

Essays in the Economics of Education

by

Ying Shi

Department of Public Policy Studies
Duke University

Date: _____

Approved:

Seth Sanders, Supervisor

Peter Arcidiacono

Charles Clotfelter

V. Joseph Hotz

Jacob Vigdor

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Public Policy Studies
in the Graduate School of Duke University
2017

ABSTRACT

Essays in the Economics of Education

by

Ying Shi

Department of Public Policy Studies
Duke University

Date: _____

Approved:

Seth Sanders, Supervisor

Peter Arcidiacono

Charles Clotfelter

V. Joseph Hotz

Jacob Vigdor

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Public Policy Studies
in the Graduate School of Duke University
2017

Copyright © 2017 by Ying Shi
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Abstract

This dissertation comprises three essays in the economics of education. I begin with a paper that evaluates the effectiveness of selective secondary schools. An unique admissions context permits identifying the causal benefits of such institutions for a more heterogeneous sample of students than previous US-based studies. The second essay examines the causes of female under-representation in STEM fields, with a focus on engineering. I decompose the gender gap into explanatory accounts such as academic preparation, ability beliefs, and preferences for prosocial values and professional goals. The final essay investigates the roles of cognitive and non-cognitive skills in explaining high school graduation gaps at the three-way intersection of race, gender, and income. A finding of particular interest is the lagging performance of disadvantaged white students relative to African American peers, even after accounting for skill disparities using a sequential model of educational attainment.

Contents

Abstract	iv
List of Tables	viii
List of Figures	xi
Acknowledgements	xii
1 Introduction	1
2 Who Benefits from Selective Schools? Evidence from High-Achieving Math and Science Students in North Carolina	4
2.1 Related evidence and potential mechanisms	9
2.2 Institutional context and data	13
2.2.1 Admissions	14
2.2.2 Data	15
2.2.3 Linking applicants to statewide administrative records	18
2.2.4 Applicant profile	20
2.3 Empirical methodology	22
2.3.1 Fuzzy regression discontinuity	22
2.3.2 Defining admission thresholds s_{dt}^*	25
2.3.3 RD validity: covariate smoothness	33
2.4 Results	34
2.4.1 Role of selection	34
2.4.2 Graphical evidence	37

2.4.3	Reduced form RD	39
2.4.4	Heterogeneous treatment effects	40
2.4.5	Mechanisms	49
2.5	Role of admissions policies	54
2.6	Conclusion	60
3	The Puzzle of Missing Female Engineers: Academic Preparation, Ability Beliefs, and Preferences	63
3.1	Determinants of STEM gender gaps	66
3.2	Data and descriptive statistics	71
3.2.1	North Carolina secondary schools	71
3.2.2	University of North Carolina	72
3.2.3	CIRP Freshmen Survey	74
3.3	The engineering pipeline	76
3.4	Empirical approach	79
3.5	Results	82
3.5.1	High school academic achievement and extracurriculars	82
3.5.2	Ability beliefs, prosocial values and other preferences	87
3.5.3	Family background and structure	94
3.6	Conclusion	97
4	A Class of Their Own: Explaining High School Attainment Gaps by Income, Race, and Gender	100
4.1	Data	105
4.1.1	NCERDC	106
4.1.2	NLSY97	109
4.2	Reduced form evidence	110
4.2.1	Attainment disparities by subgroup	110

4.2.2	Alternative explanations	112
4.2.3	OLS	117
4.3	Factor model	119
4.3.1	Measurement system	120
4.3.2	Student outcomes	124
4.3.3	Estimation	125
4.4	Results	126
4.5	Conclusion	129
A	CHAPTER 1 - TABLES	132
B	CHAPTER 1 - ROBUSTNESS CHECKS	137
C	CHAPTER 2 - SUMMARY STATISTICS	144
D	CHAPTER 3 - TABLES AND FIGURES	149
	Bibliography	154
	Biography	164

List of Tables

2.1	Match rates	19
2.2	Summary statistics - applicant sample	21
2.3	Variation in applicant quality by CD	21
2.4	Cutoff definition and predictive accuracy	28
2.5	Predicted enrollment for all applicants and compliers	31
2.6	Summary statistics for marginal applicants	32
2.7	SAT math outcomes and selection bias	36
2.8	Reduced form results	40
2.9	Heterogeneous treatment effects	43
2.10	Heterogeneous treatment effects by CD	46
2.11	11th grade characteristics of compliers	48
2.13	Admissions rules and summary statistics	57
2.14	Admissions rules and summary statistics by CD	58
2.15	Marginal treatment effects across 3 admission rules	59
3.1	Share of recent college graduates in STEM fields	67
3.2	The engineering gender gap from high school to college	77
3.3	Attrition among engineering students	78
3.4	High school gender gap in engineering intentions	83
3.5	Gender gap after 30 credit hours in UNC	84
3.6	Gelbach decomposition: High school and UNC samples	86

3.7	Gender gap in CIRP Freshmen Survey	88
3.8	Gelbach decomposition: CIRP Freshmen Survey	90
3.9	Engineering orientation among twin sample	96
4.1	NCERDC - Summary statistics	111
4.2	NCERDC - Summary statistics for alternative SES definitions	116
4.3	NLSY97 - Summary statistics	117
4.4	Linear probability model: Complete 12th grade	119
4.5	Structural high school continuation parameter estimates	127
4.6	Structural factor parameter estimates	128
4.7	Structural cognitive measure parameter estimates	128
4.8	Structural non-cognitive measure parameter estimates	129
A1	Cutoff definition 2	132
A2	Cutoff definition 3	133
A3	Balance of black applicants across cutoff	133
A4	Baseline statistics	134
A5	Heterogeneous treatment effects - nonparametric estimates	135
A6	Heterogeneous treatment effects - nonparametric estimates using race-based cutoffs	136
B1	Robustness - inclusion of race	139
B2	Robustness - balance under race-based cutoffs	140
B3	Robustness - unique matches	142
C1	Engineering orientation in second half of high school	145
C2	Majoring in engineering at UNC	146
C3	Freshmen engineering intentions: CIRP Survey	147
C4	Engineering orientation: twin sample	148
D1	Grade progression by year using the NCERDC	149

D2	Sample attrition and 12th grade graduation outcomes	150
D3	Sample attrition and share of each student group	150
D4	LPM by year: Complete 12th grade	152
D5	Results by block group characteristic	153

List of Figures

2.1	CD boundaries: 2009 - 2012 application cohorts	15
2.2	Admission thresholds in 2012	27
2.3	First stage relationship	29
2.4	RD validity - covariate smoothness	34
2.5	Outcomes - average residuals	38
2.6	Heterogeneous treatment effects by entering SAT math scores	45
2.7	Discontinuities in peer achievement and percentile rank	50
3.1	Decomposition by race	92
3.2	Decomposition by SAT math scores	93
B1	McCrary density around the cutoff	138
B2	Distribution of STEM-intensity coefficient in bootleg sampling	143
D1	On track and graduating by Year 4	151

Acknowledgements

The path to putting together this dissertation involved the unflagging support and advice of many. I must begin with my advisor Seth Sanders, who gave his time, advice, and encouragement throughout. His mentorship has played and continue to play a pivotal role in my professional and academic development. I have also been the beneficiary of several generous committee members: Peter Arcidiacono, Charles Clotfelter, V. Joseph Hotz, and Jacob Vigdor. Their input and feedback enriched my learning and growth as a researcher. Also not to be overlooked are my interactions with Sunny Ladd and Marcos A. Rangel. Instances of their advice and support number too many over the years to enumerate.

I would like to extend my gratitude to those who have provided helpful feedback on various drafts of my research, including but not limited to Elizabeth Ananat, Philip Cook, William “Sandy” Darity, Robert Garlick, Hugh Macartney, Manoj Mohanan, as well as seminar participants in the Duke Labor Lunch Group, Sanford Seminar Series, Graduate Research Workshops, IZA Brown Bag Seminar, AEFPP and APPAM annual conferences, School of Public and Environmental Affairs at Indiana University, University of Hawaii-Manoa, the Urban Institute, RAND Corporation, Humboldt-Universität zu Berlin, and the Center for Education Policy Analysis at Stanford University. This work has been generously supported by the American Education Research Association (AERA) Minority Dissertation Fellowship and a Horowitz Foundation for Social Policy grant. My research is also enabled by access

to data provided by the North Carolina Education Research Data Center (facilitated with the help of Ken Dodge, Clara Muschkin, and Kara Bonneau), the University of North Carolina, the Higher Education Research Institute, and the administration of the selective North Carolina high school.

This dissertation would not have been possible without the contributions of family and friends. I am grateful for the unconditional support of my parents, who provided unwavering encouragement for me to pursue my interests in public policy and economics. These past few years would not have been the same without being able to share the joys and tribulations of the Ph.D. with my friends at Duke, as well as friends outside of graduate school (a special mention goes out to Jesse Saloom). Lastly, I cannot overstate the role played by my partner Johannes Himmelreich, whose patience, humor, and support kept me on the path to completion. I dedicate this dissertation to him.

1

Introduction

The three essays that comprise my dissertation are motivated by an unifying interest in understanding the causes of educational inequities and evaluating the consequences of policies for these disparities and student outcomes.

In Chapter 2, I examine the effectiveness of one increasingly prevalent form of school choice: selective secondary schools. Existing evidence finds mostly null or minimal gains from attending these elite institutions. I clarify the prevalence of null findings using new data from a selective North Carolina secondary school. Previous regression discontinuity studies use school-level score cutoffs to estimate treatment effects on a relatively homogeneous group of marginal students. A key contribution of this paper is the ability to derive causal estimates for a more heterogeneous sample, using a district quota-based admissions procedure that generates multiple cutoffs per cohort. I find that null or small treatment effects across multiple outcomes disguise sizable differences by student background. More disadvantaged students who have lower quality options outside of elite education experience larger test score gains and apply to a greater share of selective colleges. In contrast, treatment effects are null among more privileged students. Attending the selective school also leads

students to apply to colleges that confer more STEM bachelor's degrees, with the largest gains accruing to those from more academically advantaged backgrounds. These patterns of heterogeneous effects underscore an efficiency-based rationale for inclusive elite education that is compatible with equity. An admissions policy that enrolls more under-represented students achieves higher test score gains while bridging gaps. Meanwhile, tradeoffs still exist for institutional goals such as increasing postsecondary STEM participation, which may be easier attained under a more advantaged student body.

Chapter 3 analyzes the under-representation of women in STEM fields, with a focus on engineering. I use administrative data from North Carolina and national surveys of high school and college freshmen to characterize the evolution and determinants of the largest STEM gender gap contributor. The sizable engineering gender disparity in grade 9 and the broadening of the gap through high school, coupled with similar rates of attrition during college, suggest that observed disparities are driven by differential entry rather than exit. I decompose the engineering gap into several explanatory accounts, including differences in academic preparation, academic ability beliefs, and preferences for prosocial values and professional goals. Disparities in SAT scores and high school GPA account for 5 to 9% of the aggregate gap. Female students' lack of academic self-confidence, even after conditioning on objective measures of performance, explain 8%. Other-regarding preferences and professional goals in the arts and sciences capture a further 15%. Empirical evidence using identifying variation in the gender composition of twins in North Carolina shows that opposite-sex pairs are more likely to pursue gender-stereotypical majors. The findings are consistent with major orientation being affected by family structure and gender-based expectations and norms.

Lastly, the co-authored Chapter 4 with Brian Clark documents the role that cognitive and non-cognitive skills play in determining students' high school graduation

outcomes. Inequalities in educational attainment beget large disparities in future economic well-being. Despite its manifest importance, little evidence document high school graduation gaps between groups defined at the three-way intersection of race, gender, and income, and also to systemically account for gaps at this level. In this paper, we use administrative North Carolina data to present a more fine-grained view of attainment disparities. Graduation rates bifurcate completely by socioeconomic status (SES), with similar attainment among advantaged black and white students, while economically disadvantaged white students lag behind the attainment of their disadvantaged black peers, particularly among females. The relative underperformance of poorer white students is concentrated among the most economically deprived, and cannot be explained by income differences.

To explain these striking differences, we consider a rich set of school-based measures that capture dimensions of cognitive and non-cognitive skills. Reduced form results accounting only for cognitive skills fully explain educational gaps between advantaged white females and all black student subgroups with the exception of poor black males. Next we specify a sequential model of educational attainment that enables us to separately identify the contributions of cognitive and non-cognitive skills. Non-cognitive skills exert greater influence on attainment outcomes relative to cognitive skills, and the explanatory power of factors decline later into schooling. We find that economic deprivation among black students translates to cognitive and non-cognitive skill gaps associated with lower high school graduation rates. In contrast, economic disadvantage among white students materialize in different ways to put downward pressure on graduation rates. For one, their markedly higher cognitive scores do not translate into academic success.

Who Benefits from Selective Schools? Evidence from High-Achieving Math and Science Students in North Carolina

Elite secondary schools are expanding in increasingly specialized and selective education sectors across the globe. Proponents of these schools argue that they give high-ability children from all backgrounds a rigorous education that prepares them for entry into competitive colleges and science- and math-oriented careers. Yet there is limited evidence of selective schools exerting meaningful causal effects on student performance and long-term outcomes. Studies yield mixed results in international settings and only null or small effects in promoting student achievement and postsecondary outcomes among US-based schools (Cullen, Jacob, and Levitt, 2006; Jackson, 2010; Pop-Eleches and Urquiola, 2013; Abdulkadiroglu, Angrist, and Pathak, 2014; Dobbie and Fryer, 2014).

A major challenge in evaluating elite schools is the non-random selection of students, with schools often using entrance exams or sets of performance criteria to determine admissions. A naive comparison between enrollees and non-enrollees of selective schools almost surely overstate their causal effects. The literature overcomes

this selection by relying on a regression discontinuity (RD) approach or by comparing the outcomes of students admitted or rejected via lottery among over-subscribed schools. RD designs identify treatment effects using discontinuities in admissions probabilities. Since these studies typically rely on one score cutoff per institution, conclusions about the effectiveness of elite schools are drawn from a relatively homogeneous sample of applicants in the neighborhood of the single admissions cutoff.¹ Interpretation of treatment effects are then limited to marginal students who, if admitted, belong to the bottom of the selective school cohort in relative baseline ability.

This paper innovates on existing literature by investigating a selective North Carolina secondary school with an admissions process yielding a more heterogeneous set of marginal students. The study's strength derives from legislative rules mandating the school to equitably serve all thirteen congressional districts (CD) in the state. This pits applicants against peers from the same residential district. As a result, the within-district nature of the admissions process guarantees a range of admissions thresholds based on varying levels of over-subscription and applicant quality. Causal effects are then identifiable for marginal students from different socio-demographic backgrounds, ability levels, and geographic areas who are exposed to the same selective school environment.

The opportunity to study effect heterogeneity prompts questions on the ideal set of students to admit. Previous studies have taken the admissions policy as given, even though schools have substantial autonomy over the selection process. Here I explicitly consider the admissions policy's role in estimating average treatment effects. If all students benefit uniformly from selective school attendance, then altering the admissions policy to enroll a different student composition has no effect on the

¹ An exception is Lucas and Mbiti (2014), which uses multiple district-specific cutoffs for each government secondary school in Kenya.

school's effectiveness. However, treatment effects will change under alternative admissions rules if different subgroups of students derive varying levels of benefit from attendance. The paper models student composition and counterfactual treatment effects under common admissions policies to emphasize that the way schools select their students matters for evaluations of effectiveness.

Another contribution of this paper is that it directly tests for the effect of attending a science- and math-intensive school on students' predilections for these fields. One justification for the existence of selective institutions is that they increase the ranks of science, technology, engineering, and math (STEM) professionals vital to a knowledge economy. STEM academic tracks also bolster students' future economic well-being. Limited evidence exists, however, on the causal effect of science- and math-oriented schools on students' STEM trajectories.² By estimating the causal impact on major and postsecondary STEM orientation, this study is well suited to comment on one of the primary goals of STEM institutions.

To conduct the analyses, I merge newly obtained applicant data from the selective school with administrative files from the North Carolina Department of Public Instruction. The resulting longitudinal dataset permits me to follow students from primary education through the end of high school. In addition to socio-demographic attributes and detailed academic histories, I am able to characterize applicants' sending school environments using rich student- and school-level data. I examine outcomes including standardized test scores, major intentions, and college application behavior at the end of high school. RD findings at the aggregate level echo the null or small treatment effects common in this literature. Students predicted to enroll in the selective school based on their admission scores do not perform meaningfully better in SAT math or reading at the end of high school. However, I find some evidence

² Wiswall et al. (2014) is one of the few studies focused solely on the causal impact of STEM institutions. The authors find insignificant effects under the assumption of selection-on-observables.

of an increase in the share of selective institutions in students' college application portfolios. These applications also comprise more STEM-intensive institutions as measured by the share of bachelor's degrees awarded in STEM. Attending the selective school raises the average STEM-intensity of college applications by 1.8 to 2.4 percentage points, an increase of 8 - 11%. Importantly, these average effects disguise important differences in who benefits from exposure to selective education.

I separately evaluate treatment effects by students' baseline math scores, demographic characteristics such as race and socioeconomic status (SES), and sending school quality. I find that students who are economically disadvantaged or come from lower-achieving schools reap the largest gains in SAT scores at the end of high school. These students achieve math and reading gains of approximately 3 to 5 percentile points, compared to mostly null results among more privileged peers. Disadvantaged students experience a 4 to 6 percentage point increase in the share of selective colleges to which they apply. Since disadvantaged students have lower baseline scores and apply to fewer selective universities, the selective school experience is in effect bridging existing disparities in these outcomes. A different pattern emerges for the STEM-intensity of students' preferred colleges. Significantly positive effects are observed for almost all subgroups, with the largest effects accruing to students who have higher baseline math scores or are from more advantaged academic backgrounds. The preponderance of positive effects among disadvantaged students shows that contrary to the mismatch hypothesis, under-represented students with relatively low baseline performance reap significant rewards from exposure to elite education.

What drives these differences in treatment effects? Two observations are expedient. Students who record null test score gains in my sample come from high quality public schools with lower shares of minority and low income peers. Within these institutions, applicants from the most privileged districts are tracked into math classes with peers who score 1.5 or more standard deviations above the statewide average

in standardized exams. Since these students have ample exposure to competitive academic settings, one explanation consistent with the null results is that the ‘treatment’ of attending a selective institution is simply smaller for advantaged students. The second observation is that mechanisms matter. Different students are sensitive to different dimensions of the selective school experience. Previous studies have focused on discontinuous jumps in average peer achievement when the student enrolls in a selective school. I consider this channel and another that is receiving increasing attention: relative rank. Given the unique nature of the quasi-natural experiment, enrolling in the selective school leads to higher-performing peers and a decrease in relative rank that vary by cohort and congressional district. I use these variations to instrument for both peer achievement and rank via two-stage least squares (2SLS).

Results show that the contributions of each mechanism to school effectiveness depend on the outcome studied. Higher performing peers, but not rank, induce students to apply to more selective colleges. Both better performing peers and higher rank encourage students to choose more STEM-intensive colleges. The contribution of rank to the latter outcome is notable, because STEM is commonly perceived as an academic track reserved for relatively high performing students in a school. Subgroups with the largest drops in rank, namely black students, did not apply to postsecondary institutions with a greater STEM focus. These results show that rank affects student outcomes independently of the peer quality channel and can inform variations in treatment effects across student backgrounds.

The presence of effect heterogeneity emphasizes that student composition matters for evaluating school effectiveness. I model how changes in student composition affect the magnitude of treatment under alternative admissions policies. Three policies are considered, the status quo, a rule that admits students solely on baseline math scores, and a rule that admits only those with the highest GPA-based class rank. The math score-based rule admits the fewest minority and economically disad-

vantaged students. Admitted students have lower gains in SAT math and the share of selective college applications, but apply to colleges that are more STEM-intensive than the two alternative rules.³ Taken together, the results suggest a tension governing efforts to close test score gaps and promote STEM participation. Greater gains in test scores and selective college applications among diverse students establish an efficiency-based argument for more inclusive selective schools. In addition to greater gains, enrolling more under-represented students also serves equity goals by reducing existing gaps. Greater STEM participation, in contrast, results from enrolling more over-represented students under the math score-based rule. These students' high baseline levels of preparation may permit them to take better advantage of the STEM opportunities afforded by the selective school, which translate into science- and math-oriented postsecondary aspirations. Different student compositions are therefore suited for attaining different types of institutional goals. An admissions policy requires weighing tradeoffs to accommodate the multidimensional and sometimes competing values of an elite institution.

The paper proceeds as follows. Section 2.1 discusses the findings and incongruities in related literature alongside a conceptual framework for how selective high schools may influence student performance. Section 2.2 describes the admissions process and dataset. Section 2.3 on empirical methodology outlines the fuzzy RD approach. I report reduced form and 2SLS results in Section 2.4 and examine the consequences of alternative admissions rules in Section 2.5 before concluding.

2.1 Related evidence and potential mechanisms

The preponderance of studies on selective schools finds null or relatively modest treatment effects. In particular, assessments of US-based exam schools yield little

³ An important caveat is that I do not explicitly model potential endogenous changes in student and school behavior in response to the changing student compositions, including but not limited to the formation of friendship and study groups and the ways students are assigned into classrooms.

evidence to suggest that students on the margin would not have performed just as well in terms of academic achievement and longer term postsecondary outcomes if they remained in their sending schools. The RD studies of Abdulkadiroglu, Angrist, and Pathak (2014) and Dobbie and Fryer (2014) find no consistent evidence that attending exam schools in Boston and New York affects test scores, AP outcomes, and college enrollment in the former, and longer term postsecondary outcomes in the latter among three New York schools. Winning the lottery at an over-subscribed high-achieving Chicago secondary school actually lowers achievement across several categories, although not all estimates are significant (Cullen, Jacob, and Levitt, 2006).

