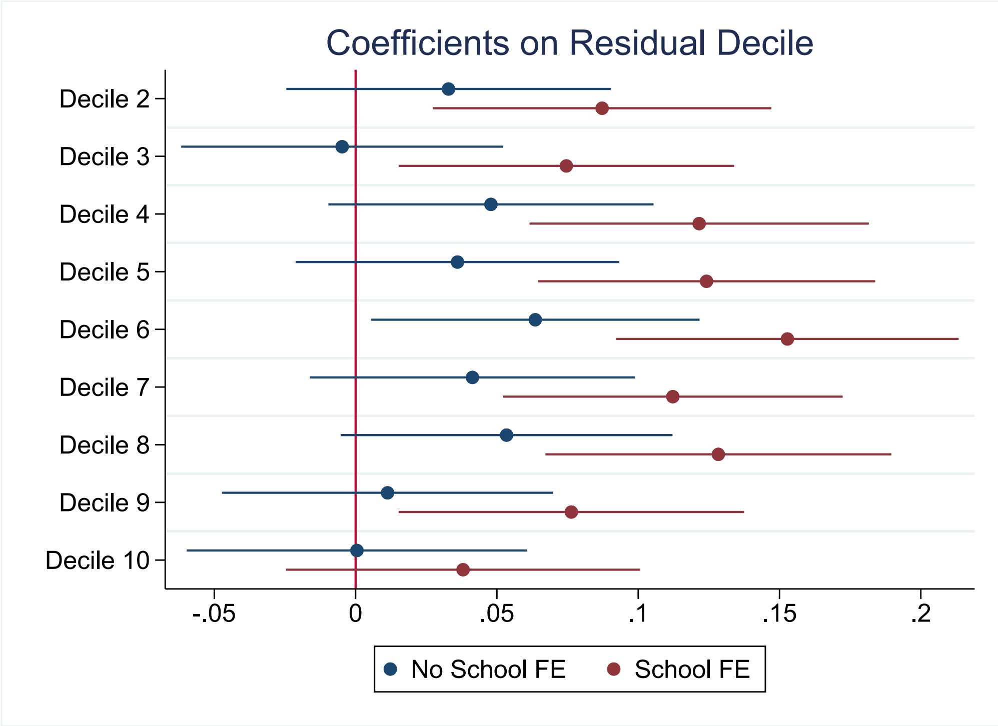


Figure 11: Coefficients on EDS and EDS × Residual Decile, Females

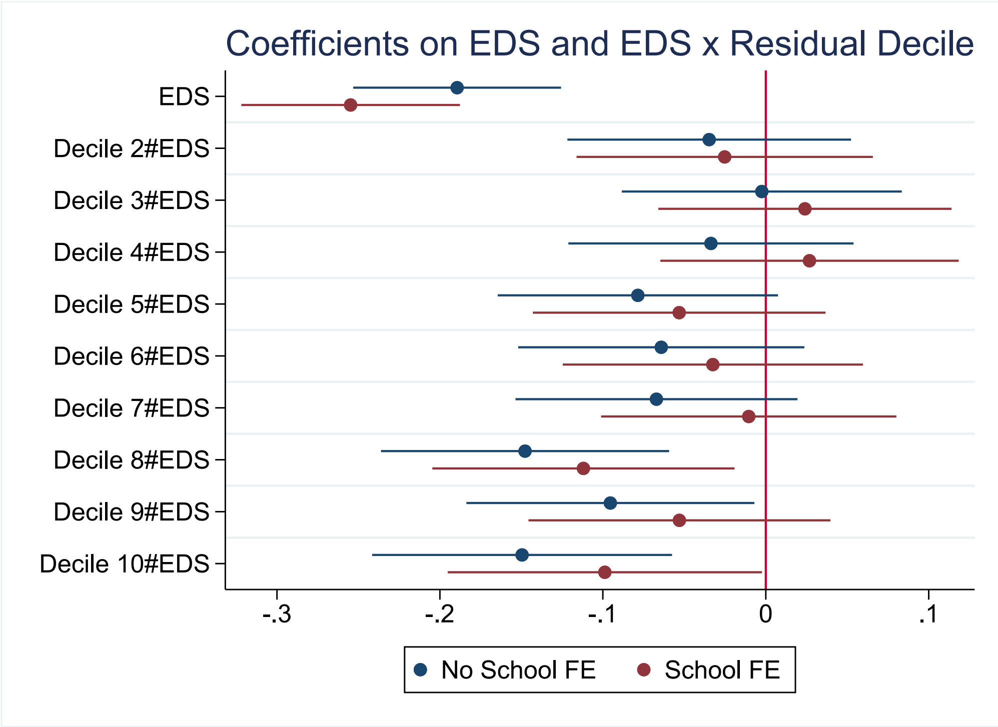


Figure 12: Coefficients on Residual Decile, Males

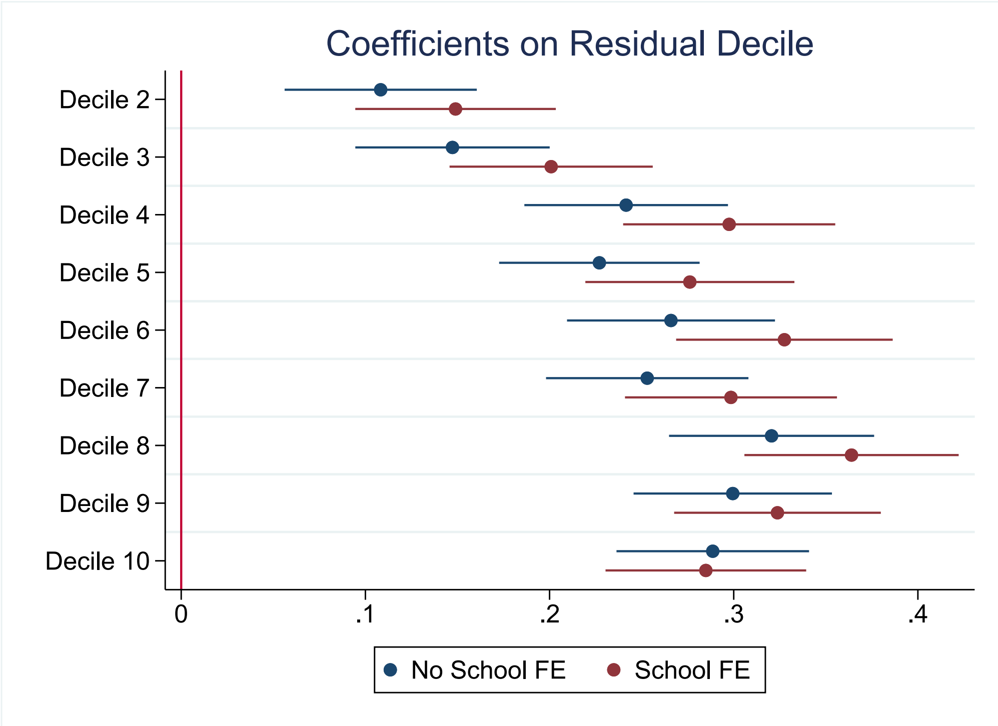
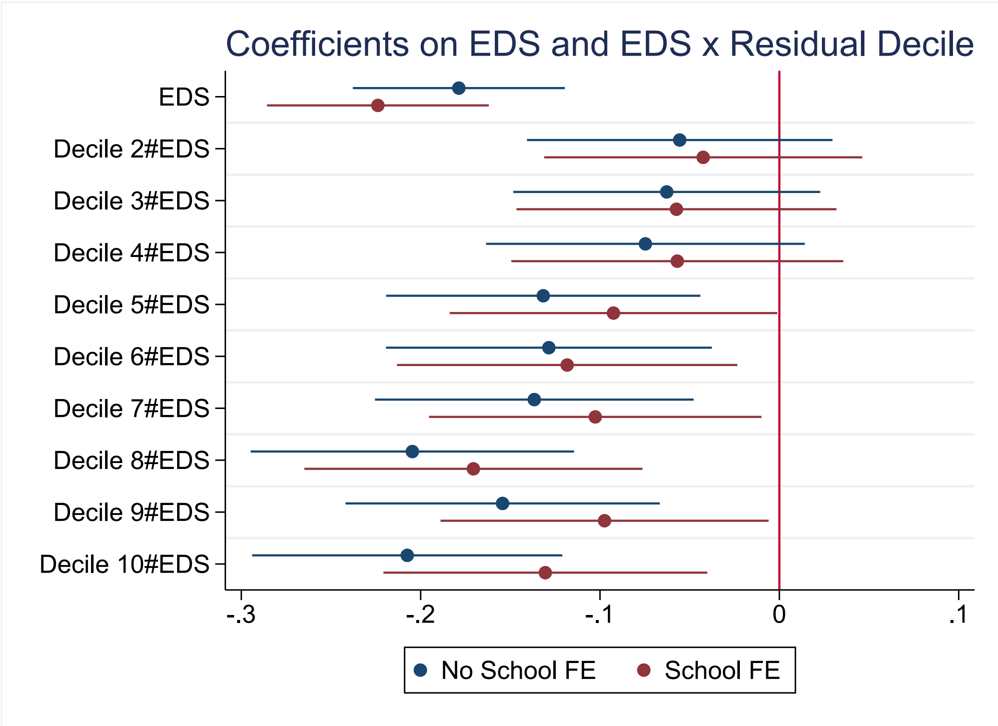