Research in non-US contexts yields mixed results. Attending a selective UK high school had a slight effect on test scores, while long-term outcomes such as university enrollment were more pronounced (Clark, 2010). Attendance at a selective Chinese middle school and elite Kenyan national schools did not measurably improve achievement scores (Zhang, 2012; Lucas and Mbiti, 2014). In contrast to these smaller effects, sizable gains in graduation exams are documented in the Romanian and Trinidad and Tobago contexts (Pop-Eleches and Urquiola, 2013; Jackson, 2010).

A parallel strand of literature on gifted and talented programs (GT) offering specialized instruction for tracked students also finds mixed effects. Bui, Craig, and Imberman (2014) use two sources of variation in a US public school district, one a discontinuous index score used to determine GT eligibility, and the other randomized lotteries in the case of oversubscribed magnet GT programs. Data evinces little evidence of higher student achievement with the exception of higher science scores among magnet GT program enrollees. Regression discontinuity estimates based on IQ thresholds in another US context also fail to record meaningful effects, although significant benefits accrue to high-performing minority students who miss the IQ-based cutoffs but still enroll in gifted classrooms (Card and Giuliano, 2014, 2016).

Meanwhile, some evidence exists on the long-term influence of these programs, from increased high school graduation and college enrollment (Cohodes, 2015) to the choice of a more challenging field of study (Booij, Haan, and Plug, 2015).⁴

In theory, there are multiple mechanisms through which attending selective schools can affect educational performance. One mechanism is the direct effect of peer-to-peer interactions in the classroom. Fellow students can impact classmates' academic interests and aspirations through social interaction and network spillovers. Multiple peer effects studies document a positive contribution of high quality peers (Hoxby, 2000; Hanushek et al., 2003; Hoxby and Weingarth, 2005; Vigdor and Nechyba, 2007; Ammermueller and Pischke, 2009; Imberman, Kugler, and Sacerdote, 2009). The dearth of "bad apples" decreases opportunities to disrupt learning, a public good that is accompanied by congestion effects (Lazear, 2001).⁵ In addition to direct benefits from the spillover effects of elite peers, students tracked to high achieving classrooms may also gain from teachers tailoring instruction to their academic level (Duflo, Dupas, and Kremer, 2011). If teachers revise pedagogy towards median student ability, a gifted student may find the average instructional level in a selective academic environment to be a better match for their academic skills. Thus elite schools can improve academic performance by providing higher average peer ability and a closer academic match via tailored instruction.

⁴ The contexts of these programs matter. For instance, Booij, Haan, and Plug (2015) studies GT program effectiveness in a selective Netherland secondary school that permits students to replace classroom lectures with self-selected projects. They conclude that this form of enrichment enables accumulation of academic skills and boosts students' academic esteem.

⁵ Related to peer inputs is a broader discussion on contributions of resource-based policies in education and the role of school quality. Some find that increased inputs into the education production function are unlikely to be effective in the absence of changed incentives (Hanushek, 1997, 2006). Among the few studies that directly examine the contributions of secondary school quality, Hoekstra, Mouganie, and Wang (2016) find that higher performance among enrollees at selective high schools in China are consistent with higher teacher quality. As such, the null or small treatment effects documented by exam school studies suggest that increases in measurable inputs are not producing better student outcomes, a conclusion that is not easily reconciled with positive effects found by many studies that examine specific inputs.

Meanwhile, opposing mechanisms may counteract these positive effects. If relative position is what matters, the big-fish-little-pond effect can erode academic self-concept and performance (Marsh, 1987). A recent literature supports a causal contribution of ordinal rank that can offset direct peer influence (Murphy and Weinhardt, 2014; Elsner and Isphording, 2015). Elite school attendance may also diminish other inputs into students' learning. There is suggestive evidence that parental inputs are substitutes with school quality, such that parents reduce effort when their children enroll in a better school (Pop-Eleches and Urquiola, 2013). If higher school quality prompts behavioral responses from parents in the form of reduced investments, the overall treatment effect can attenuate.

Another explanation for the lack of sizable treatment effects in this literature is the high quality of counterfactual schools. Studies with null findings may be estimating treatment effects for students from well-resourced institutions with experienced teachers and rigorous curricula. Among the three Boston exam schools studied in Abdulkadiroglu, Angrist, and Pathak (2014), for instance, non-admits to the elite Boston Latin School can place in the second most competitive exam institution, the Boston Latin Academy, or the high-performing John D. O'Bryant High School of Mathematics and Science. The experience of being 'treated' with higher-achieving peers is therefore more muted for top schools relative to O'Bryant, whose rejection results in the majority of applicants returning to relatively under-performing Boston public schools.⁶ The implication is that interpretation of treatment effects should consider the quality of the counterfactual school environment. The rich data environment of the North Carolina secondary school, as described below, enables a close

⁶ Among 7th grade compliers, 93% of applicants not offered a seat at the Boston Latin School is estimated to enroll in the Boston Latin Academy, 5% enrolls in O'Bryant, and only 3% enrolls in a traditional public school. While peer math scores among admitted compliers at the Latin School are 1.97 standard deviations above average, the average of non-admitted compliers is still well above state average at 1.15 standard deviations. The corresponding statistics for O'Bryant are 0.84 and -0.15 standard deviations. Among non-admits at O'Bryant, all are estimated to return to public schools.

examination of the relationship between counterfactual school environments and the magnitude of treatment effects.

2.2 Institutional context and data

The selective public high school in North Carolina that I study has a number of distinguishing features. The first is its standing as a residential institution, which enables the school to enroll students from a wider area than a comparable neighborhood public high school. The broad geographic coverage also owes to a legislative mandate that requires the school to equitably serve each of North Carolina's thirteen congressional districts. The result is a geographically diverse student body serving high achieving individuals from across the state.

Another feature is the institution's truncated grade span. The school accepts applications from 10th grade students and serves students in grades 11-12. While the exposure period is shorter than typical high schools, the application's timing during the middle of students' high school careers promotes continuity. Students who are not admitted are more likely to return to their home schools compared to a scenario with middle school applicants who are considering multiple high schools. With approximately 340 spots allocated annually, the selective school's residential enrollment totals between 650 and 700 students.

A founding vision for the school is to advance public education, with a particular focus on developing future leaders in STEM fields. All applicants must therefore demonstrate an existing interest in math and science. Once admitted, students can take advantage of a wide range of advanced courses limited not only to math and science. The school also emphasizes gender, racial and socioeconomic heterogeneity in its student population. The most recent cohort of enrolled students are 47% female, 49% Caucasian, 30% Asian American, 11% African American, and 4% Hispanic.

2.2.1 Admissions

Admission into the high school is highly competitive. The full set of application materials includes a completed form with demographic and sending school information, an academic transcript, recommendations from a counselor and three high school teachers, essay responses, and listings of extracurricular activities. In addition, applicants must take the SAT exam between October through January of their sophomore year. During the annual recruiting cycle, around 1000 tenth grade students apply for 340 slots.

At the beginning of the evaluation process, applications are sorted by the student's congressional district of residence. A legislative mandate obligates the school to equitably serve all thirteen districts in North Carolina. Within each district, three individuals from the pool of reviewers are assigned to rate each application using a common Admissions Rubric. The reviewers consider multiple criteria including academic rigor, quality of sending school, grades, extracurricular involvement, teacher and counselor evaluations, SAT scores, maturity, and interest in math and science. Points are allocated by category based on rubric guidelines on a scale of 0 to 40. The combined admissions rating score then ranges from 0 to 120. The legislative statute requires that the number of students admitted into a given district must be within 2.5 percent of the total number of bed spaces available that year.⁷ Large disparities in the number of applicants exist across districts, such that some congressional districts are significantly more over-subscribed than others. Applicants then compete with same-district peers for a limited set of open seats. Reviewers admit the highest scoring students from each district while ensuring diversity in the admitted student population. As such, they can exercise discretion in removing high-scoring students

⁷ A redrawing of congressional districts took place in 2011, and new district boundaries were adopted by the admissions committee in 2013. All cohorts in the present study were selected based on the pre-redistricting map (Figure 2.1).

or adding lower-scoring students to the final enrolled pool.

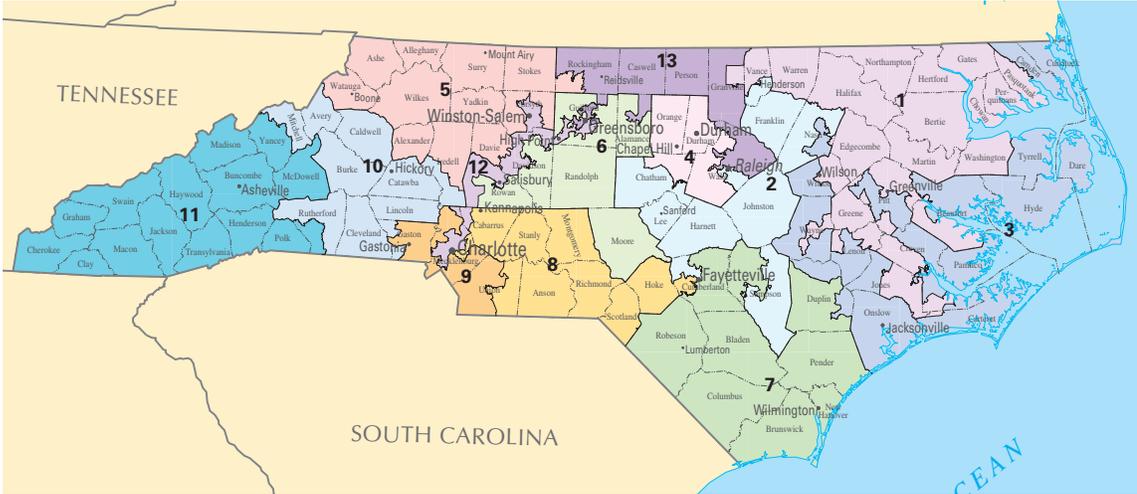


FIGURE 2.1: CD boundaries: 2009 - 2012 application cohorts

Crucially, applicants cannot predict district-specific cutoffs because these cutoffs are determined from a set of variables largely outside the student’s control. Thresholds are a function of the size, quality distribution, and demographic attributes of the applicant pool. Moreover, the school does not disclose details of the admission rubric to applicants and their families, such that they have little information on the final set of inputs into the admission scoring system, as well as the relative weight on each input. The combination of this uncertainty surrounding the placement of district-specific cutoffs in a given year and a continuous admission score provide the basis for causal identification.

2.2.2 Data

Applicant data spans the 2009 - 2012 cohorts, which correspond to high school graduation and college enrollment in 2011 - 2014. Application files include the individual’s birth month and year, ethnicity, gender, sending school, residential congressional district, county, zip code, and 10th grade SAT scores. In addition to information

provided by the applicant, the files also list the reviewers' combined score, application status (Admitted, Waitlisted, Rejected), final enrollment status, and graduation status. Importantly, the admissions office retains this data for all applicants, regardless of whether they were admitted.

The second source of data is administrative records provided by the North Carolina Education Research Data Center (NCERDC). NCERDC collects detailed information on students, teachers, schools and districts across all public and charter elementary and secondary institutions. Linked via unique and deidentified student IDs, individual-level data includes End-of-Grade (EOG) test scores from grades 3 - 8 and End-of-Course (EOC) scores corresponding to core subjects such as Algebra I and English I. Class membership rosters and grades permit construction of student grade point averages (GPA) for a particular grade, relative class ranking, and average peer achievement. School-level variables describe institutional feature such as urbanicity and student composition. NCERDC also documents family socioeconomic information such as parental education and free and reduced lunch status.

I examine three sets of outcome variables: SAT scores at the latest date of test administration, student major intentions at the end of high school, and the portfolio of student college applications. The first two sets of data come from 2011 - 2014 administrative College Board files for all seniors enrolled in North Carolina public high schools.⁸ Individualized reports document standardized scores in mathematics and critical reading at the time of the latest SAT test administration. Alongside these scores, students also fill out an accompanying questionnaire that elicits infor-

⁸ The SAT test participation rate for all college-bound seniors in North Carolina public and private high schools was 68 percent in 2012. The high participation rate may be explained in part by a preference in the state's public university system for SAT over ACT scores in the admissions process. Participation began decreasing in 2013, after the implementation of a state-wide requirement for high school juniors to take the ACT as part of the new READY accountability model. Even then, the SAT requirement as part of the STEM high school's admissions process ensures that the College Board dataset includes all applicants in the sample.

mation on individuals' first choice major.⁹ Since the overwhelming share of SAT test-takers do so during their junior or senior year, this variable captures major orientation near the end of students' high school careers. I define STEM as one of the following fields: 1) Computer and Information Sciences and Support Services, 2) Engineering, 3) Engineering Technologies/Technicians, 4) Biological and Biomedical Sciences, 5) Mathematics and Statistics, 6) Physical Sciences, or 7) Science Technologies/Technicians.

In addition to major orientation, College Board data contains codes corresponding to all colleges and universities to which students sent SAT score reports by the end of their high school careers. Students are allowed four free score reports during the registration process, which can be amended up to nine days after the test date, before SAT results are revealed to the student. Additional score reports sent after the registration period are subject to a per-report fee. The collective set closely proxies for students' college application behavior.¹⁰ A crosswalk links College Board codes to institutional attributes in the Integrated Postsecondary Education Data System (IPEDS). I measure institutional quality and selectivity using the distribution of standardized test scores for enrolled students, and STEM orientation using the share of 4-year undergraduate and graduate degrees awarded in science- and math-specific subjects.¹¹

Among 2009 - 2012 applicants, 5.6% are applying from a private, religious, or

⁹ Attributes provided on the questionnaire are made available to colleges, universities, and educational scholarship programs as part of the Student Search Service, if students choose to opt in. An incentive to participate is the possibility of receiving information about educational and financial aid opportunities from institutions.

¹⁰ Card and Krueger (2005) find a high correlation between sending SAT scores and students' actual application behavior. Score reports are especially reliable for high SAT-achievers applying to large public universities.

¹¹ STEM degrees defined as 1) Computer and Information Sciences and Support Services, 2) Engineering, 3) Engineering Technologies/Technicians, 4) Biological and Biomedical Sciences, 5) Mathematics and Statistics, and 6) Physical Sciences. Data on degrees awarded are not available for 2014, so I use institution-level data in 2013 in its place.

home school, 3.5% contain missing high school information, and 0.7% attend a public high school that does not correspond to institutional codes given by the NCERDC. I remove non-admits from private, religious and home schools because their outcomes are not observed in the administrative statewide dataset. Applicants who are missing data on their congressional district of residence and admission ratings scores are also excluded. The remaining 90% of applicants were enrolled in a recognized North Carolina public or charter school at the time of application. I link these students to NCERDC data using demographic, sending school, and residential characteristics, and weight all analyses using the inverse of the number of matches to each applicant ID.

2.2.3 Linking applicants to statewide administrative records

This section describes the process to link applicants to NCERDC records on academic outcomes and postsecondary intentions. The universe of possible NCERDC matches in a given year is trimmed significantly by one fact. All applicants must submit SAT score reports to the selective high school as part of their application. Since College Board files in the NCERDC contains a list of all institutions that a student ever submitted a score report to, I limit my NCERDC sample to 10th grade students who ever sent a score report to the selective high school. Then I use sex, birth year, and birth month as basis for the merge, as these variables are subject to less measurement error relative to attributes such as ethnicity. Supplementary information such as sending high school, enrollment status, ethnicity, and residential zip code are used to whittle down the set of possible matches. For applicants with a single match, I remove the corresponding NCERDC record from the merge set, such that the observation cannot be linked to a different applicant. The process iterates until the number of matches cannot be reduced for any applicant in the sample.

Table 2.1 shows that the distribution of matches is heavily skewed towards single

Table 2.1: Match rates

Group	2009 cohort		2010 cohort		2011 cohort		2012 cohort	
	Unique	Total	Unique	Total	Unique	Total	Unique	Total
1 match	820	820	806	806	709	709	796	796
2 matches	58	116	49	98	47	94	79	158
3 matches	12	36	33	99	23	69	15	45
4 matches	10	40	20	80	18	72	9	36
5 matches	8	40	11	55	7	35	10	50
6 matches	6	36	19	114	10	60	5	30
7 matches	8	56	0	0	2	14	5	35
8 matches	2	16	4	32	1	8	13	104
9 matches	4	36	2	18	4	36	2	18
10 matches	1	10	0	0	4	40	0	0
11 matches	0	0	0	0	0	0	6	66
12 matches	3	36	0	0	1	12	0	0
13 matches	6	78	0	0	0	0	0	0
14 matches	0	0	0	0	1	14	0	0
Total	938	1542	944	1514	827	1419	940	1562

Notes: Unique matches exclude applicants who are enrolled in a private secondary school in 10th grade and were not enrolled in the selective school.

matches, with between 85% and 88% of the applicant sample falling into this category across cohorts. 94% to 95% of the samples have three or fewer matches per applicant. The few outliers making up the remaining 5% have up to 14 matches. All subsequent analyses weight each observation by the inverse of the total number of matches. I utilize a number of robustness checks to ensure the reliability of the merge process and consistency of results, such as using only the sample of unique merges and applying bootstrap sampling and estimation methods. Finally, a small share of students did not take the SAT again after tenth grade, such that the major choice variable and final SAT scores are not measured post-treatment. When using these outcome variables I restrict the sample to only students who took the SAT in eleventh or twelve grade. The final sample amounts to 3612 weighted applicants.

2.2.4 Applicant profile

Table 2.2 documents the characteristics of the 2009 - 2012 applicant cohorts. Students in the sample skew slightly female. 12% are African American, 7% are Hispanic, and one quarter are Asian. Asians are significantly over-represented while blacks are under-represented relative to public school 10th graders across the state over the same period. 14% of all applicants qualify for free and reduced lunch, signaling an advantaged sample relative to the statewide population.¹² The school admits slightly over a third of all applicants out of the high-performing applicant pool. The average entering SAT math score is 617, 40 points higher than the average SAT verbal score. The majority of applicants place at least above 90th percentile of their 9th grade class and come from feeder schools that are above average in math performance.

As Table 2.3 shows, these numbers disguise substantial heterogeneity across congressional districts. On the academic ability dimension, students applying from CD 1 average 547 in SAT math compared to scores that are well over one standard deviation higher on average among CD 4 applicants.¹³ Another attribute that varies significantly across geographic areas is the number of applications. CD 4 has more than five times the application volume of CD 12. Admission rates span a long range across districts, with the most competitive district taking 22% of the pool while the least competitive district accepts 60%.

¹² The statewide sample of 10th grade students in feeder schools who eventually take the SAT exam comprises 24% African American, 5% Hispanic, 4% Asian, and 30% economically disadvantaged students.

¹³ In 2012, 620 corresponds to the 80th percentile in math and 580 corresponds to the 70th percentile in reading. An additional 10 points in either dimension implies a 2 percentile point increase. 10 point increments correspond to at most a 1 percentile point gain starting at the 92th percentile, or 660 for reading and 690 for math.

Table 2.2: Summary statistics - applicant sample

	Mean	P10	P25	P50	P75	P90
Binary characteristics						
Female	0.53					
Black	0.12					
Hispanic	0.07					
Asian	0.25					
White	0.52					
Economically disadvantaged	0.14					
Admitted	0.38					
Continuous characteristics						
Admission rating	90.59	70.00	82.00	92.40	101.40	108.90
Entering SAT math	617	500	560	620	680	740
Entering SAT verbal	577	460	520	580	640	690
Weighted 9th grade GPA	4.05	3.45	4.11	4.39	4.56	4.67
9th grade class rank	90.96	76.71	88.89	95.68	98.62	99.66
Quality of feeder school 10th grade	0.53	0.05	0.26	0.50	0.79	1.12
Feeder school % low SES	0.35	0.11	0.23	0.34	0.47	0.58
Feeder school % black	0.28	0.05	0.11	0.23	0.40	0.56
Feeder school % Hispanic	0.08	0.03	0.04	0.07	0.11	0.15
Feeder school % AP/honors curriculum	0.22	0.10	0.13	0.20	0.28	0.35
Share of rural households in residential ZIP	0.30	0.00	0.02	0.16	0.49	0.96

Notes: 9th grade class rank is based on weighted GPA. Feeder school 10th grade quality is calculated using the average standardized 8th grade EOG math scores of the 10th grade class. Weighted sample sizes for all binary variables and admissions rating are 3612. Sample sizes for SAT math and verbal are 3581 and 3582, respectively. Remaining sample sizes range from 3140 to 3585.

Table 2.3: Variation in applicant quality by CD

	N of applicants	Entering SAT math	Share admitted
CD 1	237	547	0.48
CD 2	260	587	0.37
CD 3	276	618	0.43
CD 4	828	659	0.22
CD 5	231	613	0.37
CD 6	242	620	0.48
CD 7	185	575	0.52
CD 8	212	590	0.43
CD 9	321	641	0.33
CD 10	243	620	0.40
CD 11	175	610	0.44
CD 12	148	610	0.60
CD 13	253	604	0.38
Total	3612	617	0.38

2.3 Empirical methodology

2.3.1 Fuzzy regression discontinuity

A challenge to estimating treatment effects of selective school attendance is the non-random selection of students. This paper relies on an admissions process that generates quasi-experimental assignment into the selective school when accounting for the underlying admissions score. In addition to the score, the committee also considers the congressional district of residence and student demographics to ensure adequate racial or socioeconomic representation among admitted students. This suggests fitting the following parametric “reduced-form” specification:

$$Y_{idt} = \pi_0 + \pi_1 D_{idt} + (1 - D_{idt})f_0(s_{idt} - s_{dt}^*) + D_{idt}f_1(s_{idt} - s_{dt}^*) + \Omega' X_{idt} + \phi_{dt} + \epsilon_{idt} \quad (2.1)$$

Here i indexes individuals, d the congressional district of residence, and t indexes cohorts. Y_{idt} is an outcome variable such as the intention to major in a STEM field. Treatment status D_{idt} assumes a value of 1 if the student’s admissions score s_{idt} surpasses the district- and cohort-specific cutoff s_{dt}^* : $D_{idt} = I(s_{idt} > s_{dt}^*)$.¹⁴ Since the school has limited allotments of residential space for men and women, I furthermore define this cutoff by gender. Under this construction, applications above the threshold are much more likely to be admitted than students below the threshold, such that there is a jump in the probability of treatment at s_{dt}^* .