Figure 13: Coefficients on EDS and EDS × Residual Decile, Males



6 Closing the Income Gap in Multiple Test-Taking

How much would the income gap in admissions-relevant test scores close if disadvantaged students took multiple admissions tests at the same rate as observably similar non-disadvantaged students? To understand this question, I perform a simple simulation exercise. I equate multiple test-taking rates by disadvantaged status among students in the same cohort with the same initial ACT score as well as the same GPA, gender, and race. To understand how initial ACT score, GPA, gender, and race impact multiple test-taking rates among non-disadvantaged students in the true data, I estimate a probit regression of multiple test-taking on initial ACT score, weighted GPA, gender, race, and cohort fixed effects, estimated on the sample of non-disadvantaged students only. I compare students across schools rather than including school fixed effects; therefore, this simulation exercise can be partially interpreted as closing cross-school gaps in multiple test-taking behavior.

$$\mathbb{1}(\text{took multiple tests})_i = \Phi(\beta_{ACT}ACT_i + \beta_{GPA}GPA_i + \beta_X X_i + \gamma_t) \tag{11}$$

Here, X_i includes race and gender. I simulate multiple test-taking among both disadvantaged and non-disadvantaged students using the estimated coefficients from this regression, $\hat{\beta}$ and $\hat{\gamma}$, and students' true characteristics taken from the data. This equates multiple test-taking rates of disadvantaged students with those of non-disadvantaged students with the same observable characteristics. Specifically, I simulate multiple test-taking for student i in cohort t using a Bernoulli random variable with mean $\hat{\pi}$ defined as follows.

$$\hat{\pi}_i = \Phi(\hat{\beta}_{ACT}ACT_i + \hat{\beta}_{GPA}GPA_i + \hat{\beta}_X X_i + \hat{\gamma}_t) \tag{12}$$

I simulate score improvements conditional on multiple test-taking analogously, equating mean score improvements by disadvantaged status among students in the same cohort with the same initial ACT score as well as the same GPA, gender, and race. To understand how score improvements vary with initial ACT score, GPA, gender, and race among non-disadvantaged students in the true data, I estimate an OLS regression of score improvement on initial ACT score, weighted GPA, gender, race, and cohort fixed effects (but not school fixed effects), estimated on the sample of non-disadvantaged students only.

$$\text{score change}_i = \delta_{ACT}ACT_i + \delta_{GPA}GPA_i + \delta_X X_i + \xi_t + \epsilon_{it} \tag{13}$$

X_i includes race and gender. I simulate score improvements, conditional on multiple test-taking, among both disadvantaged and non-disadvantaged students using the estimated coefficients from this regression, $\hat{\delta}$ and $\hat{\xi}$, and students' true characteristics from the data.

$$\widehat{\text{score change}}_i = \hat{\delta}_{ACT}ACT_i + \hat{\delta}_{GPA}GPA_i + \hat{\delta}_X X_i + \hat{\xi}_t \tag{14}$$

Figure 14 plots kernel density estimates of the true and simulated test score distributions by disadvantaged status. The simulation shifts the admissions-relevant test score distribution among disadvantaged students to the right. Table 4 demonstrates that equating multiple test-taking rates and score changes by disadvantaged status, conditional on GPA, gender, and race, closes the gap in multiple test-taking rates by approximately 8.3 percentage points (33%), from 25.2 percentage points to 16.9 percentage points, and closes the income gap in admissions-relevant test scores by 8 SAT points (5%).