I experiment with polynomials of the distance between a student’s admission score and the cutoff, $s_{idt} - s_{dt}^*$, to ensure that π_1 is robust to various functional forms. The interaction between D_{idt} and a parametric function of $s_{idt} - s_{dt}^*$ permits the slope to be different on both sides of the cutoff. X_{idt} is a vector of individual attributes, while ϕ_{dt} refers to CD- and cohort-specific intercepts.¹⁵ The choice of CD-

¹⁴ The definition of s_{dt}^* will be discussed in detail in the following section.

¹⁵ I use an expanded set of individual attributes including race, gender, socioeconomic status

and cohort-specific fixed effects addresses the possibility of district-specific changes in applicant quality over time. The composition of students applying from a given district can vary for multiple reasons, such as fluctuations in information availability on the selective high school and changes in the curriculum of sending institutions relative to the school.

The use of D_{idt} underscores the ‘fuzzy’ part of the regression discontinuity. Admissions scores and CD-cohort intercepts do not predict the probability of enrollment with certainty. Selection and sorting behavior lead to possible correlations between enrollment and the error term. Reviewers may use discretion to remove or add applicants to the offer list. Furthermore, students who were accepted may not always enroll, and those who decide not to attend the school may differ systematically from enrollees in their individual characteristics, counterfactual high schools, and family backgrounds. Given these considerations, I rely on an indicator for exceeding score cutoffs to randomize at the intent-to-treat level. Estimates quantify the effect of becoming eligible for enrollment in the selective STEM school.

The reduced form specification estimates the cumulative effect of enrolling in the selective high school but stops short of parsing the treatment effect into its components. I therefore use two stage least squares (2SLS) to examine the roles of two commonly cited channels in the literature: peer achievement and relative rank. While average higher peer quality may induce higher scores through teamwork-based classroom learning or induce teachers to tailor curriculum for a more gifted population, a “small fish in a big pond” mentality may erode students’ academic self-concept and effort, thereby hurting achievement. In addition to these key inputs, other factors that may matter for student outcomes include parental involvement, teacher quality, and curriculum content that are more difficult to measure and unlikely to be

measured by free and reduced lunch eligibility, and entering SAT math scores to increase precision of estimates.

directly observed by the econometrician. A simplified representation of these inputs is expressed by the following education production function:

$$Y_{idt} = \lambda' \tau_{idt} + \delta' \eta_{idt} + v_{idt}$$

Student outcomes Y_{idt} depend on vectors of observed and unobserved inputs captured by τ_{idt} and η_{idt} , respectively. τ_{idt} refers to the two mechanisms of interest, peer achievement and class rank. This underlying relationship between academic outcomes and the full set of inputs informs the interpretation of estimates using a 2SLS framework that only considers the peer achievement and rank channels. Such a parsimonious model is written as:

$$Y_{idt} = \alpha' \tau_{idt} + \Gamma' \mathbf{X}_{idt} + \zeta_{idt} \quad (2.2)$$

$$\tau_{idt} = \beta' z_{idt} + \Lambda' \mathbf{X}_{idt} + \xi_{idt} \quad (2.3)$$

Vector z_{idt} instruments for both peer achievement and rank. The presence of two endogenous variables means that the model cannot be identified using a single instrument of treatment status, expressed as an indicator for surpassing the admissions score cutoff. Instead I use interactions between treatment status and the applicant's cohort and district of residence as the vector of instruments z_{idt} . \mathbf{X} includes individual socio-demographic characteristics, cohort and district fixed effects, and a linear function of $s_{idt} - s_{dt}^*$.

α assumes an unbiased causal interpretation if there are no omitted η_{idt} or if these unobserved inputs are orthogonal to τ_{idt} . In the more likely case of correlated inputs, determining the direction of bias requires careful consideration of the contents of η_{idt} . It seems reasonable to assume that an academic environment enrolling higher-performing peers will also have more resources at its disposal that manifest in more experienced teachers and non-teaching staff or updated facilities enabling more re-

search opportunities in the lab. On the other hand, if parental inputs are substitutes for school quality, parents may scale back on time and financial investments when their children enroll in a selective school. This implies that 2SLS estimates of peer achievement, for example, are identifying the cumulative effect of peer quality and multiple hard-to-observe factors. Any upward bias introduced by the positive correlation between peer quality and school resources may be tempered by the negative correlation with other channels, namely parental involvement.

Under imperfect compliance, $\frac{\alpha}{\beta}$ recovers the local average treatment effect if the usual instrumental variable assumptions hold. The exclusion restriction requires that outcomes are only affected by the treatment, not the treatment assignment. As such, exceeding the cutoff cannot influence outcomes independently of increasing the probability of treatment. This assumption is likely satisfied given applicants do not know their admission scores and will not be able to act on this information in a manner that influences their future academic achievement or major intentions. Monotonicity mandates that exceeding the cutoff cannot simultaneously lead some individuals to take up the treatment while others to reject it. This effectively rules out the existence of defiers. Under these assumptions, 2SLS results are interpreted as causal for the sample of compliers, defined as applicants who enroll in the selective school when their scores exceed the threshold and do not enroll when scores reside below the threshold.¹⁶

2.3.2 *Defining admission thresholds s_{dt}^**

Eligibility status depends crucially on the district-level cutoff score. Based on discussions with school personnel on the admissions process, I assume that there is a

¹⁶ If D_z represents potential treatment status given the binary instrumental variable $Z = z$, then compliers must have $D_1 > D_0$ (or equivalently, $D_0 = 0$ and $D_1 = 1$). It is possible to nonparametrically identify statistical characteristics for compliers in models with a binary endogenous regressor, which in this case corresponds to the enrollment variable (Abadie, 2003).

pre-determined range of slots reserved for applicants from each district. This range is subject to some variation from year to year depending on the total applicant pool, with the number of seats constrained separately by gender due to limits on the available number of dorm spaces.¹⁷ The precise district- and gender-specific admission thresholds are unknown, but admission score and applicant status data permit empirically testing alternative assignment rules.

I define the cutoff value as the admission score that maximizes the share of applicants that are correctly classified as admitted if their scores exceed the threshold, or not admitted for those at or below the threshold. As such the cutoff s_{dt}^{1*} minimizes the combined probability of type I and type II errors. Figure 2.2 shows the classification of students across CDs and the most recent application cohort, with red dots denoting those who are misclassified. According to Table 2.4, this assignment rule correctly places 92% of all applicants into either the admitted or non-admitted categories.

I consider two alternative procedures and juxtapose their classification of applicants with the preferred approach. In the second procedure, the admissions committee selects a score threshold and unconditionally accepts everyone above the cutoff, while selectively admitting individuals at or below the cutoff to ensure sociodemographic diversity. As such, all applications exceeding $x = s_{dt}^{2*}$ are admitted with probability 1. If the cutoff is set at the highest non-admitted applicant's score, the procedure becomes susceptible to a large number of false negatives if this individual is an outlier with an unusually high admissions rating. To minimize this source of measurement error, I equate the cutoff to the score of the *second* highest non-admitted student. This trades off mis-assigning the highest-scoring non-admit to the list of admitted students with the threat of misclassifying relatively low-scoring

¹⁷ By legislative statute, the number of students served for each congressional district must fall within 2.5 percent of the average number of bed spaces available that year.

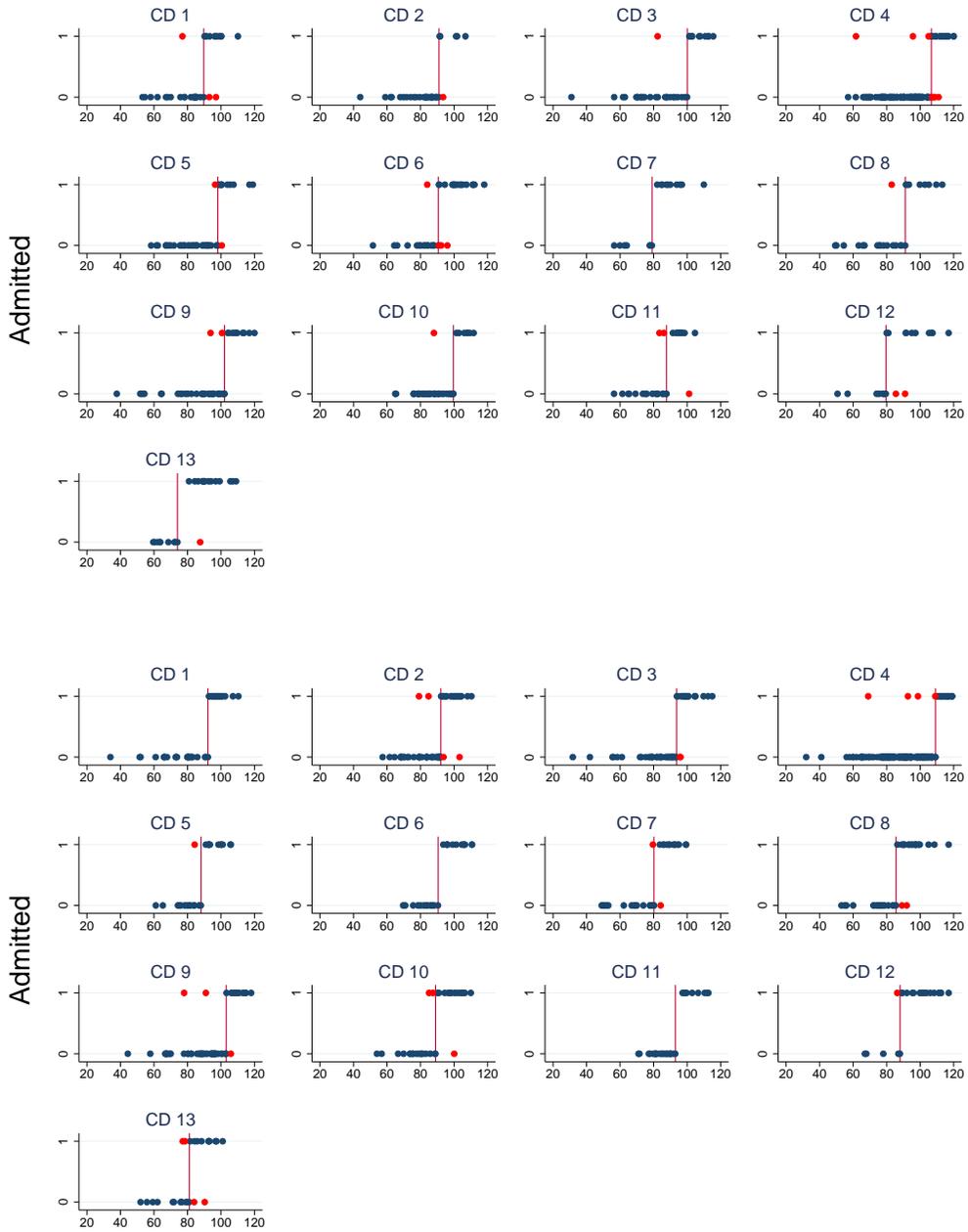


FIGURE 2.2: Admission thresholds in 2012

Notes: Males are in top panel and females are in bottom panel.

admitted students as rejected. The third assignment rule is an inverse of the second. Instead of adding applicants to an initial list of students with scores exceeding the second highest non-admitted student, the admissions committee removes applicants

Table 2.4: Cutoff definition and predictive accuracy

CD	Application year				Total
	2009	2010	2011	2012	
1	0.98	0.85	0.98	0.89	0.92
2	0.96	0.84	0.90	0.91	0.90
3	0.89	0.90	0.92	0.97	0.92
4	0.87	0.90	0.96	0.94	0.92
5	0.97	0.72	0.99	0.95	0.91
6	0.98	0.94	0.96	0.95	0.96
7	0.99	0.84	1.00	0.88	0.93
8	0.94	0.88	1.00	0.94	0.94
9	0.90	0.91	0.97	0.90	0.92
10	0.99	0.95	0.99	0.96	0.97
11	1.00	0.96	1.00	0.84	0.94
12	0.93	0.88	0.92	0.90	0.91
13	0.96	0.83	0.97	0.87	0.90
Total	0.93	0.88	0.96	0.92	0.92

from an initial list instead. The corresponding threshold s_{dt}^{3*} is the second lowest score among admitted applicants.

The second assignment procedure nearly matches the accuracy of predicted admission decisions of the original definition, yielding a correct classification in 91% of cases (Table A1). By comparison, the third procedure correctly predicts only 82% of cases (Table A2). This drop in relative performance is particularly pronounced in congressional district 4, where multiple admitted students had excessively low scores, thereby resulting in many misclassified non-admitted students. These results suggest that the admissions committees likely set a high bar for admissions and accepted those above the bar unconditionally, while making some exceptions for applicants below the cutoff to ensure a diverse student population. For the remainder of the paper, I use the optimal cutoffs derived from the first definition.

Under this definition, I pool applicants across cohorts and CDs and relate their

proximity to the admissions cutoff to enrollment in the selective high school. Figure 2.3 shows a large discontinuity in the likelihood of enrollment when an applicant exceeds the cutoff. 6.7% of applicants who miss the threshold by 2 points or less eventually enroll, compared to 87.6% of applicants who barely exceed the threshold. The fuzzy nature of the discontinuity is partially attributable to misclassified students under the optimal assignment rule. Moreover, some students who are admitted choose not to enroll.¹⁸

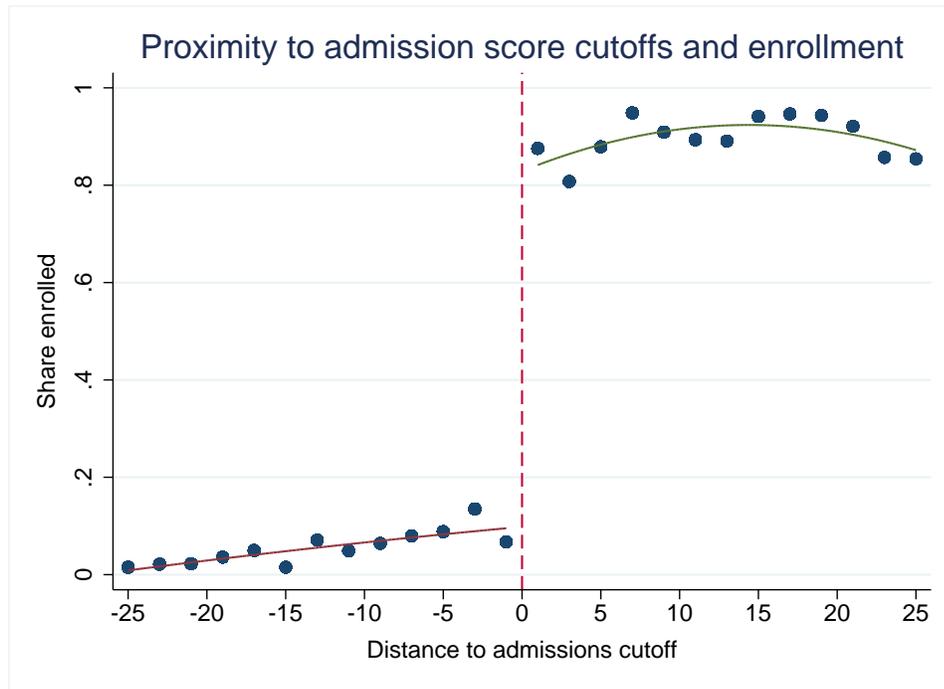


FIGURE 2.3: First stage relationship

To verify that proximity to admission score cutoffs matters for enrollment, Table 2.5 estimates enrollment for applicants and the subset of compliers using local linear regressions within the neighborhood of the cutoff. The first two columns in the upper panel describe the share of all applicants on both sides of the admissions

¹⁸ The selective school maintains an admissions waitlist and students can come off the waitlist if spots become available before Labor Day weekend during the fall semester of 11th grade. Only 4 students on the waitlist eventually enroll in my full applicant sample. Since their admission scores are at or below the cutoffs, they are classified as not treated in my analyses.

cutoff that are estimated to enroll in public or charter schools during the fall of 11th grade. The third column characterizes the enrollment behavior of compliers, defined as applicants who enroll in the selective school when scores exceed the cutoff and do not enroll when scores fail to meet the cutoff. The bottom panel examines enrollment in the selective high school. The counterfactual for 89% of all applicants missing the cutoff is a public high school, with estimated public school enrollment rates differing notably by residential CD. 9% enroll in the selective high school, representing the vast majority of remaining applicants who did not score sufficiently high. More telling are results associated with compliers who miss the cutoff. Public and charter high schools are the institutions of choice for 98% of these applicants, making these schools the de-facto counterfactual institutions in my analyses.

Predicted cutoffs enable a closer look at the population of students near the score threshold. Table 2.6 summarizes these individuals' sociodemographic characteristics and academic backgrounds for the pooled sample and by congressional district. Students within 10 points of the cutoff are less likely to be a minority and economically disadvantaged relative to the full applicant sample, and have entering SAT scores that are approximately 15 points higher. There is also marked cross-CD variation. The share of black or Hispanic marginal students is lowest in CD 4 (6%) and highest in CD 1 (40%). The average CD 4 marginal applicant has the lowest 9th grade class rank but applies from the highest quality sending school. Despite coming from lower parts of the class rank distribution, marginal students from CD 4 have mean entering SAT math scores that are almost 170 points higher than peers from CD 1. These spatial disparities in academic background are similarly pronounced when assessing sending school quality. I compute mean math test scores at the broader grade level as well as individuals' fall semester 10th grade math class. CD 4 students enroll in courses with students averaging 1.68 standard deviations above the state average in math performance, compared to 0.97 standard deviations among CD 1 applicants.

Table 2.5: Predicted enrollment for all applicants and compliers

	All applicants		Compliers
	Missed cutoff	Exceeded cutoff	Missed cutoff
Public/charter schools			
CD 1	0.97	0.15	0.97
CD 2	0.93	0.23	0.96
CD 3	0.95	0.12	0.99
CD 4	0.87	0.14	0.99
CD 5	0.92	0.11	0.98
CD 6	0.92	0.21	0.97
CD 7	0.90	0.07	0.96
CD 8	0.87	0.15	0.95
CD 9	0.85	0.13	0.97
CD 10	0.90	0.13	0.96
CD 11	0.93	0.07	0.98
CD 12	0.86	0.10	0.95
CD 13	0.88	0.19	0.95
Total	0.89	0.15	0.98
Selective high school			
CD 1	-0.00	0.91	-
CD 2	0.04	0.84	-
CD 3	0.01	0.88	-
CD 4	0.16	0.87	-
CD 5	0.08	0.92	-
CD 6	0.05	0.78	-
CD 7	0.06	0.94	-
CD 8	0.08	0.89	-
CD 9	0.13	0.92	-
CD 10	0.10	0.90	-
CD 11	0.04	0.96	-
CD 12	0.08	0.89	-
CD 13	0.10	0.84	-
Total	0.09	0.88	-

Notes: This table shows the estimated likelihoods that students who miss or exceed the admissions cutoff will return to public schools or enroll in the selective school. Enrollment in public schools is measured during the fall term of grade 11. Compliers are defined as applicants who enroll in the selective school when their scores exceed the threshold and do not enroll when scores miss the threshold. Results for compliers are estimated using the IV strategy defined in Abadie (2003) with optimal bandwidths and local linear regressions.

Table 2.6: Summary statistics for marginal applicants

	All	1	2	3	4	5	6	7	8	9	10	11	12	13
CD														
Demographics														
% Female	0.55	0.60	0.58	0.58	0.56	0.52	0.48	0.60	0.66	0.50	0.53	0.49	0.45	0.51
% Black	0.08	0.32	0.10	0.03	0.02	0.04	0.06	0.13	0.11	0.08	0.02	0.02	0.23	0.13
% Hispanic	0.06	0.08	0.10	0.08	0.04	0.04	0.03	0.10	0.06	0.05	0.08	0.04	0.11	0.07
% Asian	0.24	0.08	0.09	0.08	0.57	0.13	0.22	0.10	0.23	0.42	0.09	0.11	0.26	0.18
% White	0.57	0.49	0.66	0.76	0.32	0.78	0.66	0.62	0.54	0.40	0.76	0.77	0.38	0.58
% low SES	0.12	0.22	0.15	0.08	0.05	0.11	0.07	0.19	0.16	0.09	0.12	0.15	0.23	0.13
Academic performance														
Entering SAT math	632	550	593	618	717	620	623	568	597	675	636	621	594	624
Entering SAT verbal	592	525	566	575	661	580	578	552	574	624	595	585	556	586
Weighted 9th grade GPA	4.10	3.97	4.36	4.40	3.99	3.43	4.38	4.32	4.45	4.48	3.70	4.30	4.01	3.75
9th grade class rank	93.40	95.65	95.41	95.53	89.39	91.19	93.90	95.94	96.22	95.20	93.25	93.99	91.10	92.56
Feeder school attributes														
Quality of feeder school	0.53	0.21	0.42	0.44	0.77	0.56	0.46	0.30	0.29	0.76	0.62	0.56	0.48	0.45
Dist to feeder school mean	1.36	1.26	1.33	1.41	1.40	1.32	1.43	1.31	1.45	1.29	1.46	1.35	1.20	1.39
Quality of feeder math class	1.37	0.97	1.24	1.32	1.68	1.39	1.27	1.16	1.10	1.54	1.46	1.43	1.22	1.37
Dist to math class mean	0.52	0.51	0.51	0.54	0.49	0.49	0.60	0.44	0.62	0.51	0.60	0.51	0.47	0.45
Feeder school % low SES	0.35	0.50	0.41	0.36	0.21	0.36	0.35	0.48	0.39	0.28	0.36	0.42	0.43	0.32
Feeder school % black	0.26	0.52	0.30	0.24	0.26	0.14	0.23	0.31	0.36	0.24	0.12	0.10	0.39	0.32
Feeder school % Hispanic	0.08	0.04	0.13	0.06	0.09	0.08	0.09	0.08	0.08	0.07	0.07	0.06	0.09	0.07
Feeder school % AP/honors	0.21	0.16	0.17	0.17	0.30	0.18	0.22	0.15	0.18	0.27	0.18	0.18	0.25	0.23
% rural households in ZIP	0.31	0.57	0.47	0.38	0.11	0.55	0.29	0.41	0.22	0.10	0.33	0.49	0.09	0.33

Notes: summary statistics computed for applicants within 10 points of the admissions cutoff. 9th grade class rank is computed using weighted grade point averages. Feeder school and math class quality are the average 8th grade standardized math scores of students in the entire 10th grade and the fall semester math class, respectively. The distances to school and class means measure the proximity of applicant's own 8th grade math scores to these averages.

2.3.3 RD validity: covariate smoothness

If treatment is as good as randomized in the neighborhood of the admission score cutoff, then the distribution of observed baseline characteristics should be smooth around the threshold. Graphical representations of baseline covariates enable a direct test of the identifying assumptions implicit in a regression discontinuity framework. Figure 2.4 shows student characteristics across the admissions score. A quadratic polynomial is fitted on either side of the pooled sample. For students' entering SAT math and verbal scores at the time of application, there are no observable discontinuities at the cutoff among these scores. The same holds for two supplementary measures of cognitive ability available for the majority of applicants: standardized scores on 8th grade End-of-Grade math and reading exams.

Next I examine whether demographic attributes such as gender and race are continuous near the threshold. The share of female students is continuous, although the gender share covariate exhibits more variation across the running variable relative to earlier measures of academic performance. In contrast, there is a clear discontinuity in the share of black students on both sides of the cutoff.¹⁹ Table A3 shows that increases in the share of black applicants past the cutoff are widespread and not limited to particular districts. These patterns suggest that the majority of admissions committees are making discretionary choices to diversify the admitted student population. I address this form of selection in two ways. I include indicators for each racial category in all specifications to control for non-random selection correlated with race. I also perform robustness checks by replacing existing cutoffs with separate thresholds computed for each race in a given cohort and district. In the pooled sample the shares of black students are balanced across the race-specific cutoffs.

¹⁹ In results not shown, regressing each baseline covariate on treatment status, a cubic function of admissions scores on both sides of the cutoff, and other controls yield the same results. The coefficient on treatment status is significant only for race indicators.

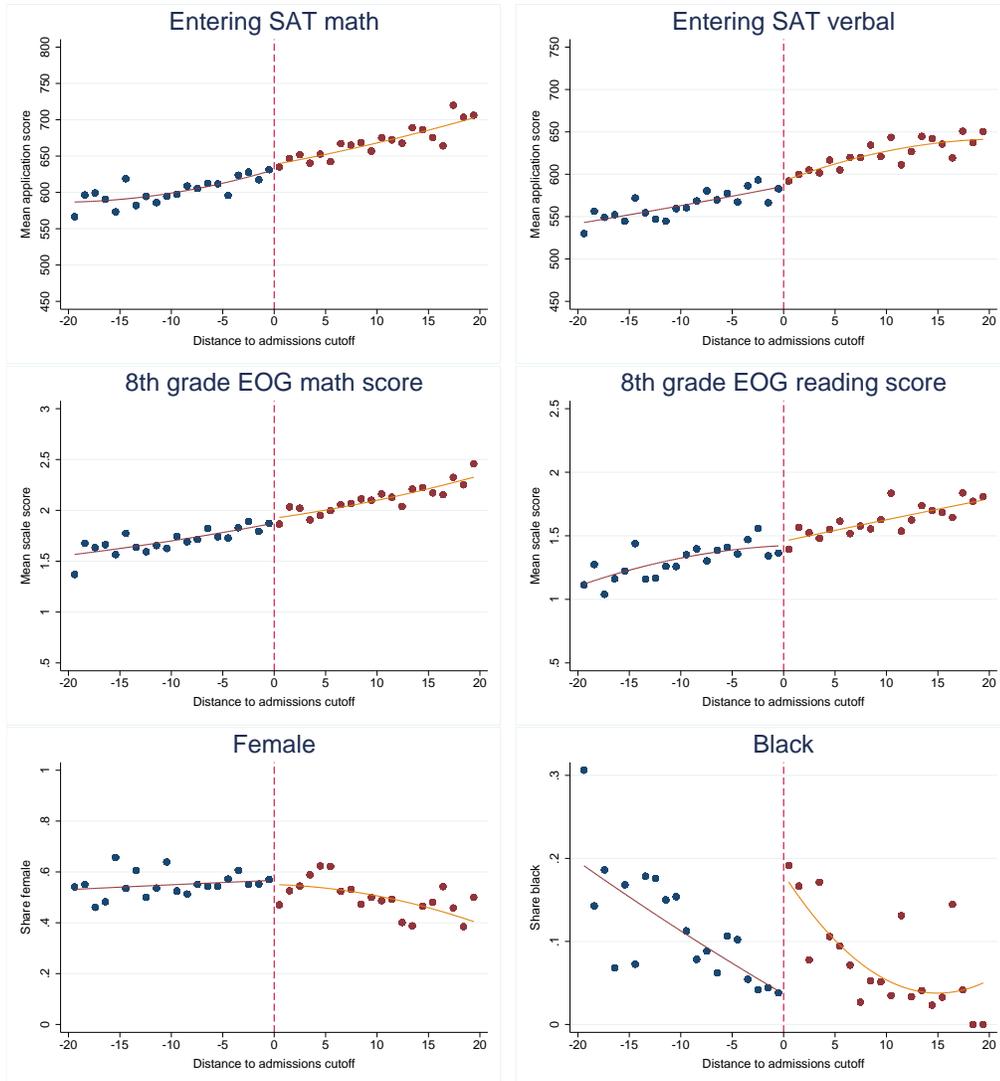


FIGURE 2.4: RD validity - covariate smoothness

2.4 Results

2.4.1 Role of selection

The first set of results juxtaposes the effect of enrollment using both the full and applicant samples. Table 2.7 compares differences in SAT math outcomes between students who did and did not enroll in the selective school after taking covariates into account. The full sample is the universe of 10th grade students enrolled in a school

with at least one applicant to the selective institution and who eventually take the SAT exam. Column 1 controls for gender, race, socioeconomic status, the average of 8th grade EOG math and Algebra I scores, high school fixed effects, and cohort fixed effects. Students enrolled in the selective school score 6.7 percentile points higher in SAT math at the end of high school. A causal interpretation of the adjusted coefficient conditional on the vector of covariates requires strong assumptions ruling out selection on unobservables. Yet there are plausible factors underlying selection into these schools that are difficult to observe. For one, attributes that prompt students to undertake a cumbersome application process affect academic trajectories in ways that are hard to account for using standard administrative variables.

Newly obtained data in the form of application variables permits evidence on the plausibility of the selection-on-observables assumption. The next two specifications restrict the sample to applicants only. Column 2 controls for a cubic function of the admissions score, a composite index consisting of students' academic performance, interest in science and math subjects, and other intangible individual attributes not typically documented in administrative data.²⁰ The coefficient on enrollment falls to 1.8 percentile points. Column 3 supplements these parametric results using nonparametric methods to ensure that functional form assumptions are not driving the key findings. It computes optimal bandwidths following the methods of Imbens and Kalyanaraman (2012) and uses a tent-shaped edge kernel centered around the admissions cutoff in local linear regressions.²¹ The results are consistent in magnitude with parametric estimates. The significant drop in the treatment coefficient from

²⁰ Alternative specifications using different functional form assumptions such as quadratic and quartic polynomials yield similar results. Thus the findings are robust to the degree of the polynomial.

²¹ The edge kernel is constructed as follows:

$$K(s_{idt}, h) = \mathbb{I}\left\{\frac{s_{idt}}{h} \leq 1\right\} \cdot \left(1 - \frac{s_{idt}}{h}\right)$$

s_{idt} represents the admissions score and h the optimal computed bandwidth.

Table 2.7: SAT math outcomes and selection bias

	Full	Parametric	Nonparametric
	(1)	Applicant	Applicant
		(2)	(3)
Enrolled	6.698*** (0.567)	1.826*** (0.548)	1.734*** (0.602)
Female	-4.348*** (0.085)	-4.419*** (0.339)	-4.224*** (0.443)
Black	-6.403*** (0.184)	-6.771*** (0.860)	-7.093*** (1.042)
Hispanic	-4.052*** (0.192)	-0.961 (0.964)	-1.055 (1.067)
Asian	1.296*** (0.386)	2.051*** (0.523)	2.017*** (0.550)
EOG8 math and Algebra I scores	Yes	Yes	Yes
Cubic function of admission scores		Yes	
Observations	134690	2866	1275
R^2	0.731	0.636	0.434

Notes: The full sample includes 1) 10th grade students in 2009-2012 that eventually take the SATs, 2) went to a school with at least one applicant to the STEM school in the same cohort, and 3) have non-missing gender, race, socioeconomic status, and EOG8 math and Algebra I scores. Specification (1) controls for gender, race, socioeconomic status, math scores, high school fixed effects, and cohort fixed effects. The applicant sample spans the 2009-2012 cohorts and includes individuals with at least one match in the public school micro data who have non-missing admission scores and congressional district information, as well as admission scores that were less than 40 points around the cutoff. Column (2) restricts the sample to only applicants. It includes gender, race, SES, math scores, CD-cohort fixed effects, and flexibly controls for a cubic polynomial of admission scores on either side of the cutoff. Column (3) controls for a linear function of admissions scores and uses the optimal bandwidths suggested by Imbens and Kalyanaraman (2012) to construct nonparametric estimates. Robust standard errors are clustered at the cohort, CD, and gender level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6.7 percentage points suggests that the adjusting solely for observed characteristics leaves an estimate that is likely upwardly biased. In the next set of results I aim for a causal estimate of selective school attendance using a fuzzy regression discontinuity approach. The treatment status variable, defined as having an admissions score in excess of the cutoff, replaces the enrollment variable to ensure that treatment is randomized at the intent-to-treat level.

2.4.2 Graphical evidence

Figure 2.5 uses the pooled sample of applicants to depict the relationship between multiple outcomes and the distance of an applicant's admissions score to the cutoff. The Y-axis shows residuals after controlling for sex, race, SES, entering SAT math scores, and CD-cohort fixed effects.²² A cubic polynomial is fitted on either side of the admissions threshold. The first three outcomes report SAT performance and major intentions at the time of the latest SAT test administration. A small jump occurs at the cutoff for SAT math percentile, although the residuals are measured with some noise. A more pronounced discontinuity near the cutoff is visible for the verbal SAT. The graph on consistent STEM intention lacks a clear discontinuity.²³ Residuals for applicants exceeding the threshold assume a U-shape, with students in the middle of the high school's ability distribution are least likely to express interest in STEM subjects at the end of high school.

College application behavior is examined along three dimensions: 1) the total number of postsecondary institutions the student applied to, 2) the selectivity of those institutions, and 3) the STEM-intensity of institution measured by the share of all Bachelor's degrees that are awarded in these fields. Visual inspection concludes that attending the high school had no measurable effect on the total number of university applications as proxied by the number of score reports. The next graph examines the effect of enrollment on applications to selective universities. Institutions are selective if the 25th percentile of their average undergraduate SAT math score is at least 500.²⁴ The share of these selective institutions appears discontinuous,

²² Each point on the scatter plot shows the average of residuals in a given bin, where residuals are computed as

$$r = Y - E(Y|X, \beta)$$

²³ Since an individual can choose up to five potential fields, the consistent STEM intention variable takes a value of 1 if the individual only indicates STEM majors in their preferred list.

²⁴ The most selective among these institutions have at least 600 or above on their 25th percentile undergraduate math SATs. In the North Carolina context, only Duke University, Davidson College,

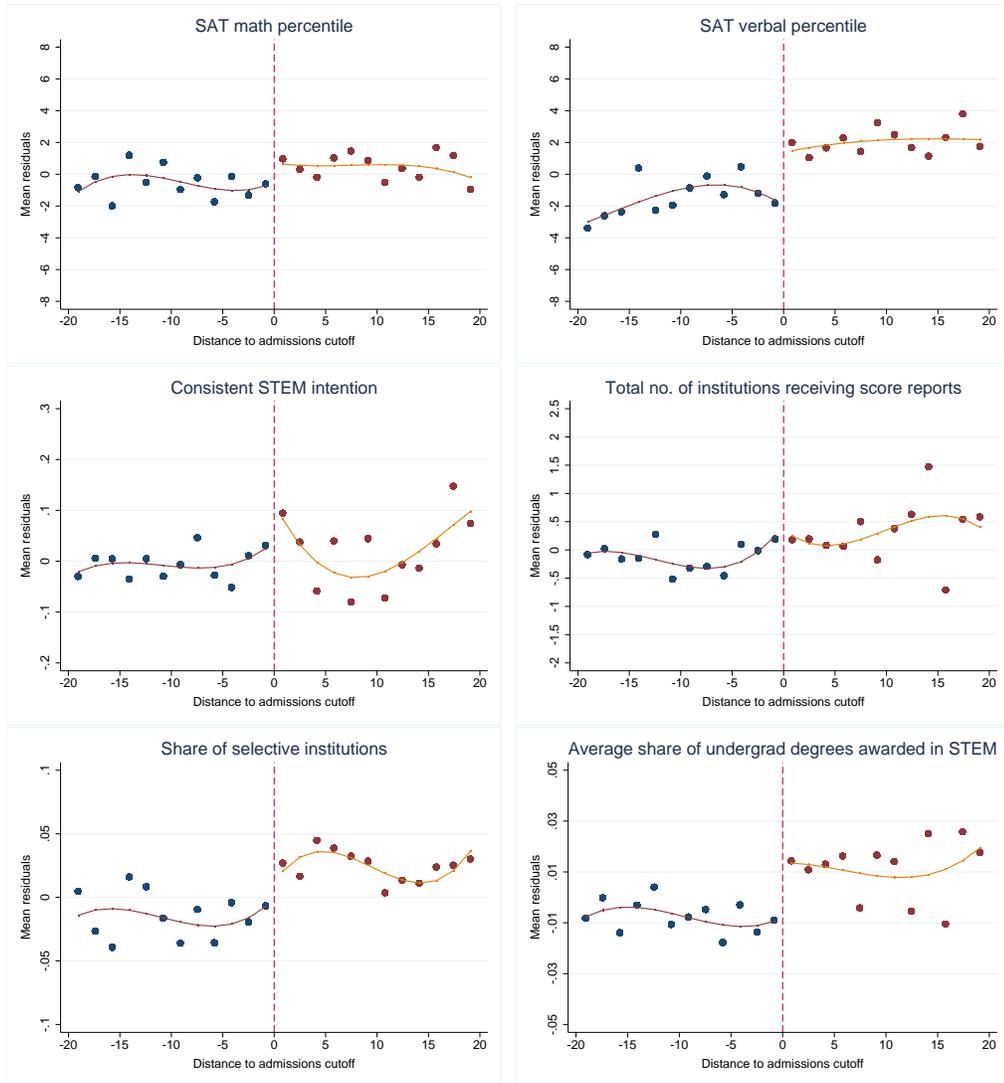


FIGURE 2.5: Outcomes - average residuals

although the difference is relatively small in magnitude. Finally, I examine whether an individual is more likely to apply to a STEM-intensive institution as measured by the share of undergraduate degrees conferred in these majors. The scatter plot strongly suggests a discontinuous increase in STEM-intensity among institutions to

and the University of North Carolina - Chapel Hill qualify under this category. Other institutions that count as selective in North Carolina include Appalachian State University, Elon University, North Carolina State University, High Point University, UNC - Asheville, UNC - Charlotte, and UNC - Wilmington.

which the students applied.

2.4.3 *Reduced form RD*

Table 2.8 shows two sets of reduced form results corresponding to graphical evidence: parametric estimates using cubic polynomials on either side of the cutoff and nonparametric estimates using optimal bandwidths and local linear regressions.²⁵ Students who exceed the admissions cutoff increase their SAT math and verbal performance by 1 to 2 percentile points, although none of the coefficients are significant. Consistent with noisiness observed in graphical evidence, the standard errors associated with STEM major preferences are large, and there is no indication that students are more likely to track into STEM fields as the result of enrolling in the selective school.

The next three outcomes summarize institutional attributes of all schools to which students ever sent SAT score reports. There is no evidence to support the view that students are changing the number of college applications. However, the share of very selective institutions increases by 3.3 percentage points in the nonparametric model. Curriculum offerings among institutions receiving score reports are also changing, as both parametric and nonparametric specifications point to an increase in the STEM intensity of receiving institutions. The effect of 1.8 - 2.4 percentage points is equivalent to an increase of 8 - 11% relative to the baseline of 22% (Table A4).²⁶

²⁵ I restrict all analyses to applicants with at least one match in the NCERDC data with non-missing admission scores and CD information. I furthermore restrict the sample to those who last took the standardized exam during junior or senior year to ensure sufficient time for the treatment to take effect.

²⁶ Interpretation of these average results depend on the extent of heterogeneity in treatment effects. If local treatment effects vary, then normalizing the cutoff to 0 and combining across cohorts and CDs results in an average estimate that is weighted by the relative density of individuals proximal to each cutoff (Cattaneo et al., 2016).

Table 2.8: Reduced form results

Dependent variable	Parametric (1)	Nonparametric (2)
SAT math percentile	0.745 (0.819)	0.866 (0.608)
SAT verbal percentile	1.638 (1.459)	1.488 (1.154)
Consistent STEM intention	0.064 (0.047)	0.021 (0.035)
Total no. of colleges receiving score reports	0.021 (0.421)	0.231 (0.278)
Share of selective colleges	0.018 (0.020)	0.033** (0.013)
Avg. share of undergrad degrees awarded in STEM	0.024*** (0.008)	0.018*** (0.005)

Notes: Results report the coefficient on treatment status, defined as an indicator for exceeding the cutoff. All models control for treatment status, a vector of individual characteristics including gender, race, SES, 10th grade SAT math, and CD-cohort fixed effects. Column (1) reports estimates from the parametric model that flexibly allows for different cubic functions of distance to the cutoff on either side of the cutoff. Applicants with admission scores more than 40 points removed from the cutoff are excluded. Column (2) reports nonparametric estimates using optimal bandwidths and a linear control of distance to the cutoff. The outcomes of SAT math and verbal percentile refer to student performance at the latest SAT administration near the end of high school. Consistent STEM intention takes a value of 1 if the individual listed only STEM majors out of five possible major choices. Total institutions refer to the number of postsecondary institutions receiving a SAT score report by the end of the individual's high school career. Average share of STEM degrees refers to average share of bachelor's degrees awarded that belong to a STEM field among all institutions receiving score reports. Robust standard errors are clustered at the cohort, CD, and gender level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.4.4 Heterogeneous treatment effects

An unique advantage of this study is the relatively heterogeneous population of students generated by the district-based admissions rule. Districts with unusually strong applicant pools, namely CD 4, have higher cutoffs than less competitive districts like CD 1. Marginal students thus vary in terms of individual attributes and academic preparation by their residential district. I investigate heterogeneous treatment effects of exposure to elite education along several dimensions: socio-demographic

background, neighborhood characteristics, sending high school attributes, baseline math ability, and residential CD. Variables corresponding to the first category include gender, race and socioeconomic status. The choice of residential ZIP urbanicity as a neighborhood variable reflects the possibility that denser areas have more high-quality alternative schools, such that enrolling in the selective school is a smaller ‘treatment’ relative to students with few schooling options. High school attributes are captured using two variables: the quality of the feeder school as measured by average student math scores and its share of honors and Advanced Placement curriculum.²⁷ I run a separate model for each category, interacting attributes such as gender with the treatment status indicator.²⁸

Table 2.9 reports treatment effects pertaining to each subgroup across all outcomes. Gains in SAT math scores are most prominent among academically and economically disadvantaged students.²⁹ Effect sizes are 6.0 percentile points for the economically disadvantaged and 3.6 percentile points for students from the bottom half of sending schools in average math achievement. Similar but more muted patterns of achievement are apparent for SAT verbal outcomes. Gains among disadvantaged groups range between 3 to 5 percentile points in SAT math and verbal,

²⁷ Low SES status is defined as ever being eligible for the national subsidized lunch program between grades 8 and 10, which corresponds to household income of 185% of the federal poverty level or less. Rural status is defined by the share of all households within a zip code located in rural areas. More urban zip codes contain 0 - 16% of rural households. I measure school quality using average 8th grade EOG math scores among 10th grade students at the sending high school. I divide the sample into the bottom and upper halves and call them low- and high-scoring schools, respectively. The lower half ranges between -0.83 and 0.58 standard deviations relative to the statewide average, while the upper half ranges between 0.58 and 2.34 standard deviations. High school curriculum offerings is measured by the share of all courses at the honors or AP level. High schools with more advanced curricula offer at least 20% of courses at honors or AP levels.