Figure 14: Simulated and Observed Test Score Distributions

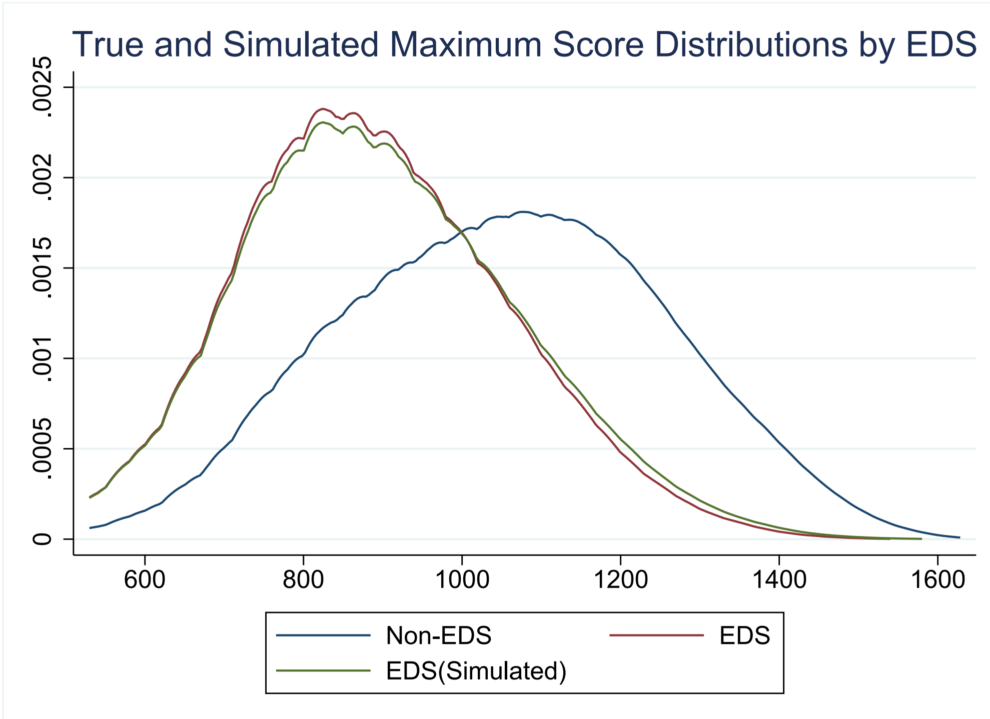


Table 4: Simulated Multiple Test-Taking by EDS

	Non-Disadvantaged Mean	Disadvantaged Mean	Difference
Observed			
Took Multiple Tests	.5647	.3125	.2522
Score Change	54.83	59.60	-4.77
Maximum Test Score	1053	889.5	163.5
Simulated			
Took Multiple Tests	.5762	.4071	.1691
Score Change	57.69	62.50	-4.81
Maximum Test Score	1053	897.3	155.7
Observations	226972	161661	65211

7 Conclusion

Using administrative data on the universe of public school students in North Carolina, I provide the first evidence of income disparities in strategic multiple test-taking behavior under universal ACT testing. I find that low-income students are less likely than their peers to take the SAT in addition to the state-mandated ACT, and that they improve upon their initial ACT scores by less when doing so. These disparities persist within schools and after conditioning on prior academic achievement; as a result, the widely used admissions practice of superscoring moves economically disadvantaged students down in test score rank relative to observably similar non-disadvantaged students.

Additionally, I show that disparities in multiple test-taking behavior by gender and disadvantaged status are partially driven by differential responses to the ability signal sent by a student's initial ACT score. I leverage detailed information on students' past academic performance to predict their scores on the initial ACT test. I find that, while non-disadvantaged boys who overperform on the initial ACT test are more likely to take additional tests, responses to initial ACT performance are muted for girls and disadvantaged students.

Finally, I use a simple simulation exercise to show that closing the income gap in multiple test-taking behavior and the resulting score improvements, conditional on initial ACT score, race, and gender, would reduce the income gap in admissions-relevant test scores by roughly 5%. The remainder of the income gap in admissions-relevant test scores can be accounted for by disparities in initial ACT scores, suggesting that the benefits of universal ACT testing in North Carolina are highly uneven.

References

Altonji, J., Arcidiacono, P., and Maurel, A. (2016). The Analysis of Field Choice in College and Graduate School. In *Handbook of the Economics of Education*, volume 5, pages 305–396. Elsevier.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Bailey, M. and Dynarski, S. (2011). Gains and Gaps: Changing Inequality in U.S. College Entry and Completion. Working Paper w17633, National Bureau of Economic Research, Cambridge, MA.

Belley, P. and Lochner, L. (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital*, 1(1):37–89.

Bloem, M. D., Pan, W., and Smith, J. (2021). College entrance exam-taking strategies in Georgia. *Southern Economic Journal*, 88(2):587–627.

Bond, T. N., Bulman, G., Li, X., and Smith, J. (2018). Updating Human Capital Decisions: Evidence from SAT Score Shocks and College Applications. *Journal of Labor Economics*, 36(3):807–839.

College Board (2015). SAT Score-Use Practices by Participating Institutions. Technical report.

Cook, E. E. and Turner, S. (2019). Missed Exams and Lost Opportunities: Who Could Gain From Expanded College Admission Testing? *AERA Open*, 5(2):1–20.

Foote, A., Schulkind, L., and Shapiro, T. M. (2015). Missed signals: The effect of ACT college-readiness measures on post-secondary decisions. *Economics of Education Review*, 46:39–51.

Frisancho, V., Krishna, K., Lychagin, S., and Yavas, C. (2016). Better luck next time: Learning through retaking. *Journal of Economic Behavior & Organization*, 125:120–135.

Goodman, J., Gurantz, O., and Smith, J. (2020). Take Two! SAT Retaking and College Enrollment Gaps. *American Economic Journal: Economic Policy*, 12(2):115–158.

Goodman, S. (2016). Learning from the Test: Raising Selective College Enrollment by Providing Information. *Review of Economics and Statistics*, 98(4):671–684.

Hoxby, C. M. and Turner, S. (2015). What High-Achieving Low-Income Students Know About College. *American Economic Review*, 105(5):514–517.

Hurwitz, M., Smith, J., Niu, S., and Howell, J. (2015). The Maine Question: How Is 4-Year College Enrollment Affected by Mandatory College Entrance Exams? *Educational Evaluation and Policy Analysis*, 37(1):138–159.

Hyman, J. (2017). ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice. *Education Finance and Policy*, 12(3):281–311.

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., and Dunlop Velez, E. (2015). The condition of education 2015 (NCES 2015- 144). *Washington, DC: Department of Education, National Center for Education Statistics.*

Klasik, D. (2013). The ACT of Enrollment: The College Enrollment Effects of State-Required College Entrance Exam Testing. *Educational Researcher*, 42(3):151–160.

Krishna, K., Lychagin, S., and Frisancho, V. (2018). Retaking in high stakes exams: Is less more? *International Economic Review*, 59(2):449–477.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Cataldi, E. F., and Mann, F. B. (2018). The Condition of Education 2018. NCES 2018-144. *National Center for Education Statistics.*

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

Papay, J. P., Murnane, R. J., and Willett, J. B. (2016). The Impact of Test Score Labels on Human-Capital Investment Decisions. *Journal of Human Resources*, 51(2):357–388.

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121(1-2):297–317.

Vigdor, J. L. and Clotfelter, C. T. (2003). Retaking the SAT. *Journal of Human Resources*, 38(1):1–33.

Westrick, P. A., Marini, J. P., Young, L., Ng, H., Shmueli, D., and Shaw, E. J. (2019). Validity of the SAT for Predicting First-Year Grades and Retention to the Second Year. *College Board*.