²⁸ The sample is slightly majority female, and by construction evenly divided by neighborhood urbanicity, feeder school quality and curriculum offerings. 12% and 14% of the applicant sample are black and ever eligible for free and reduced lunch, respectively. I use both SES status and the quality of the sending high school to distinguish between individual and school-level indicators of disadvantage. 9% of applicants from lower achieving high schools are also economically disadvantaged, compared to 21% of applicants from higher achieving high schools.

²⁹ See Table A4 for baseline statistics by student subgroup.

and remain robust to alternative specifications including non-parametric regressions using optimal bandwidths and the construction of CD-, cohort-, and race-specific cutoffs. Meanwhile, advantaged students from higher-performing schools or those offering more advanced curriculum record no improvements or even see a decrease in SAT performance. Attending the elite school also encourages under-represented students to apply to more selective institutions. Gains among low-income individuals or those from rural neighborhoods and lower-quality sending schools average 4 to 6 percentage points, relative to a baseline rate of 70 to 80%.

The relatively noisy measurement of STEM intentions makes it difficult to draw conclusions about the influence of selective schools on STEM major orientation. In contrast, findings on individuals' propensity to apply to STEM-oriented universities are unambiguous. Institutional STEM intensity increased across almost all subgroups, at a magnitude of 1.9 to 3.0 percentage points. Black students were the only group that did not experience a significant increase, although the possibility of a smaller treatment effect under 2 percentage points cannot be ruled out given the standard error size. In contrast to SAT and college selectivity outcomes, gains appear more concentrated among relatively advantaged students, such as those attending schools with a higher share of advanced course offerings.³⁰

³⁰ These heterogeneous treatment effects are robust to alternative specifications. Table A5 shows corresponding estimates using the nonparametric model. Under-represented and disadvantaged groups benefit in standardized test score and college selectivity outcomes, while advantaged peers are more likely to prefer STEM-oriented postsecondary institutions. Another specification check uses CD-, cohort-, and race-specific thresholds to minimize the effect of race-based selection around the original cutoffs. The shares of black students are now balanced on both sides of the threshold. Corresponding estimates documented in Table A6 are highly consistent with baseline results.

Table 2.9: Heterogeneous treatment effects

	Gender		Race		SES		ZIP urbanicity		HS quality		HS adv. curriculum	
	Female (1)	Male (2)	Black (3)	Non-black (4)	Low (5)	High (6)	Rural (7)	Urban (8)	Low (9)	High (10)	Low (11)	High (12)
SAT math percentile	2.206** (0.942)	-1.047 (0.948)	3.909** (1.572)	0.173 (0.824)	5.985*** (1.442)	-0.185 (0.797)	2.731*** (0.927)	-0.815 (0.880)	3.594*** (0.967)	-2.350*** (0.859)	3.136*** (1.002)	-1.741* (0.908)
SAT verbal percentile	1.986 (1.427)	1.211 (1.735)	3.205 (2.285)	1.354 (1.446)	4.369* (2.214)	1.153 (1.448)	3.140** (1.565)	0.923 (1.549)	4.313*** (1.620)	-1.474 (1.453)	2.794* (1.632)	0.072 (1.518)
Consistent STEM intention	0.072 (0.051)	0.055 (0.047)	0.069 (0.069)	0.064 (0.046)	0.082 (0.069)	0.061 (0.046)	0.075 (0.049)	0.057 (0.049)	0.096* (0.055)	0.044 (0.049)	0.088 (0.055)	0.050 (0.052)
Total no. of colleges	0.007 (0.431)	0.038 (0.452)	-0.266 (0.556)	0.076 (0.430)	-0.396 (0.519)	0.099 (0.423)	-0.395 (0.433)	0.286 (0.447)	-0.127 (0.421)	0.123 (0.454)	-0.334 (0.466)	0.343 (0.447)
Share of selective colleges	0.031 (0.022)	0.002 (0.021)	0.069* (0.036)	0.008 (0.021)	0.064** (0.027)	0.009 (0.021)	0.038* (0.023)	0.001 (0.020)	0.049** (0.022)	-0.021 (0.021)	0.041* (0.023)	-0.008 (0.020)
Average share of STEM BAs	0.023** (0.009)	0.025*** (0.009)	0.010 (0.012)	0.026*** (0.008)	0.023* (0.012)	0.024*** (0.008)	0.022** (0.009)	0.027*** (0.009)	0.023** (0.009)	0.021** (0.008)	0.019** (0.009)	0.030*** (0.009)

Notes: Reduced form coefficients are the interactions between the individual or HS attributes (e.g. female) and an indicator for exceeding the score cutoff. All specifications include a vector of individual characteristics including gender, race, SES, 10th grade SAT math, and CD-cohort fixed effects, as well as a cubic of distance to the cutoff on each side of the score cutoff. Applicants with admission scores more than 40 points removed from the cutoff are excluded. Robust standard errors are clustered at the cohort, CD, and gender level. * p<0.1, ** p<0.05, *** p<0.01

In addition to results by binary characteristics, I also evaluate treatment effects along the continuum of baseline math ability and by residential CD. Figure 2.6 shows point estimates and associated confidence intervals for bins of entering SAT math scores. Treatment effects are largest in magnitude among students with lower baseline ability. 3.3 and 2.1 percentile points accrue to students who began with SAT score at or below 550 and between 550 and 600, respectively. The lowest ability math students also increase their share of selective college applications by around 7 percentage points. Meanwhile, gains in the level of preferred college STEM-intensity are more uniform for students of all entering math abilities. The lowest-achieving math students gain slightly less than 2 percentage points, while the largest gains of over 3 percentage points are observed among students in the middle and top of the math ability distribution. An analysis by residential CD corroborates findings on the differential ways that disadvantaged and more privileged students benefit from selective school attendance. Table 2.10 shows that treatment effects on the science- and math-orientation of preferred colleges are concentrated among districts with higher quality feeder schools and relatively lower shares of economically disadvantaged students. Meanwhile, applicants from districts with the lowest baseline scores, CDs 1 and 7, record the largest SAT math score gains of 7 and 6 percentile points, respectively. The prevalence of positive treatment effects among disadvantaged students with lower baseline scores shows that being overmatched to a selective secondary school does not appear to negatively impact student performance.³¹ Null effects among advantaged students rather suggest that elite education can bridge existing achievement gaps.

³¹ This contributes high school evidence to a mismatch literature focused on postsecondary education. While some argue for the detrimental effects of race-based affirmative action (Sander and Taylor, 2012), others find little empirical evidence for the theory (Alon and Tienda, 2005; Bowen and Bok, 2016). To the extent there are consequences from mismatch, determining the optimal level of racial preferences requires balancing institutional diversity and the accompanying positive effects of college quality with negative matching effects (Arcidiacono and Lovenheim, 2016).

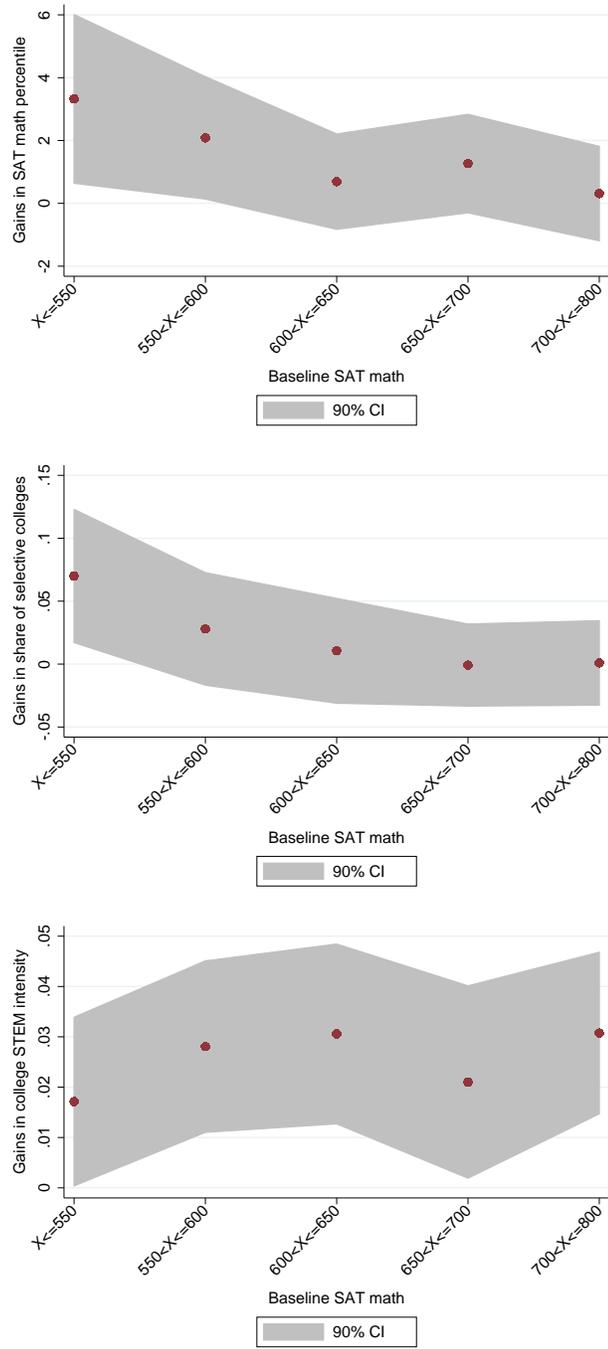


FIGURE 2.6: Heterogeneous treatment effects by entering SAT math scores

Notes: outcomes are 1) final SAT math percentile, 2) share of selective colleges, and 3) college STEM-intensity.

Table 2.10: Heterogeneous treatment effects by CD

	CD 1	CD 2	CD 3	CD 4	CD 5	CD 6	CD 7	CD 8	CD 9	CD 10	CD 11	CD 12	CD 13
SAT math percentile	6.814*** (1.409)	0.951 (1.796)	2.125 (1.742)	-4.816*** (1.112)	1.493 (1.676)	1.473 (1.434)	5.790** (2.209)	3.889** (1.857)	-2.531* (1.364)	3.155* (1.629)	3.134* (1.761)	2.981 (2.072)	-0.796 (1.972)
Share of selective colleges	0.063 (0.041)	0.020 (0.035)	0.013 (0.031)	-0.037* (0.022)	0.060* (0.036)	0.025 (0.029)	0.020 (0.027)	0.050 (0.036)	0.015 (0.025)	0.044 (0.031)	0.049 (0.036)	-0.018 (0.042)	0.058** (0.027)
Average share of STEM BAs	0.004 (0.009)	0.002 (0.011)	0.032*** (0.012)	0.028*** (0.011)	0.031** (0.014)	0.032** (0.015)	0.022 (0.017)	0.020* (0.011)	0.037** (0.014)	0.039*** (0.015)	0.007 (0.016)	-0.007 (0.016)	0.032** (0.013)

Notes: Reduced form coefficients are the interactions between the CD and an indicator for exceeding the score cutoff. Robust standard errors are clustered at the cohort, CD, and gender level. * p<0.1, ** p<0.05, *** p<0.01

The diversity of treatment effects prompts the question of why some benefit more than others. Table 2.6 provides evidence that the ‘treatment’ is larger for the most disadvantaged students. Applicants from the three CDs documenting the largest SAT math gains come from feeder schools and math classes with the lowest average math achievement relative to other districts. These same CDs of 1, 7, and 8 also have among the highest shares of low income and minority peers in their sending schools, in stark contrast to the over-representation of Asian and high SES students in the selective school. Table 2.11 offers some insight into specific channels by documenting the counterfactual 11th grade academic environments of compliers. For one, the experience of enrolling in the selective school represents a larger jump in peer quality for more disadvantaged students. Under-represented applicants predicted to not receive an admissions offer enroll in 11th grade classrooms that average 0.38 standard deviations above the state mean in math scores, compared to 0.56 standard deviations for their more privileged peers. Another dimension of the academic environment that alters is the student’s *relative* academic standing, computed as percentile rank among 11th grade classmates using standardized 8th grade math scores.³² Black students predicted to receive an admissions offer end up at the 29th percentile of the math ability distribution in the selective school, compared to the 81st percentile in their alternative schools. This drop of 52 percentile points is significantly larger than the corresponding 38 percentile points among non-black students. It is apparent from Table 2.11 that exposure to the selective school environment poses challenges and opportunities that vary by student background. The next section begins to

³² Let R_{ist} denote percentile rank computed from high school peers’ 8th grade EOG math scores, where i indexes individuals, s the high school, and t the cohort:

$$R_{ist} = 100 \times \left(\frac{n_{ist} - 1}{N_{st} - 1} \right)$$

Individual ordinal rank n_{ist} ranges from 1 at the bottom of the class to grade size N_{st} corresponding to the highest achiever. As such, the lowest ranking student has $R_{ist} = 0$ while the highest has $R_{ist} = 100$.

unbundle the importance of different mechanisms using variations in the academic context generated by cohort- and district-specific cutoffs.

Table 2.11: 11th grade characteristics of compliers

	11th grade school quality		11th grade class rank	
	Missed cutoff	Exceeded cutoff	Missed cutoff	Exceeded cutoff
Gender				
Female	0.53	2.00	89	47
Male	0.56	1.98	90	52
Race				
Black	0.21	2.00	81	29
Non-black	0.57	1.98	90	52
Socioeconomic status				
Low SES	0.38	1.98	85	45
High SES	0.56	1.99	90	50
Urbanicity of residential ZIP				
Rural	0.42	1.99	90	47
Urban	0.68	1.99	89	53
Feeder school quality				
Low quality	0.25	1.95	91	42
High quality	0.85	2.03	88	57
Feeder school curriculum				
Less advanced	0.36	1.97	91	45
More advanced	0.75	2.03	88	55
CD				
CD 1	0.18	1.88	85	30
CD 2	0.47	1.91	87	44
CD 3	0.48	2.04	90	49
CD 4	0.82	2.01	89	67
CD 5	0.51	2.01	89	52
CD 6	0.40	2.00	91	50
CD 7	0.29	2.03	89	35
CD 8	0.30	1.98	90	50
CD 9	0.75	1.95	88	58
CD 10	0.69	2.04	89	64
CD 11	0.53	1.98	89	52
CD 12	0.53	2.03	89	28
CD 13	0.35	2.01	90	49

Notes: This table shows estimated school quality and class rank for compliers in the fall of 11th grade, as measured by the average 8th grade math scores of all students in the grade. Compliers are defined as applicants who enroll in the selective school when their scores exceed the threshold and do not enroll when scores miss the threshold. Results are estimated using the IV strategy defined in Abadie (2003) with optimal bandwidths and local linear regressions.

2.4.5 Mechanisms

I focus on the contributions of both peer quality and relative rank to student outcomes.³³ Figure 2.7 plots the average math scores of peers during the fall of junior year against distance to the cutoff. Applicants to the left of the threshold predominantly have peers from their sending school, while those with scores at or exceeding the threshold are significantly more likely to have peers from the selective high school. In the pooled sample, applicants who barely miss the cutoff have peers that score above 0.5 standard deviations above the state average, compared to over 1.6 standard deviations among 11th grade peers at the selective high school. The second graph focuses on math performance relative to same-grade peers.³⁴ Students who exceed the predicted score cutoffs are above the 40th percentile rank while those who barely missed the threshold fall near the 70th percentile rank, reflecting the reality of enrolling in a more competitive institution in which everyone excels academically.

Table 2.4.5 presents non-parametric 2SLS estimates from instrumenting for the two endogenous variables of peer achievement and percentile rank in Equation 2.2. The first two columns instrument separately for each dimension using treatment sta-

³³ Other possible mechanisms include tailored instruction and parental inputs. Table 2.6 shows that the distances between applicants' own math abilities and their 10th grade math class mean vary from 0.44 to 0.62 standard deviations, although the different magnitudes do not have any strong relationships with student outcomes. Within-school tracking and classroom membership data are not available for the selective school, thereby preventing the estimation of 11th grade statistics. Data are similarly lacking for measures of parental inputs. Understanding how parental involvement differs by family background and the effect this has on returns to school quality can shed light on muted treatment effects among well-off students. If the parents of more advantaged students are better able to assist with advanced coursework and invest more temporal or financial resources into their children, moving to a boarding school likely induces a greater drop in parental involvement among this group.

³⁴ I define peer group as students in the same school and grade. I rely on empirical evidence in support of linear rank effects to make the assumption that students across the math ability distribution are evenly affected (Murphy and Weinhardt, 2014). Since schools do not communicate math rank, this implies that students can infer relative position through math course choice and performance. I am in the process of obtaining math class membership data at the selective school and will check for robustness under revised peer ability and rank variables constructed at the classroom level.

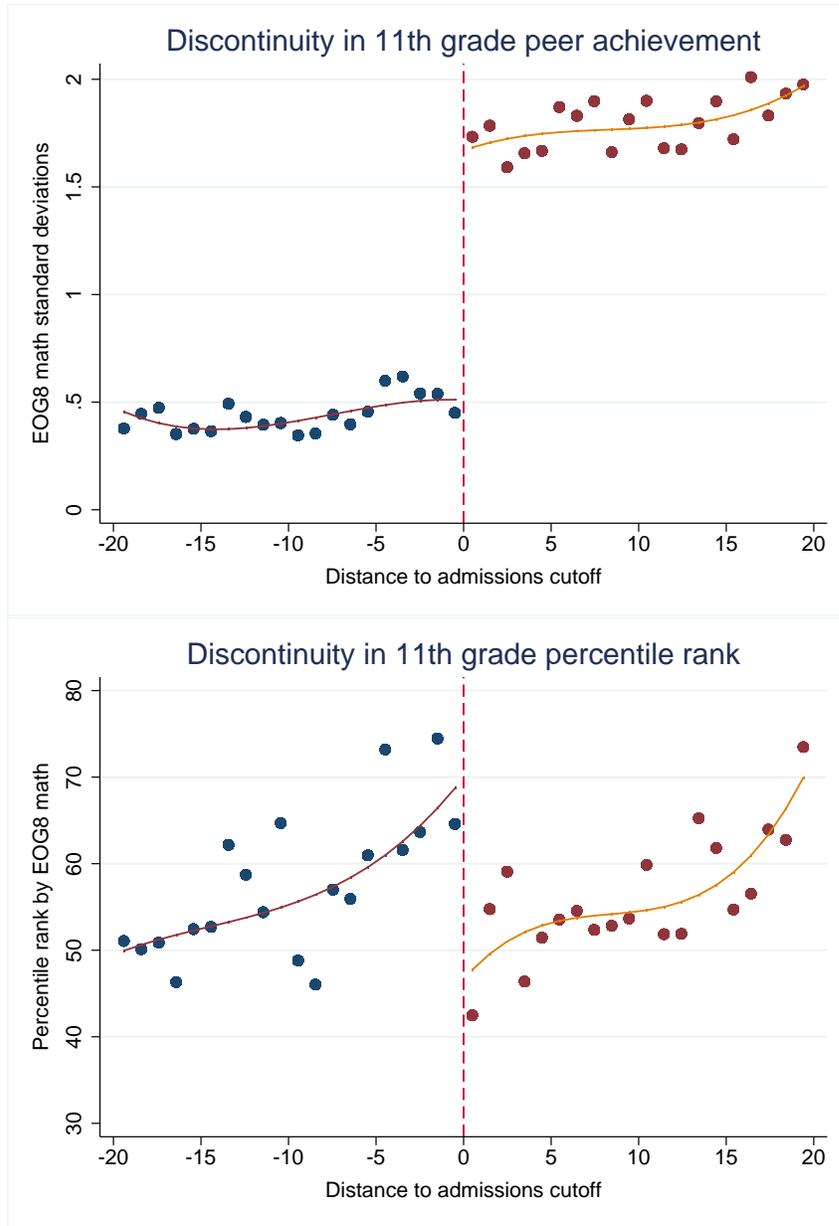


FIGURE 2.7: Discontinuities in peer achievement and percentile rank

tus expressed as a single indicator for exceeding the cutoff. Exceeding the admission score cutoff induces an increase in peer quality of over one standard deviation, and a drop in percentile rank of approximately 35 percentile points. The magnitudes and precision of these estimates are consistent with graphical evidence on the sharp

discontinuities in peer composition around the cutoff. The third column interacts the treatment status with application cohort and district indicators to identify the contributions of both endogenous variables.³⁵ I report associated F-statistics for the just- and over-identified models under each outcome, along with first stage coefficients on treatment status. The strong first-stage relationship between treatment status and the two mechanisms is apparent from three-digit F-statistics across all outcomes. The use of multiple instruments in column (3) significantly deflates the F-statistics but they remain well into the double-digits.

An increase in 11th grade peer achievement yields similarly sized and statistically insignificant improvements in SAT scores as reduced form results. A jump in peer quality leads students to apply to institutions that graduate 2.8 percentage points more STEM degrees. The coefficients on percentile rank maintain the opposite sign of results on peer achievement, because surpassing the admissions cutoff increases the probability of being lower ranked in a higher performing school.

³⁵ Identification derives from multiple moments of peer achievement. Applicants from two CDs that experience the same increase in peer quality from enrolling in the selective school can undergo different rank changes because they do not originate from the same relative position. Consider two individuals that both score 1.5 standard deviations above the statewide average in classroom $i = \{1,1,1.5\}$ and classroom $j = \{0,1.5,2\}$, respectively. Average peer quality is the same across both classrooms, but the applicant in the former is ranked at the 100th percentile, while the latter is at the 50th percentile. This example underscores the role of variance in isolating the rank effect, as more compressed score distributions lead to larger changes in relative rank. An alternative example with classrooms $i = \{0,1.5,2\}$ and $j = \{1,1.5,2\}$ illustrates a situation in which the change in rank remains constant while there is a greater increase in peer quality in classroom i .