A Appendix

A.1 Robustness of Multiple Test-Taking Regressions

Table A1: Multiple Test-Taking and Student Characteristics, Polynomial Specification

	(1) Took Multiple	(2) Score Change	(3) Rank Change
ACT Score	0.000808	-2.216***	0.0455***
	(1.07)	(-32.94)	(9.62)
ACT^2	0.000000563	0.00167***	-0.0000807***
	(0.73)	(26.65)	(-16.18)
ACT^3	-6.58e-10*	-0.000000462***	3.02e-08***
	(-2.57)	(-24.13)	(18.02)
Weighted GPA	0.771***	50.83***	-4.121***
	(5.72)	(5.12)	(-10.87)
GPA^2	0.0859*	-7.491*	3.086***
	(1.97)	(-2.56)	(24.38)
GPA^3	-0.00985*	1.059***	-0.339***
	(-2.16)	(3.80)	(-25.20)
Female	0.0251***	-22.64***	-1.285***
	(4.28)	(-59.59)	(-38.27)
Economically disadvantaged (EDS)	-0.307***	-1.568**	-0.650***
	(-32.22)	(-2.87)	(-16.80)
Black	0.484***	-4.065***	1.488***
	(35.60)	(-6.46)	(20.45)
Black × EDS	0.215***	-4.338***	-0.210*
	(14.31)	(-4.86)	(-2.52)
Hispanic	-0.175***	-4.020***	-0.835***
	(-10.99)	(-4.47)	(-10.75)
Hispanic × EDS	0.178***	-6.286***	-0.107
	(8.90)	(-5.34)	(-1.22)
Asian	0.164***	15.49***	0.953***
	(6.41)	(9.98)	(8.09)
Asian × EDS	0.0952*	-19.84***	-1.156***
	(2.34)	(-8.57)	(-5.64)
Cohort fixed effects	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes
Probit	Yes	No	No
OLS	No	Yes	Yes
N	325046	155964	325567
R^2/Pseudo-R^2	0.3320	0.290	0.191

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Multiple Test-Taking and Student Characteristics, Linear Probability Model

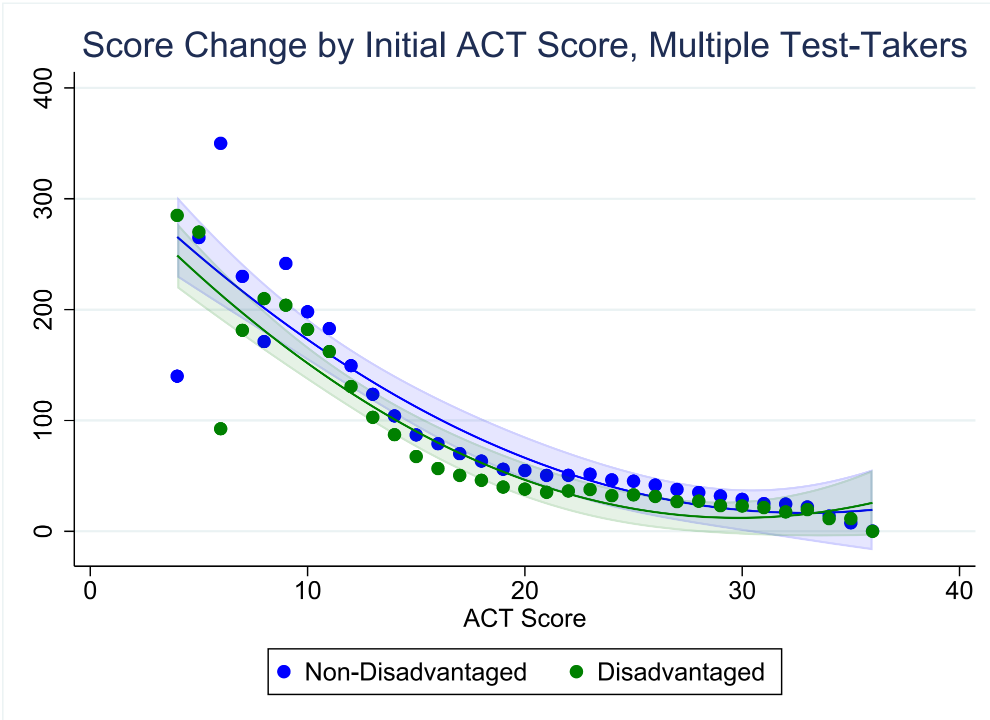
	(1) Took Multiple
ACT Score	-0.0000120
	(-1.42)
Weighted GPA	0.281***
	(160.50)
Female	0.00853***
	(5.38)
Economically disadvantaged (EDS)	-0.0828***
	(-32.42)
Black	0.133***
	(35.14)
Black × EDS	0.0430***
	(10.48)
Hispanic	-0.0506***
	(-11.45)
Hispanic × EDS	0.0466***
	(8.94)
Asian	0.0176**
	(2.95)
Asian × EDS	0.0463***
	(4.33)
Cohort fixed effects	Yes
School fixed effects	Yes
N	325567
R^2	0.377