	(1)	(2)	(3)
SAT math percentile			
Average 11th grade peer EOG8 math score	1.012* (0.537)		0.192 (1.997)
11th grade percentile rank in EOG8 math		-0.024 (0.018)	-0.050 (0.059)
F-statistic	369.867	178.943	124.573 51.574
Exceeded cutoff	1.046*** (0.054)	-34.978*** (2.615)	
Observations	3,254	2,857	2,857
SAT verbal percentile			
Average 11th grade peer EOG8 math score	1.495 (1.112)		1.485 (3.514)
11th grade percentile rank in EOG8 math		-0.065* (0.033)	-0.034 (0.098)
F-statistic	411.464	204.977	149.055 29.868
Exceeded cutoff	1.058*** (0.052)	-35.514*** (2.481)	
Observations	3,664	3,216	3,216
Consistent STEM intention			
Average 11th grade peer EOG8 math score	0.013 (0.035)		0.008 (0.097)
11th grade percentile rank in EOG8 math		-0.000 (0.001)	0.000 (0.003)
F-statistic	353.722	170.480	91.605 56.905
Exceeded cutoff	1.046*** (0.056)	-34.882*** (2.672)	
Observations	3,048	2,684	2,684

	(1)	(2)	(3)
Total no. of colleges receiving score reports			
Average 11th grade peer EOG8 math score	0.228 (0.267)		0.575 (0.838)
11th grade percentile rank in EOG8 math		-0.008 (0.008)	0.015 (0.024)
F-statistic	424.077	220.398	88.972 26.304
Exceeded cutoff	1.068*** (0.052)	-36.115*** (2.433)	
Observations	3,818	3,401	3,401
Share of selective institutions			
Average 11th grade peer EOG8 math score	0.028** (0.013)		0.097** (0.047)
11th grade percentile rank in EOG8 math		-0.001** (0.000)	0.002 (0.001)
F-statistic	429.070	212.045	38.997 17.541
Exceeded cutoff	1.061*** (0.051)	-35.509*** (2.439)	
Observations	3,651	3,244	3,244
Share of undergrad degrees in STEM			
Average 11th grade peer EOG8 math score	0.015*** (0.005)		0.068*** (0.023)
11th grade percentile rank in EOG8 math		-0.000*** (0.000)	0.002*** (0.001)
F-statistic	494.045	255.222	148.491 70.118
Exceeded cutoff	1.091*** (0.049)	-37.063*** (2.320)	
Observations	4,153	3,691	3,691

Notes: 2SLS estimates of Equation 2.2 describe the effect of average peer achievement and percentile rank on outcomes. Columns (1) and (2) use an indicator for exceeding the cutoff as instrument. Column (3) interacts the indicator with cohorts and CDs. * p<0.1, ** p<0.05, *** p<0.01

The last specification permits the possibility of multiple mechanisms acting in tandem to influence student achievement and application behavior.³⁶ The direction of signs on the percentile rank variable underscores the importance of considering both mechanisms. Whereas lower rank in the second specification implies that students are more likely to be enrolled in the selective school, the explicit control for peer achievement in the third model largely reverses the sign of rank coefficients. I find that peer achievement is the predominant mechanism for college selectivity decisions, with an increase in peer quality leading to a 9.7 percentage point increase in the share of selective institutions in students' application portfolios. Meanwhile, the outcome of college STEM-intensity is increasing in both mechanisms. An improvement in peer achievement yields a 6.8 percentage point increase while higher percentile rank leads to a 0.2 percentage point increase.

2.5 Role of admissions policies

Findings on the heterogeneous consequences of attending selective schools underscore the importance of student composition for evaluating treatment effects. Since success depends in large part on who enrolls, the decision of who to admit is an important one that shapes institutions' ability to realize their goals, ranging from improving student test scores to broadening postsecondary access.

Elite schools utilize a range of admissions criteria. Some institutions rely on a single exam score to determine placement. Many postsecondary institutions outside of the US and certain US-based secondary institutions such as New York City's selective high schools fall into this category. The composite score adopted by the North Carolina selective school belongs to the other end of the spectrum, as it eschews a targeted focus on test scores to consider academic and non-academic criteria includ-

³⁶ The use of multiple instruments to estimate each channel's contribution leads to inflated standard errors that broaden the range of effect sizes that are too small to be detected.

ing sociodemographic and geographic attributes. Explicit diversity considerations call to mind the range of affirmative action policies prevalent in the secondary and higher education sectors. Juxtaposed against a rule focused solely on test scores, both race-based and race-neutral affirmative action approaches seek to increase student body representation among historically under-represented groups. Empirical evidence shows varying degrees of success in reaching this goal. To illustrate the effects of diversity considerations, I simulate student composition and estimate counterfactual treatment effects under alternative admissions policies.

A shift in admissions policy has immediate consequences on two fronts. The first is that the admitted student mix changes, conditional on the same applicant pool. For instance, changing the weight of test scores relative to non-academic factors alters the likelihood that high-scoring individuals will be admitted. Shifts in admissions policy also change the incentives to apply. Higher expected probabilities of admissions in effect reduces the cost of applying for favored groups, thereby increasing their representation in the applicant pool. Empirical literature on higher education generally find that the first mechanism dominates.³⁷ Since the shift is largely driven by institutions' choice of who to admit rather than changes in application behavior, I focus on the first effect and abstract away from the impact of admissions policies on the applicant pool. This translates to simulating the effect of changing admissions rules on student composition and treatment effects while keeping the applicant pool constant.

I juxtapose admitted student compositions under three admissions policies: status quo, math score-based, and class rank-based. The existing policy considers a range of academic factors and personal attributes including district of residence and the

³⁷ Studies evaluate the adoption of race-based affirmative action and eventual replacement of these policies using alternative diversity considerations in many public universities. There is a preponderance of evidence for lower minority representation from affirmative action bans, with the bulk of the effect deriving from changes in who is admitted instead of large shifts in applicant composition (Arcidiacono, 2005).

quality of sending schools. The second regime admits at least 325 students in each cohort with the highest entering 10th grade SAT math scores. This contrasts with the third policy of admitting on class rank, calculated using 9th grade weighted grade point averages and standardized on a 0 to 100 scale. The reliance on rank as a sole input recalls statewide percent rules, such as guaranteeing all high school seniors graduating in the top 8% of their class a spot in an University of Texas campus. Implicit in these types of admissions policies is the assumption that equalizing access across high schools recovers the drop in minority student representation following the ban of race-based affirmative action (Long and Tienda, 2008).

The first column of Table 2.13 shows average socio-demographic characteristics, academic performance, and feeder school attributes of all observed applicants from 2009 - 2012. The next three columns report corresponding statistics for applicants predicted to receive an offer under the 3 admissions rules. As expected, each rule yields admitted student populations with the highest average in the index it prioritizes, namely admissions scores, math scores, and class rank. Both the status quo and class rank-based policies achieve greater racial, socioeconomic, and geographic diversity relative to the math score-based rule. 8% of admitted students under the math-based rule are black or Hispanic, compared to 18% and 15% under the current and rank-based policies. The rank-based rule leads to the highest proportion of economically disadvantaged students of 12%, relative to 11% under the status quo and 6% under the rank-based rule.³⁸ Moreover, the math-based rule skews the gender ratio among admitted students towards males. This is consistent with previous findings that conditional on prior achievement, females tend to under-perform in high stakes tests (Corcoran and Baker-Smith, 2016). Spatial diversity is best achieved using district-specific admissions quotas (Table 2.14). CD 4, one of the most afflu-

³⁸ The rank-based rule tend to perform well on diversity measures if feeder schools and neighborhoods are more segregated (Tienda and Niu, 2006).

Table 2.13: Admissions rules and summary statistics

	All Applicants	Admitted		
		Status quo	Math	Rank
Socio-demographics				
Female	0.53	0.50	0.42	0.54
White	0.52	0.54	0.48	0.59
Black	0.12	0.09	0.02	0.08
Hispanic	0.07	0.09	0.06	0.07
Asian	0.25	0.24	0.41	0.22
American Indian	0.01	0.01	0.00	0.01
Other	0.04	0.04	0.04	0.04
Economically disadvantaged	0.14	0.11	0.06	0.12
Academic performance				
Adm rating	90.59	103.05	100.89	97.48
9th grade class rank	90.96	95.10	91.86	98.94
SAT math	616.91	665.62	708.42	637.75
Feeder school attributes				
Quality of feeder school	0.53	0.53	0.72	0.44
% AP/honors courses	0.22	0.22	0.26	0.20
% black	0.28	0.28	0.23	0.29
% hispanic	0.08	0.08	0.07	0.08
% free-reduced lunch	0.35	0.37	0.28	0.39
Observations	3612	1309	1365	1304

Notes: Status quo sample defined as observations with non-missing admission scores, and CD information. Of the 3612 weighted observations, 3581, 3372, 3414, and 3439 have non-missing 10th grade SAT math scores, 9th grade class rank, feeder school quality, and feeder school share of AP/honors courses, respectively. The math-based policy admits no fewer than 325 students per year on the basis of 10th grade SAT math scores. The rank-based admissions policy takes no fewer than 325 students per year on the basis of only class rank as computed from 9th grade weighted GPA.

ent districts, supplies 10% of all students eligible to be admitted under the current policy. This rises to 17% using the rank-based approach and more than triples to 36% under the math score-based regime. More economically deprived areas such as CD 1 are increasingly absent from the admitted pool as the admission criteria shift to measures of academic preparation. Thus available evidence show that a heavy reliance on standardized tests in the admissions process yields a more academically advantaged admitted student sample with lower shares of under-represented student

groups.

Table 2.14: Admissions rules and summary statistics by CD

CD	All applicants		Admitted - current		Admitted - math		Admitted - rank	
	N	Column Pct	N	Column Pct	N	Column Pct	N	Column Pct
1	336	6.56	132	8.88	31	1.81	122	7.14
2	369	7.20	116	7.78	70	4.09	155	9.09
3	392	7.64	142	9.52	122	7.12	177	10.38
4	1174	22.93	155	10.42	611	35.54	295	17.32
5	328	6.40	91	6.07	91	5.30	90	5.30
6	343	6.70	138	9.24	122	7.09	113	6.66
7	263	5.14	108	7.25	52	3.00	105	6.18
8	301	5.87	102	6.87	73	4.25	117	6.86
9	455	8.89	108	7.25	193	11.23	144	8.43
10	344	6.72	101	6.80	118	6.85	120	7.07
11	248	4.85	86	5.78	69	4.03	92	5.39
12	210	4.10	107	7.18	59	3.44	76	4.45
13	359	7.01	104	6.95	107	6.23	98	5.75

What would be effect of attending the selective school if students were admitted under these alternative policies? Under certain assumptions, it is possible to re-weight the pooling of local effects across cohorts and districts to derive counterfactual treatment effects (Bertanha, 2015). I assume that treatment effects vary smoothly across the forcing variable and student characteristics such as baseline math ability. I then average over the distribution of these characteristics for students to estimate treatment effects under the new rule. This approach does not address general equilibrium consequences. Put differently, I assume that an alternative admissions policy changes the admitted student mix, but not the treatment effect associated with a given characteristic. In doing so I am abstracting away from student and institutional behavioral responses to the shifting student body composition.³⁹ Specifically, I use

³⁹ Students may respond to different peers by socializing or studying with new groups, and schools may react by adapting curricula or sorting students differently into classrooms. Such endogenous patterns may lead to unintended consequences, such as the lowest ability students clustering in friendship and study groups within a school and fewer cross-group interactions than would be intended (Arcidiacono, Khan, and Vigdor, 2011; Carrell, Sacerdote, and West, 2013). A structural model that considers peer effects and other school inputs, combined with more selective school data

local linear regressions and optimal bandwidths and allow treatment effects to vary by baseline math ability, race, socioeconomic status, and distance to the admissions cutoff:

$$\begin{aligned}
 Y_{idt} = & \pi_0 + \pi_1 D_{idt} + \gamma_1 D_{idt} \times \text{Math} + \gamma_2 D_{idt} \times \text{Race} + \gamma_3 D_{idt} \times \text{SES} \\
 & + \beta_1 (1 - D_{idt})(s_{idt} - s_{dt}^*) + \beta_2 D_{idt}(s_{idt} - s_{dt}^*) + \Omega' X_{idt} + \phi_{dt} + \epsilon_{idt}
 \end{aligned}
 \tag{2.4}$$

Effect heterogeneity is operationalized by interactions with the treatment status variable ($D_{idt} = I(s_{idt} > s_{dt}^*)$). Table 2.15 summarizes marginal treatment effects for two groups of students: those near the threshold of treatment and those predicted to be admitted because their scores exceed the cutoff value.⁴⁰ I report marginal treatment effects across three student achievement and postsecondary outcomes: end-of-high-school SAT math percentile, the share of selective college applications, and the average STEM-intensity of students' preferred colleges.

Table 2.15: Marginal treatment effects across 3 admission rules

	Distance to adm. score cutoff (S)		Distance to SAT math cutoff (M)		Distance to rank cutoff (R)	
	Marginal students	Admitted	Marginal students	Admitted	Marginal students	Admitted
SAT math percentile	0.993 (0.611)	0.946 (0.877)	0.687 (0.598)	-1.345* (0.716)	1.193* (0.611)	0.987 (0.636)
Share of selective colleges	0.031** (0.013)	0.027** (0.013)	0.028** (0.013)	0.011 (0.012)	0.033*** (0.013)	0.032** (0.012)
College STEM-intensity	0.018*** (0.005)	0.020*** (0.006)	0.020*** (0.006)	0.025*** (0.006)	0.018*** (0.005)	0.019*** (0.005)

Notes: All results estimate Equation 2.4 using the sample of 2009 - 2012 applicants. Marginal students are defined as $|S| \leq 3$, $|M| \leq 20$, and $|R| \leq 1$, while admitted students are defined as having scores exceeding cutoffs. All specifications include treatment status, interactions between treatment status and 10th grade SAT math, race, SES, and distance to the cutoff, and a vector of individual characteristics including race, SES, 10th grade SAT math, gender, and CD-cohort fixed effects. Marginal effects are estimated using local linear regressions, optimal bandwidths and edge kernels. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Treatment effects for marginal students under the current admissions policy are similar in magnitude to non-parametric results presented in Table 2.8. When comparing across the alternative policies, marginal students under the rank-based rule

on variables such as classroom membership, can provide more nuanced estimates of counterfactual treatment effects.

⁴⁰ To determine the sample of marginal students, I select a bandwidth under each rule that provides a similarly sized group of students near the margin of admissions. These correspond to 3 points for the 0 - 120 admission score index, 20 points for the 200 - 800 SAT point scale, and 1 percentile point for the 0 - 100 class rank scale. Treatment effects are averaged over the distribution of students near the margin of admissions. For the admitted student sample, treatment effects are averaged over the distribution of students who are predicted to be admitted under a given rule.

observe a significant increase of 1.2 percentile points, while the remaining policies cannot rule out a smaller increase of less than 1 percentile point. Among admitted students who are higher-achieving on average, there is a 1.3 percentile point drop in SAT math performance under the math-based rule. This reflects the earlier finding that treatment effects are decreasing in baseline math ability as suggested by Figure 2.6. Coefficients for admitted students under the two other admissions rules are positive but insignificant. A more diverse student sample also yields higher shares of selective college applications. Marginal students under the status quo and rank-based rules increase their application share by 3.1 and 3.3 percentage points, respectively. Students near the neighborhood of the math score-based rule experience a smaller treatment effect of 2.8 percentage points.

While larger gains in SAT math and college selectivity outcomes accrue to more socio- demographically diverse students, a somewhat different pattern emerges when evaluating college STEM outcomes. Among both marginal and admitted student samples, the math score-based admissions policy produces effects of the largest magnitude. This is consistent with larger effects found among students near the middle or the top of the ability distribution, which is a group disproportionately admitted under the math-based rule. The evidence suggests that high achieving math students are more influenced by the selective school’s academic environment to choose postsecondary institutions that are similarly STEM-oriented. Their relatively higher levels of baseline preparation may enable them to better take advantage of advanced curriculum and research opportunities in STEM, thereby promoting a science- and math-focused track in postsecondary education.

2.6 Conclusion

This paper uses discontinuous probabilities generated by an unique admissions policy to estimate the impact of selective secondary schools on students’ test scores,

major orientations, and postsecondary intentions. I rely on the district-based quasi-experimental setting to evaluate how a diverse group of marginal students fare from exposure to selective education. This contributes to an exam schools literature that is not well positioned to examine effect heterogeneity, owing to the fact that regression discontinuity designs are mostly limited to one cutoff per school. The exception of Lucas and Mbiti (2014) evaluates selective schools with multiple cutoffs per institution in the Kenyan context, and finds no differential effects by baseline ability, gender, or socioeconomic status. In contrast, this paper provides substantial US-based evidence that high achieving students from diverse backgrounds benefit in meaningful but different ways from attending a selective secondary school.

The findings shed light on the null or small aggregate treatment effects prevalent in the literature. The treatment effect of attending a selective school is muted among advantaged students who can exercise the option of returning to a better school. This suggests that previous null findings are consistent with estimating treatment effects on a relatively privileged student sample. Meanwhile, disadvantaged students experience a larger increase in peer quality and other hard-to-observe but positively correlated measures of school quality. The advantages of attending a selective school are tempered by channels such as a large drop in relative rank, which are shown to lead to fewer applications to STEM-focused colleges.

Patterns of differential benefits underscore the importance of the admissions process to evaluations of school effectiveness. I show that admissions rules considering sociodemographic and geographic diversity yield bigger test score gains and a higher share of selective school applications, while a rule that admits on baseline math scores alone results in null or even negative effects among marginal and admitted students. This implies an efficiency-based argument for inclusive selective schools. Admitting a diverse pool of students is compatible with the goals of maximizing score gains and encouraging access to selective postsecondary institutions. The findings

also deny any equity-efficiency tradeoffs, as greater gains need not come at the cost of fewer under-represented students and widening achievement gaps. A more nuanced picture emerges for other institutional goals. For example, an accompanying objective might be to encourage graduates to enroll in universities heavily invested in engineering and related STEM programs. While under-represented students are applying to more STEM-intensive institutions, the magnitudes of gains are slightly larger among students with higher baseline math scores. Therefore the optimal design of admissions policies requires weighing between the fulfillment of different institutional goals.

Overall, the evidence suggests that expanding elite education to serve more disadvantaged groups has the potential to significantly benefit these students' test scores and selective college aspirations. The best way to operationalize such an expansion — through new selective schools or increased enrollment at existing schools — remains open to debate. Replicating the existing model by opening similar schools is one option supported by this study's findings. An increase in the proportion of disadvantaged students at existing institutions may also be effective, assuming that the shift in composition does not induce student behavioral responses that can negatively impact school effectiveness. The main emphasis is that conclusions on the benefits of selective school attendance should consider a wide range of outcomes and individual characteristics. Students from diverse backgrounds have qualitatively different experiences from enrollment, and this bears on the schools' ability to improve their learning and ensure future academic and professional success.

The Puzzle of Missing Female Engineers: Academic Preparation, Ability Beliefs, and Preferences

The past several decades bore witness to significant changes in education-related gender gaps. In the years immediately following World War II, only one female enrolled in college for every 2.3 males. Yet by the 1980s, women surpassed men in both college enrollment and completion (Goldin, Katz, and Kuziemko, 2006). These successes in postsecondary educational attainment, however, failed to translate into higher rates of female participation in select science, technology, engineering and math (STEM) fields, where women are still vastly underrepresented (Turner and Bowen, 1999; Griffith, 2010). In particular, gender disparities are most glaring in the subfield of engineering (Legewie and DiPrete, 2014). That the sizable engineering gap persists despite reversals of gender gaps in non-STEM subjects and STEM ones such as biology is a prevailing puzzle.

Gender disparities in fields of study have lasting consequences for longer-term earnings and skill distributions. Engineering graduates enjoy a substantial pay premium relative to peers in other fields such as education. The size of this difference

rivals the earnings gap between college and high school graduates (Altonji, Blom, and Meghir, 2012). Differential take-up of science and math-intensive fields account for a notable share of the male-female earnings gap, such that achieving gender parity on major choice could significantly reduce earnings inequality (Paglin and Rufolo, 1990a; Brown and Corcoran, 1997; Blau and Kahn, 2000). Gender gaps in major or occupational choice can also lead to differential accumulation of STEM-focused human capital among men and women that matter for tomorrow’s workforce.

This paper uses new administrative data from North Carolina and a pooled national survey of college freshmen to investigate the largest contributor to STEM disparities: engineering. While a plethora of economics studies focus on the aggregate STEM gender gap, comparatively little research examines specific STEM subfields. However, the divergent patterns by subfield, from postsecondary gender parity in biology to striking gaps in computer science and engineering, necessitate a more targeted approach. This work contributes evidence on the engineering gender gap along three dimensions. The first is to document the size of the gender gap and its evolution from the beginning of high school through postsecondary schooling. The second is to differentiate between the roles of entry versus exit during this period. The final and most substantial component takes advantage of several rich datasets to examine contributors to the gender gap, ranging from individuals’ ability beliefs to the role of family structure and gender-based norms. The availability of a statewide longitudinal dataset linking secondary schooling to postsecondary education enables the paper to provide the most comprehensive account of factors underlying the engineering gender gap.

The datasets’ temporal coverage permits a closer look at entry into engineering beginning in high school. Using engineering orientation or choice as outcome variables, I document a disparity of over 8 percentage points in 9th grade which expands during high school and is 11 percentage points through the first year of postsecondary

education. The magnitude of this gap is especially striking in light of baseline female engineering participation rates between 2 to 4%. Once students have declared an engineering major in the North Carolina public university system, I find no evidence to support systematically higher attrition among female students. Results indicate that the gap is mainly attributable to lower entry among female students rather than higher exit during this period. Efforts to increase the rate of female entry and reduce gender divergence in STEM orientation, in particular engineering, should begin no later than high school.

Tailored policies are informed by evidence on what determines the persistent gender gap in engineering. I investigate four explanatory accounts: differences in academic preparation, differences in academic ability beliefs, differences in prosocial values and professional goals, and the role of family structure and gender-based norms. Decomposition evidence shows that SAT scores and high school GPA account for between 5 to 9% of the overall disparity. Course-taking patterns in the first half of high school betray few clues on eventual major orientation, but over 12% of the postsecondary gap can be explained using linked transcript data in the second half of high school, when inclined students invest more in math and physical science coursework. Meanwhile, beliefs about lower academic ability dissuade women from entering the field even after controlling for academic performance. Elevating women to the same belief levels as their male counterparts would bridge the gender gap by 8%. Female preferences for prosocial responsibilities and contributing to the arts over sciences explain over 14% of the gap.

One common denominator across the main explanatory accounts is the role of environmental and social influences. Differential skill investments or math ability beliefs may reflect gender-specific norms conveyed to students by parents, teachers, and peers. To investigate the extent of these influences, I make use of identifying

variation in the gender composition of twins in North Carolina.¹ Males from opposite-sex pairs are substantially more likely than males from same-sex pairs to choose engineering as a preferred major. These results cannot be explained by differential investment in math skills or a relative advantage story, in which STEM is chosen by the twin with higher math performance. Males from opposite-sex pairs have lower math achievement and the findings are robust to using only higher math-scoring same-sex male twins as a control group. The collective evidence suggests that gender-specific roles and expectations play a role. The salience of gender in opposite-sex pairs can encourage boys to invest in stereotypically male pursuits such as computer skills and engineering.

This paper is organized as follows. The next section highlights engineering’s contribution to the overall STEM gap and grounds this study in related literature. Section 3.2 details the three main administrative and cross-sectional datasets utilized for decomposition. Section 3.3 describes the role of entry vs. exit, while the subsequent section outlines the empirical strategy. Section 3.5 presents evidence on the relevance of each explanatory account. I conclude with a discussion of implications.

3.1 Determinants of STEM gender gaps

In the United States, gender gaps in major choice are large and persistent. Table 3.1 uses Census data to document the share of recent college graduates across all STEM subjects and by subfield.² On aggregate, male graduates are twice as likely

¹ Twins are preferred over siblings because the choice of gender is arguably exogenous within the pair, and is not affected by gender considerations in fertility decisions that are correlated with gender-based investments in children. To my knowledge this is the only STEM orientation study based on a twin sample, although previous work have used siblings (see, for example, Anelli and Peri (2015a)).

² I define a field as STEM if it belongs to one of the following categories: 1) Agriculture, 2) Computer and Information Sciences, 3) Engineering, 4) Engineering Technologies, 5) Biology and Life Sciences, 6) Mathematics and Statistics, 7) Physical Sciences, and 8) Nuclear, Industrial Radiology, and Biological Technologies, abbreviated as Science Tech.

to graduate from college with a degree in STEM. This result, however, disguises large variations by subfield. Degree attainment in biology nears gender parity, while fields such as engineering and computer science still exhibit sizable gaps. Engineering is the most popular STEM major among men and second most popular major among women. 11.7% of male graduates elect this field, compared to only 2.5% of females. The magnitude of the gender differential in engineering participation exceeds all other subfields. Over 9 percentage points of the 16 percentage point gap is attributable to gender disparities in engineering, while computer science contributes an additional 4 percentage points. Since engineering plays an outsized role in informing the STEM gender gap, it is the central focus of this paper.

Table 3.1: Share of recent college graduates in STEM fields

	STEM	Engineer	Comp Sci	Biology	Math and Stats	Physical Sciences	Engineer Tech	Ag	Science Tech
Female	14.8	2.5	0.9	7.0	0.9	2.5	0.2	0.9	0.1
Male	30.5	11.7	4.7	6.9	1.7	3.8	0.7	1.1	0.0

Notes: Source: 2008-2012 American Community Survey. Includes college graduates who are between 22 - 23 years old.

A key objective is to decompose the aggregate gap into explanatory accounts. In choosing these factors I draw upon a wealth of literature that explores cross-gender differences. Previous research in the STEM context concentrates on several sources of disparity: academic skills and preparation, family background and expectations, tastes or preferences such as those related to pecuniary payoffs or the work environment, and psychosocial attributes such as ability beliefs. Within academic preparation, the bulk of the literature focuses on math skills. Evidence on these inputs vary by the time period covered and the ages of individuals studied. Earlier studies often cite differences, although gaps in performance have closed in recent years (Xie and Shauman, 2003; Hyde et al., 2008). Recent studies find small or insignificant gender differences in math standardized tests across elementary and secondary schools

(Hyde et al., 2008; Hyde and Mertz, 2009; Sass, 2015), while others find that gaps only materialize several years into school entry (Fryer and Levitt, 2010). Among the mathematically gifted, evidence for higher variability among males shows diminishing gaps over time. A 13:1 ratio of men to women among high SAT math achievers in the early 1980s has since bridged to approximately 2:1 at the top end of the distribution (Benbow and Stanley, 1983; Ellison and Swanson, 2010). Conditioning on performance still leaves a large unexplained residual in the STEM gender gap, suggesting that academic preparation plays a relatively minor role (Turner and Bowen, 1999; DiPrete and Buchmann, 2013).

Family background is another potential source of influence on individuals' STEM orientation. Parental expectations of children's math and science abilities and academic trajectories may differ by child's gender, thereby affecting students' investments in such skills (Eccles, Jacobs, and Harold, 1990). These expectations can be shaped by parents' own educational and occupational experiences. However, empirical evidence on the role of parental influence for STEM orientation is limited in the economics literature (Fryer and Levitt, 2010).

Pre-labor market skill accumulation also depends on individual preferences. Men may value work in STEM fields higher because they prefer specific tasks performed on the job. Women may enjoy taking non-STEM courses more and sort into those fields on the basis of non-pecuniary factors. Over time, differences across gender preferences can lead to clearly differentiated human capital acquisition. There is growing evidence affirming the important role of preferences. Zafar (2013), for instance, finds that differences in coursework and workplace enjoyment and gaining parents' approval are the primary determinants of divergent major choices among male and female college students.³

³ An important question beyond the scope of this paper is how these preferences develop and evolve over the life course. Evidence shows that environmental factors such as academic context

One explanatory account receiving increasing attention is psychosocial attributes. There is accumulating evidence that gender-based differences in confidence and mind-sets influence individuals' academic behavior (Sax, 1994; Beyer and Bowden, 1997; Dar-Nimrod and Heine, 2006). Conditional on performance, higher confidence can drive men to enter competitive arenas at greater rates than women (Gneezy, Niederle, and Rustichini, 2003; Niederle and Vesterlund, 2007). Beliefs about one's own abilities depend in part on attitudes toward intelligence. Women are more likely to hold a fixed view of intelligence, where ability is intrinsic and cannot be easily gained, while males are more likely to ascribe to an incremental theory of intelligence that enables augmentation through hard work (Dweck, 2000, 2008).⁴ Research shows that these psychosocial attributes affect actual academic decisions. Confidence and an inclination for competitiveness influence females' performance in high-level math tests (Niederle and Vesterlund, 2010). Females who ascribe to a fixed view of ability do worse than those who emphasize an experiential and malleable account of math ability (Dar-Nimrod and Heine, 2006).⁵ The preponderance of math-based curricula in STEM majors bring these issues to the fore. Insofar as students perceive STEM majors to require technical mastery, gender gaps in beliefs about one's own ability can lead men and women to sort into different academic tracks.

matter. Attending single-sex schools or classrooms with higher shares of females can encourage more women to choose STEM majors (Solnick, 1995; Billger, 2002; Favara, 2012; Anelli and Peri, 2015b). Similarly, exposure to female teachers and faculty can increase female students' participation STEM courses and majors (Rothstein, 1995; Bettinger and Long, 2005; Dee, 2007; Carrell, Page, and West, 2010), although some studies find non-existent or only temporary effects (Canes and Rosen, 1995; Sass, 2015). Thus the social or institutional context can shape gender gaps by influencing individual preferences and subsequently affecting students' investments in science and math skills. In this respect, preferences are not independently determined, but rather dynamically connected to academic preparation and social context in its development.

⁴ Emphasizing the importance of fixed intelligence leads to reduced confidence when the task becomes difficult, while focusing on effort promotes long-run confidence (Dweck, 2008).

⁵ The academic context can interact with and activate these attitudes. A recent paper found that academic fields that believe intrinsic, raw talent are important for success exhibit particularly large gender disparities (Leslie et al., 2015).

While the above accounts have been tested in a wide number of studies to illuminate STEM gender disparities, significantly less is known about the interplay of determinants for any given subfield, such as engineering. Existing studies predominantly focus on the engineering gap in the postsecondary context and document the roles of factors such as achievement beliefs or academic performance using survey data (Vogt, Hocevar, and Hagedorn, 2007; Heyman, 2002; Sax et al., 2016). This paper uses statewide administrative data and two national surveys to advance existing evidence along several fronts. Rather than focusing on the gender gap at a point in schooling, I document the evolution of the engineering gender gap from the beginning of high school through college. The focus on earlier schooling years inform the appropriate timing of interventions designed to bridge STEM gender gaps. Within K-12, programs up to middle school are more generally focused on promoting girls' interests in math and science, while high school programs target participation and achievement in gateway STEM courses or specific college-preparatory experiences (Valla and Williams, 2012). Later-timed interventions may be most effective if there is no evidence of significant differential male-female interest by early high school. If men and women may already differ significantly in beliefs, preferences, and other factors by grade 9, earlier interventions are preferable.

In addition to understanding the temporal demarcation of differential entry, this paper uses administrative postsecondary data to examine the role of exit. If women attrit at higher rates during postsecondary, the data is well suited for examining postsecondary factors that lead to learning about one's engineering-specific abilities and tastes. For instance, new inputs such as course grades can unseat previous beliefs about one's aptitude for graduating with an engineering degree. Lastly, the longitudinal nature of North Carolina data permits linking engineering orientation to pre-college academic preparation, from achievement on standardized test scores to course selection and performance. These data sources complement existing surveys

using self-reported achievement data to provide a more comprehensive account of the engineering gender gap.

3.2 Data and descriptive statistics

3.2.1 North Carolina secondary schools

The first dataset comprises administrative records spanning all public and charter secondary schools from the North Carolina Education Research Data Center (NCERDC). Beginning in 2009, NCERDC supplemented the database with College Board data on SAT scores and major intentions at the end of high school. One set of outcome variable derives from students' responses to a question on "First Choice Major." Answers are recorded during the latest administration of the SAT taken by the high school student. I limit the sample to students who do so during their junior or senior year, such that the variable describes major intentions during the second half of high school.⁶

Students' academic achievement variables derive from high school transcript files that detail courses taken and grades associated with each. It is possible to construct cumulative GPA for the first two years of high school. I also focus on earned credit hours in reading, math, physical science, and computer programming during the first half of high school.⁷ The longitudinal combination of students' academic achievement history and forward-looking plans render these data elements suitable for exploring the academic determinants of major orientations. In addition, the SAT questionnaire solicits information on extracurricular activities in grades 9 and 10

⁶ One caveat is that the outcome is conditional on taking the SAT. Students who do not intend to enroll in a 4-year institution or took the ACT in place of the SAT are excluded from this sample. Using the 2009 cohort, 47% of high school seniors in the NCERDC database took the SAT at least once.

⁷ Each student is given a 2-year period for accumulating credits in each subject under the assumption of regular academic promotions. For example, course history information for a graduating senior in 2009 derives from 2007 10th grade and 2006 9th grade transcript files. The physical science category describes cumulative credits earned in physical science, chemistry, and physics courses.

including participation in computer and musical activities.

Table C1 summarizes SAT and GPA performance, earned course credits, and extracurricular participation using the 2009 - 2014 cohorts of graduating seniors. Asterisks denote means among students who do not aspire to engineering that significantly differ from same-gender students who intend to major in engineering. Aspiring female engineers are more selected in academic ability than male counterparts. While female non-engineers have lower SAT math and verbal scores than male non-engineers, females who indicate an engineering interest have the highest SAT math and verbal scores of any group. Aspiring engineers of both genders differentiate themselves by earning more physical science and computer programming credits in the first half of high school.

In addition to rich academic data, NCERDC files provide some context on family structure by flagging twin pairs using identifying information such as exact birth date. I use twin IDs from End-of-Grade exams to construct a sample of twin siblings ever enrolled in a North Carolina public or charter school between 2004 and 2007. The analytic sample includes pairs who are matched to 2009 - 2014 College Board data with non-missing SAT scores. Of the 9569 pairs in the full sample, approximately one-third are opposite-sex twins, with the remaining split between same-sex female twins and same-sex male twins. Sex allocation within these pairs provides the identifying variation for subsequent analyses.

3.2.2 University of North Carolina

Secondary school data in the NCERDC are complemented with administrative records from the University of North Carolina (UNC) system. A unique ID enables researchers to track each student's progress from public school entry through college graduation at any one of UNC's 16 campuses. This permits linking students' secondary school transcripts with UNC records. I expand the set of covariates to include

earned credits in reading, math, physical science, and computer programming courses in grades 11 and 12.

Outcome variables derive from enrollment and graduation records, which include students' term-by-term credit accumulation, GPA and major choice. This allows for tracking course credit milestones and exploring the evolution of major choice. For instance, I can compare engineering take-up rates by gender after students attain 30 or 60 credit hours at their home institution. Enrollment records also contain student qualifications including SAT math and verbal scores and high school cumulative grade point average.⁸

The base sample comes from the 2003-2010 enrollment files, with approximately 80% of records linking to NCERDC data. The majority of the analyses uses the 2009-2010 enrollment cohorts due to the relatively recent availability of high school transcript files. I drop unmatched observations that correspond to out-of-state students, in-state residents who attended private school, and public school attendees with missing transcript information.

Summary statistics in C2 juxtapose the academic achievements and high school course-taking patterns of the 22,174 students who declared an engineering major by the time they have earned 30 credit hours at their home institutions. Consistent with the high school sample, there is more selection on both SAT math and verbal scores among women. Aspiring female engineers are primarily distinguished by more earned credits in math and physical science courses. The course-taking trajectories of aspiring male engineers are particularly discernible by the second half of high school, marked by fewer reading credits and more math, physical science, and computer programming courses.

⁸ High school GPA in the UNC sample comes from university enrollment records and runs along a scale of 0 to 6.

3.2.3 CIRP Freshmen Survey

The third dataset is an annual survey of entering full-time freshmen students at colleges and universities administered by the Cooperative Institutional Research Program (CIRP). As the largest continuous national survey of college students, the CIRP Freshmen Survey provides a snapshot of incoming students' background characteristics and college expectations.⁹ The dependent variable on engineering intentions comes from a question eliciting students' probable fields of study. Family background variables include parents' occupational categories and total income, while measures of academic preparation derive from student self-reports of SAT and ACT scores and high school GPA.¹⁰

A key advantage of the survey is the breadth of self-assessments that permits evaluating several prevalent factors for their contribution to the gender gap. The first involves self-confidence and academic ability beliefs. The survey includes a spectrum of self-reported abilities, including academic and mathematical.¹¹ Conditioning on standardized test scores, high school GPA, and other objective measures of academic performance allows for examining the role of academic self-confidence independent of true academic ability. Apart from ability beliefs, some students may prefer engineering for its expected pecuniary benefits or compatibility with preferences for problem-solving and scientific inquiry. These tastes and preferences are partially captured via questions on personal goals and expected future acts.¹² I in-

⁹ Most universities and colleges administer the survey during student orientation, although the survey is typically made available between March and October annually.

¹⁰ Previous literature on the relationship between self-reported and actual GPAs find reasonable validity, with a particularly strong positive correlation for higher ability students (Kuncel, Credé, and Thomas, 2005).

¹¹ Students are asked to rate themselves on each trait along a scale of 1) lowest 10%, 2) below average, 3) average, 4) above average, and 5) highest 10%.

¹² Students evaluate the personal importance of each social, political, academic or economic goal by describing them as 1) not important, 2) somewhat important, 3) very important, and 4) essential. The survey also asks students to guess the probability of undertaking a future action, such as

investigate whether economic considerations exert influence by using a question on the importance of “being very well off financially.” The role of social and other-regarding values are embedded in questions on the importance of “helping others who are in difficulty,” “influencing social values,” and likelihood that the student will “participate in volunteer or community service work.” Professional goals in the arts and sciences are captured by the importance of “creating artistic works” and “making a theoretical contribution to science.” Finally, I evaluate the role of family considerations. The relationship between family considerations and major choice relies on a variable on the importance of raising a family. If men believe that an engineering career is compatible with family life, while women find it difficult to balance an engineering career with raising children, then these expectations can affect their participation.

While the CIRP Freshmen Survey extends as far back as 1965, I constrain the analytic sample to more recent cohorts to ensure the consistency of survey variables over time. The sample retains all students in four-year colleges or universities with non-missing demographics information, SAT scores, and parental occupational categories.¹³ The pooled cross-sectional base sample of 2,042,832 students spans the 1990-1999, 2001, 2004, 2006, 2008 and 2010 cohorts. Average SAT scores in the survey are higher than the UNC sample, reflecting likely compositional differences in the sample of participating universities and students (Table C3). Aspiring engineers are highly selected on attributes such as mathematical ability beliefs and interest in making a theoretical contribution to science.

changing their major or dropping out of college. Students choose between 1) no chance, 2) very little chance, 3) some chance, and 4) very good chance.

¹³ For the small subset of students with only ACT composite scores and missing SAT values, I input SAT math and verbal scores using the sample of individuals who took both standardized exams and a quartic of ACT scores. Across the sample, less than 5% of academic beliefs, personal goals, and expected future acts covariates have missing values. I include indicators for missing data and sample means in place of missing data.

3.3 The engineering pipeline

Table 3.2 traces the engineering gender gap from the beginning of high school to postsecondary education. Multiple data sources converge on sizable and persistent gaps in engineering orientation and major choice throughout this period of schooling. The earliest data point comes from the High School Longitudinal Study of 2009. 9th grade students exhibit a 8.2 percentage point gender disparity in engineering orientation, defined by a preference for an engineering job or occupation at age 30. By the second half of high school, data from the College Board shows a 14.7 percentage point gap, comparable to the 14.0 percentage gap documented at the beginning of students' postgraduate careers using the CIRP Freshmen Survey.

The suggestive broadening of the gender gap from early high school through the beginning of college is consistent with several explanations. Patterns could be attributable to differences in the sampled population, from the nationally representative HSLS:09 to aspiring college students in North Carolina and a national sample of college freshmen. Furthermore, questions soliciting engineering orientation varied from planned occupational choice in the HSLS:09 to preferred majors in the two other samples. Despite these differences, evidence exists to support an actual broadening of the gender disparity during high school. The first HSLS follow-up in 2012 found an engineering gap of 9.8 percentage points among students expected to be in 11th grade. The longitudinal nature of the survey and the stability of the outcome variable across waves point to growing gender differences in engineering orientation during high school.

One concern with high school data is that major intentions indicated on a college entrance exam will not endure into college and translate into actual major choice. The last specification follows students' decisions using UNC administrative files that document term-by-term major choice. Column (4) presents a snapshot of the full

Table 3.2: The engineering gender gap from high school to college

	High school		Postsecondary	
	HSLs: Gr. 9	Gr. 11 & 12	CIRP: freshmen	UNC: 30 credit hrs
	(1)	(2)	(3)	(4)
Female	-0.082*** (0.003)	-0.147*** (0.002)	-0.140*** (0.000)	-0.107*** (0.030)
Black	-0.008 (0.005)	0.001 (0.002)	0.015*** (0.001)	-0.035** (0.013)
Hispanic	-0.013*** (0.004)	0.009*** (0.003)	0.021*** (0.001)	-0.009 (0.007)
Asian	0.028*** (0.009)	0.026*** (0.003)	0.044*** (0.001)	-0.008 (0.010)
American Indian	-0.016 (0.022)	-0.011*** (0.004)	-0.012*** (0.005)	-0.017 (0.013)
Other	-0.004 (0.006)	-0.012*** (0.003)	0.006*** (0.001)	-0.005 (0.008)
Observations	14668	266910	2042832	22174
R^2	0.040	0.070	0.072	0.129

Notes: HSLs:09 comprises a nationally representative sample of 9th graders in 2009. The outcome variable is constructed from weighted student responses to the question “what is the job or occupation that you expect or plan to have at age 30?” using the SIOCC30 variable. The College Board sample comprises students who took the SAT exam in 2009 - 2014 at least once as juniors or seniors. The outcome variable is constructed from a variable eliciting students’ first choice majors. The sample includes observations with non-missing SAT scores and those who can be linked to 9th or 10th grade transcript data. The base specification uses indicator variables for cohort year and academic level during the last SAT test administration. The Freshmen Survey sample spans 1990-1999, 2001, 2004, 2006, 2008, and 2010 academic years. The outcome variable comes from a variable eliciting freshmen’s probable field of study or major. The base specification uses year and college type indicators and student weights. The UNC sample comprises 2009-2010 enrollees who declared a major by the time of attaining 30 credit hours at their home institution, with non-missing high school GPAs, SAT scores, and transcript data. The specification also includes campus fixed effects. Standard errors clustered by campus. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

sample of UNC students who attained at least 30 credit hours (approximately one year of study) at their home institutions. Among this group, females are 10.7 percentage points less likely to choose engineering. The magnitude of this disparity becomes more apparent when it is compared to baseline participation rates: just over 3 percent for women.