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2 Score Improvements by Initial ACT Score

Figure A1: Mean Score Improvement by Initial ACT Score and EDS



A.3 Robustness of Multiple Test-Taking and Belief Updating

Table A3: Multiple Test-Taking and ACT Score Residual, Linear Probability Model

	(1) Female	(2) Male
Predicted ACT Score	0.0525***	0.0518***
	(51.12)	(50.73)
ACT Score Residual	0.00163	0.0104***
	(1.38)	(10.47)
Economically disadvantaged	-0.0707***	-0.0749***
	(-18.64)	(-19.93)
Economically disadvantaged × Residual	-0.00275	-0.00482**
	(-1.54)	(-3.15)
School Fixed Effects	Yes	Yes
N	67776	65002
R^2	0.164	0.157

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Multiple Test-Taking and ACT Score Residual, Full Sample

	(1) Female	(2) Male
Predicted ACT Score	0.138***	0.139***
	(114.60)	(118.29)
Act Score Residual	-0.0111***	0.00770***
	(-4.55)	(3.67)
Economically disadvantaged	-0.203***	-0.214***
	(-24.29)	(-24.65)
Economically disadvantaged × Residual	0.00427	0.00700*
	(1.14)	(2.02)
School Fixed Effects	Yes	Yes
N	134312	130392
R^2	.1994	.2095

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Multiple Test-Taking and ACT Score Residual, Full Set of ACT Score Predictors

	(1) Female	(2) Male
ACT Score Residual	0.00998**	0.0242***
	(2.82)	(7.85)
Economically disadvantaged	-0.210***	-0.222***
	(-17.19)	(-16.85)
Economically disadvantaged × Residual	-0.00486	-0.00740
	(-0.89)	(-1.45)
School Fixed Effects	Yes	Yes
Lagged Test Score Covariates	Yes	Yes
GPA Covariate	Yes	Yes
N	67526	64655
Pseudo-R^2	0.2257	0.2456

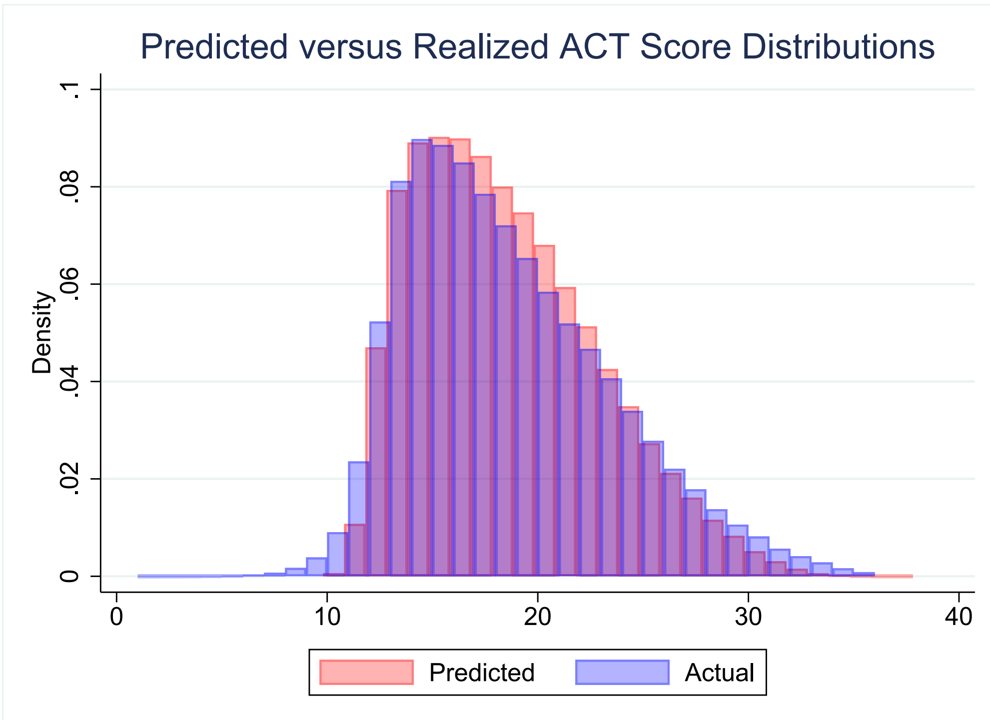
t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.4 ACT Score Prediction Model Fit

Figure A2 compares the predicted ACT score distribution with the true ACT score distribution. The prediction model fits the shape of the ACT score distribution well with an R^2 of 0.795, despite slightly underpredicting mass at the upper and lower tails.

Figure A2: Prediction Model Fit



A positive value of the residual indicates that a student overperformed on the ACT test relative to his or her past performance, and vice versa. The standard deviation of the residual is 2.20. The standard deviation of both the predicted ACT score and the ACT score residual is slightly smaller among disadvantaged students, as demonstrated in Figures A3 and A4, respectively.

Figure A3: Distribution of Prediction ACT Score $\hat{ACT}_{ist}$

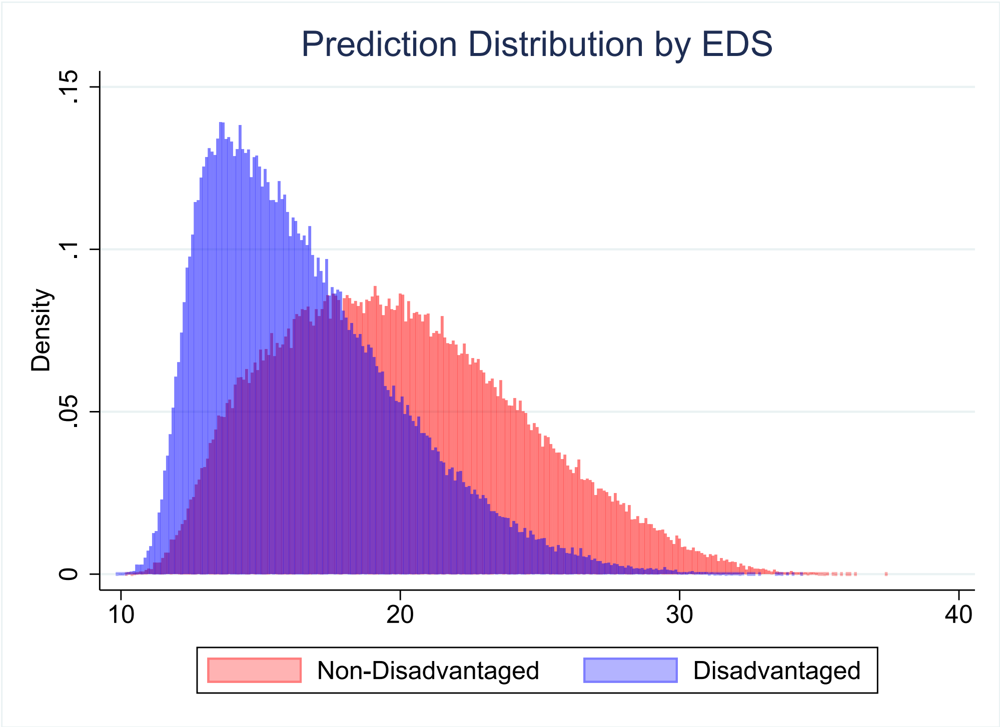


Figure A4: Distribution of the Residual $\hat{\epsilon}_{ist}$ By EDS