The gender gap’s magnitude during high school and early postsecondary years underscores the role of differential entry rates among male and female students. Alongside entry, researchers also conjecture that females switch out at higher rates due to any combination of academic, social, and environmental factors. The longitu-

Table 3.3: Attrition among engineering students

	Raw gap	Conditional gap	
Female	0.023 (0.014)	0.013 (0.012)	0.015 (0.011)
Black		-0.049 (0.030)	-0.008 (0.007)
Hispanic		-0.053** (0.018)	-0.048** (0.015)
Asian		0.020 (0.016)	0.022 (0.014)
American Indian		0.011 (0.032)	0.017 (0.025)
Other		0.028 (0.015)	0.031 (0.017)
SAT and high school GPA		Yes	Yes
Cohort FE		Yes	Yes
UNC campus FE			Yes
Observations	14889	14889	14889
R^2	0.000	0.036	0.040

Notes: Sample comprises students whose initial declared major is engineering. Models are augmented with indicators for SAT math and verbal scores for each 10-point bin, and for each high school GPA decile. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

dinal nature of the UNC sample provides a suitable context for examining persistence in and attrition from engineering. I follow a sample of full-time, first-time freshmen enrollees in 2003-2010 whose initial declared major was engineering. Table 3.3 shows a raw attrition gap among females that is statistically insignificant from male peers. Conditioning on cohort fixed effects and academic performance as measured by SAT scores and high school GPA attenuates the still insignificant gender disparity in persistence. Taken together, the evidence suggests that higher attrition among aspiring women is not a first-order explanation for postsecondary gender disparities in engineering. Women have already sorted away from the field at the point these datasets begin documenting major intentions.¹⁴

¹⁴ The present timespan precludes studying attrition in the labor market. As such, subsequent

3.4 Empirical approach

The prominence of entry differentials begs the question of what determines students' decisions to pursue engineering. In order to characterize inputs into the major choice decision that draw on common hypotheses cited in the literature, we can express the utility attached to engineering as:

$$U = f(X, A, S, \Lambda) \quad (3.1)$$

X is a set of sociodemographic characteristics such as gender, race, and parental occupations. A captures academic preparation and achievement reflecting individuals' ability to perform well in engineering-related coursework. This variable also encompasses the effort exerted to lay the academic foundations for entering this major, as well as exposure to key math and science courses in secondary education or earlier that can shape students' engineering interests. S is a set of beliefs about ability shaping students' expectations for success. S is subject to interactions with individual attributes such as gender and ability, to the extent that parental and classroom influences shape ability beliefs differentially along individual characteristics and academic achievement measures. Finally, Λ describes preferences and work-related tastes such as pecuniary vs. non-pecuniary motivations and prosocial orientation. This furthermore reflects the cumulative effect of exposure to social context and its interactions with gender.

The baseline model I estimate conditions on only a parsimonious set of individual characteristics X to characterize the raw gender gap in engineering:

$$Y_i = \alpha^{base} + \beta^{base} Fem_i + \phi^{base} X_i + \epsilon_i^{base} \quad (3.2)$$

attrition from the engineering workforce due to professional, social, and family reasons may play a significant role but are outside the scope of this study.

Y_i indicates whether individual i expresses an engineering orientation or is observed to major in the subject. X is a set of individual characteristics such as race and timing of student responses. The parameter of interest, β , expresses the average gender difference in engineering participation. In order to determine the contribution of ability beliefs, prosocial preferences, and other contextual factors to the gap, I augment the baseline model with several determinants in the full specification:

$$Y_{ij} = \alpha^{full} + \beta^{full} Fem_{ij} + \phi^{full} X_{ij} + \gamma_1^{full} A_{ij} + \gamma_2^{full} S_{ij} + \gamma_3^{full} \Lambda_{ij} + \eta_j + \epsilon_{ij}^{full} \quad (3.3)$$

The subscript j denotes high school. A is a vector of academic preparation variables including SAT math, SAT verbal, high school GPA, and high school credits earned in reading, math, science, and computer programming. I assume that the effect of ability beliefs S is estimable via survey questions that elicit self-assessments of mathematical, academic, and writing abilities. Λ represents a vector of survey responses that capture pecuniary, other-regarding, work-based, and family preferences. There are likely other inputs into the major decision-making process not captured by existing covariates. For instance, students with unobserved preferences for engineering coursework may select into different high schools offering contrasting academic environments that shape major orientation. I include high school fixed effects η_j to address selection on unobservable characteristics that may be correlated with A_{ij} , S_{ij} , and λ_{ij} .¹⁵ Due to data constraints, this specification abstracts away from the dynamic relationships between human capital investment, ability beliefs, and preference formation.¹⁶

¹⁵ I estimate the model using high school fixed effects in the College Board and UNC samples only because the CIRP Freshmen Survey does not elicit high school IDs.

¹⁶ An individual with higher math ability beliefs, *ceteris paribus*, will invest more in STEM-related human capital, which can further shape engineering-related ability beliefs and preferences for engineering coursework. A lack of panel data on preferences and beliefs constrain the analyses to provide a static perspective.

A hurdle in estimating the contribution of each covariate to β is that results are non-robust to the sequence in which covariates are added to the base regression. I rely on the decomposition technique of Gelbach (2016) to render the factor contributions order-invariant.¹⁷ The decomposition relies on the sample omitted variable bias formula to explain the sensitivity underlying the relationship between β and included covariates. The portion of the gender gap explained by new explanatory variables is $\hat{\delta}_{Fem} = \hat{\beta}^{base} - \hat{\beta}^{full}$. This total difference is separable into k additional covariate groups:

$$\hat{\delta}_{Fem} = \sum_k \hat{\delta}_{k,Fem} = \sum_k \hat{\Gamma}_{k,Fem} \hat{\gamma}_k^{full} \quad (3.4)$$

This setup makes clear that $\hat{\delta}_{k,Fem}$, the contribution of the k -th covariate (group), is the product of two channels of influence. The first is the male-female difference in this factor after partialling out all other explanatory elements in the base regression. $\hat{\Gamma}_{k,Fem}$ is the coefficient on Fem from an auxiliary regression of the k -th covariate on all explanatory variables in the base model. The amount explained by SAT math scores, for instance, depends on the raw gender difference in this attribute after conditioning on the basic set of individual characteristics. The second channel is how correlated the k -th covariate is to the outcome under the full model. If the coefficient associated with SAT math is sufficiently small, it will not be a meaningful determinant of the gender gap.

¹⁷ The approach of Gelbach (2016) generalizes the Oxaca-Blinder technique while ensuring path independency. A more detailed presentation of the technique is available in Appendix A.

3.5 Results

3.5.1 High school academic achievement and extracurriculars

All three analytic samples permit an examination of the role of academic preparation on the engineering gender gap. The first model in Table 3.4 replicates the unconditional gap of 14.7 percentage points from Table 3.2 and is augmented with standardized test scores, cumulative GPA during the first half of high school, and earned credits in reading, math, physical science, and computer programming.

Female students score on average 28 and 4 points lower in the mathematics and verbal portions of the SAT while receiving higher grades. Controlling for only SAT performance reduces the gender gap to 13.3 percentage points, since students are positively selected on better math scores to enter engineering and negatively selected on superior verbal performance. The addition of GPA indicators widens the gap to 13.6 percentage points to allow for higher male engineering participation if they had the same grades as their female peers. Replacing these covariates with earned credits in grades 9 and 10 shows that both male and female students inclined towards engineering invest in fewer reading credits and more physical science and computer engineering credits. Accounting for these covariates, however, leads to little change in the female coefficient.

I examine whether gender differences in the choice of high school extracurriculars can inform students' major intentions. Seven activities in the first half of high school spanning computer, journalism, government, music, theater, dance, and ROTC attenuate the gap to 14.2 percentage points in Table 3.4. Female participation is higher in the theater and arts, although aspiring female engineers are similarly distinguished as their male counterparts by their reduced involvement in these activities relative to same-gender peers. Participation in computer-related activities is strongly associated with higher engineering take-up rates among both genders.

Table 3.4: High school gender gap in engineering intentions

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.147*** (0.002)	-0.146*** (0.002)	-0.136*** (0.002)	-0.146*** (0.002)	-0.142*** (0.002)	-0.129*** (0.002)
Earned reading credits in Grade 9				-0.004** (0.002)		-0.006*** (0.002)
Earned math credits in Grade 9				-0.002 (0.002)		0.003** (0.001)
Earned physical science credits in Grade 9				0.012*** (0.003)		0.004 (0.003)
Earned computer programming credits in Grade 9				0.032*** (0.011)		0.013 (0.013)
Earned reading credits in Grade 10				-0.007*** (0.001)		-0.007*** (0.001)
Earned math credits in Grade 10				0.015*** (0.002)		0.009*** (0.001)
Earned physical science credits in Grade 10				0.021*** (0.002)		0.002 (0.001)
Earned computer programming credits in Grade 10				0.045*** (0.010)		0.029*** (0.011)
High school fixed effects		Yes				Yes
SAT and high school GPA			Yes			Yes
Extracurriculars in 9th and 10th grades					Yes	Yes
Observations	266910	266910	266910	266910	266910	266910
R^2	0.070	0.077	0.091	0.072	0.072	0.101

Notes: all specifications include indicators for race, cohort and academic level during the last SAT administration. Models also augmented with indicators for SAT math and verbal scores for each 10-point bin, and for each high school GPA decile. Extracurriculars include indicators for participation in computer, music/vocal, theater, junior ROTC, dance, government/political, and journalism/literary activities during both 9th and 10th grades. * p<0.1, ** p<0.05, *** p<0.01

Table 3.5: Gender gap after 30 credit hours in UNC

	(1)	(2)	(3)	(4)	(5)
Female	-0.107*** (0.030)	-0.094*** (0.024)	-0.107*** (0.029)	-0.092*** (0.024)	-0.084*** (0.021)
Earned reading credits in Grade 9				-0.009*** (0.003)	-0.007** (0.003)
Earned math credits in Grade 9				-0.007 (0.005)	0.002 (0.003)
Earned physical sci. credits in Grade 9				0.004 (0.006)	0.003 (0.007)
Earned comp. prog. credits in Grade 9				0.056 (0.037)	0.031 (0.040)
Earned reading credits in Grade 10				-0.006* (0.003)	-0.001 (0.002)
Earned math credits in Grade 10				0.007** (0.003)	0.009** (0.004)
Earned physical sci. credits in Grade 10				0.025** (0.010)	0.017* (0.009)
Earned comp. prog. credits in Grade 10				-0.013 (0.012)	-0.018 (0.014)
Earned reading credits in Grade 11				-0.018** (0.007)	-0.013*** (0.003)
Earned math credits in Grade 11				0.013* (0.006)	0.012** (0.004)
Earned physical sci. credits in Grade 11				0.028** (0.011)	0.028** (0.010)
Earned comp. prog. credits in Grade 11				0.019 (0.032)	0.015 (0.033)
Earned reading credits in Grade 12				-0.008*** (0.002)	-0.005* (0.003)
Earned math credits in Grade 12				0.029*** (0.010)	0.027*** (0.009)
Earned physical sci. credits in Grade 12				0.058** (0.021)	0.052*** (0.016)
Earned comp. prog. credits in Grade 12				0.050** (0.023)	0.056** (0.022)
SAT and GPA		Yes			Yes
High school fixed effects			Yes		Yes
Observations	22174	22174	22174	22174	22174
R^2	0.129	0.160	0.149	0.150	0.193

Notes: sample comprises enrolled UNC students from 2009-2010 matched to transcript files who have declared a major by the time of attaining 30 credit hours at their home institution. Models include indicators for race, cohort, each SAT math and verbal 10-point bin, high school GPA decile, and UNC campus fixed effects. Standard errors are clustered at the campus level. * p<0.1, ** p<0.05, *** p<0.01

The longitudinal nature of North Carolina data permits studying the contribution of high school academic preparation for a supplementary population: first-time undergraduates in the UNC system. Table 3.5 uses the sample of students who have

completed at least 30 credit hours at their home institutions. Accounting for SAT scores and high school grades reduces the disparity from 10.7 to 9.4 percentage points. Conditioning on a detailed set of earned course credits using detailed transcript data in grades 9-12 reduces the gap to 9.2 percentage points. Fewer reading credits are consistently associated with a proclivity towards engineering across grades, while more math and physical science credits take on significance beginning only in 10th grade.

A shortcoming of OLS results is that the contributions of individual observables are yet unknown. Table 3.6 decomposes the full models into the contributions of its constituent parts. I juxtapose the explanatory power of several sets of covariates across the high school and UNC samples: SAT scores and GPA, credits earned, extracurricular activities, and high school choice. The full high school model explains 12.2% of the baseline 14.7 percentage point cross-gender disparity, compared to 21.5% of the UNC sample. In both cases substantial variation is left unexplained by mean differences in men and women's observable pre-college academic trajectories.

SAT scores and cumulative GPAs explain 7.5% and 8.8% of the gap in the high school and UNC samples, respectively. This net result disguises the oppositional effects of SAT and GPA. Conditioning on SAT math and verbal scores attenuates the gap, while the addition of grades exacerbates it due to higher average GPA among female students. While SAT scores taken alone explain more than 7.5% and 8.8%, the cumulative magnitude is significantly less than comparable studies using earlier cohorts. For example, nearly one-third of the engineering gender gap among college graduates in the 1989 entering cohort are explained via differences in SAT performance (Turner and Bowen, 1999).¹⁸ This is consistent with the literature showing

¹⁸ SAT scores may play a greater role in this study due to the composition of high-ability students in the College and Beyond Database. They originate from 12 institutions: Stanford, Yale, Princeton, Kenyon, Oberlin, Swarthmore, Hamilton, Williams, Wesleyan, Bryn Mawr, Smith, and Wellesley. On the other hand, the use of categorical variables for SAT scores in place of 10-point indicators

Table 3.6: Gelbach decomposition: High school and UNC samples

	High school sample		UNC sample	
	Contribution	% explained	Contribution	% explained
High school fixed effects	-0.000*	-0.3%	-0.000	-0.2%
SAT and high school GPA	-0.011***	-7.5%	-0.009***	-8.8%
SAT math				
SAT verbal				
Cumulative high school GPA				
Earned credits in Grades 11 & 12			-0.013***	-12.2%
Reading				
Math				
Physical science				
Comp. programming				
Earned credits in Grades 9 & 10	-0.001***	-0.6%	-0.000	-0.2%
Reading				
Math				
Physical science				
Comp. programming				
Grades 9 and 10 extracurriculars	-0.006***	-3.8%		
Computer activity				
Music/vocal activity				
Theater activity				
Junior ROTC				
Dance activity				
Government/political activity				
Journalism/literary activity				
Total explained	-0.018***	-12.2%	-0.023***	-21.5%

Notes: Gender gap in the College Board sample reduces by 1.8 percentage points, from 14.7% to 12.9% when using the full model. The gap attenuates from 10.7% to 8.4% in the UNC sample. * p<0.1, ** p<0.05, *** p<0.01

convergence in standardized test scores over recent decades despite a persistent engineering gender gap.

Earned credits across multiple subjects in the first half of high school have little tangible effect on male-female differences in engineering orientation, explaining 0.6% and 0.2% of the gaps in the high school and UNC samples, respectively. Curriculum exposure in grades 11 and 12, in contrast, is markedly more meaningful. Male-female variation in reading, math, physical science, and computer programming credits earned account for 12.2% of the disparity in the UNC sample. Later high school years permit greater flexibility in upper level math and science course electives, such

may underestimate the contribution of SAT scores to the gender gap.

that students with existing tastes for engineering can invest in related courses. In addition to curriculum choice, extracurricular activities explain 3.8% of the gap in the high school sample. The single largest contributor is participation in computer-related activities. 15% of aspiring female engineers distinguish themselves early by engaging in these extracurriculars during grades 9 and 10, compared to 9% of non-engineers. Analogous statistics among males are 16% and 12%.

The availability of rich administrative data linking secondary and postsecondary permits a thorough exploration of the roles of academic and extracurricular trajectories. The choice of high school accounts for almost none of the gender gap in both samples, suggesting that the gender gap phenomenon is consistently driven by within-school variation and not by males disproportionately selecting into high schools exhibiting higher engineering participation. Skill gaps as measured by standardized test scores and GPA explain no more than 9% of female under-representation in engineering. More variation in engineering orientation is accounted for by the combination of extracurriculars and math, science, and reading curricular momentum in the second half of high school.

3.5.2 Ability beliefs, prosocial values and other preferences

The CIRP Freshmen Survey expands the set of explanatory accounts to include select ability beliefs alongside professional and personal preferences. I classify survey covariates into five categories: beliefs in academic abilities, pecuniary goals, other-regarding values, professional contributions in the arts and sciences, and family considerations. The unconditional gap of 14.0 percentage points in the first specification presented in Table 3.7 attenuates to 12.8 percentage points when conditioning on SAT scores, high school GPA, parental occupations, and the total number of college applications. A greater reduction of the disparity, to 10.2 percentage points, occurs

when including ability belief and preference covariates.¹⁹ Self-confidence in math ability and interest in making a theoretical contribution to science greatly increase the probability of choosing engineering. Other-regarding preferences and tastes for artistic or creative careers deter entry into engineering.

Table 3.7: Gender gap in CIRP Freshmen Survey

	Full sample			Female	Male
	(1)	(2)	(3)	(4)	(5)
Female	-0.140*** (0.000)	-0.128*** (0.000)	-0.102*** (0.000)		
Self-confidence: academic ability					
Self Rating: Academic ability			-0.003*** (0.000)	-0.000 (0.000)	-0.008*** (0.001)
Self Rating: Mathematical ability			0.030*** (0.000)	0.017*** (0.000)	0.046*** (0.000)
Self Rating: Writing ability			-0.016*** (0.000)	-0.006*** (0.000)	-0.026*** (0.000)
Pecuniary goals					
Goal: Being very well off financially			0.005*** (0.000)	0.004*** (0.000)	0.005*** (0.000)
Social and other-regarding values					
Goal: Helping others who are in difficulty			-0.022*** (0.000)	-0.013*** (0.000)	-0.031*** (0.001)
Goal: Influencing social values			-0.020*** (0.000)	-0.007*** (0.000)	-0.032*** (0.000)
Future Act: Participate in volunteer/service work			-0.004*** (0.000)	-0.001*** (0.000)	-0.005*** (0.000)
Professional goals in arts and sciences					
Goal: Become accomplished in the performing arts			-0.010*** (0.000)	-0.004*** (0.000)	-0.017*** (0.000)
Goal: Creating artistic work			-0.009*** (0.000)	-0.004*** (0.000)	-0.017*** (0.001)
Goal: Make a theoretical contribution to science			0.049*** (0.000)	0.027*** (0.000)	0.071*** (0.000)
Family considerations					
Goal: Raising a family			0.001*** (0.000)	-0.003*** (0.000)	0.005*** (0.000)
SAT and high school GPA		Yes	Yes	Yes	Yes
Parental occupations		Yes	Yes	Yes	Yes
Number of colleges applied to		Yes	Yes	Yes	Yes
Observations	2042832	2042832	2042832	1083838	958994
R^2	0.072	0.116	0.151	0.071	0.143

Notes: Survey sample spans 1990-1999, 2001, 2004, 2006, 2008, and 2010 academic years. All specifications include race, year, and college type dummies. Academic performance covariates include indicators for each 10-point SAT bin and GPA category. All models use student weights. * p<0.1, ** p<0.05, *** p<0.01

The ensuing two columns estimate the model separately for males and females to

¹⁹ I assume a linear interpretation for ease of presentation, such that an increase in self-rating from 3 (Average) to 4 (Above average) is given identical interpretation as an increase from 4 to 5 (Highest 10%). Results remain robust to the use of dummies for each response category.

assess whether the effect of each attribute on engineering decisions is similar across genders. Coefficients are generally larger in magnitude for the male sample because the baseline engineering participation rate among men of over 18% far exceeds the 4% for women. Preferences affect both genders' choices in a similar direction with the exception of family considerations. Females who assign high importance to the goal of raising a family sort away from engineering while the opposite holds true for men, although the coefficient on the aggregate sample is small in magnitude.

Table 3.8 decomposes the 14.0 percentage point gap into the relative contributions of academic preparation, ability beliefs, and preference variables. The full model explains over 27% of the aggregate gap. Cross-gender differences in SAT scores and high school GPA contribute 4.8% to the overall variation, less than the 8-9% of the two previous samples. One explanation for this disparity is differences in student composition, with a significantly lower share of African American students in the national sample relative to North Carolina. Another reason is classical measurement error that can be introduced via self-reported academic achievement scores.

Beliefs in academic, mathematical, and writing ability explain 7.5%. The majority of this effect is driven by male-female differences in math ability beliefs. A female student exhibiting identical academic performance as a male peer rates herself significantly lower on math skills. Since students who are confident along this dimension are more likely to select into engineering, equalizing females' math ability beliefs with males' bridges the gender gap by an amount that nearly rivals the cumulative contribution of standardized test scores and GPA. Notably, neither pecuniary goals nor family considerations had a sizable impact. While men were more likely to elevate the importance of financial gain, pecuniary goals only contributed 0.5% to the overall gap. The small mean difference in men and women's survey responses led to family considerations contributing a negligible amount to the gender disparity.

Next I turn to two accounts focused on preferences: prosocial values and profes-

Table 3.8: Gelbach decomposition: CIRP Freshmen Survey

	Contribution	Share of gap
SAT and high school GPA	-0.007***	-4.8%
Self-confidence: academic ability	-0.011***	-7.5%
Self Rating: Academic ability		
Self Rating: Mathematical ability		
Self Rating: Writing ability		
Pecuniary goals	-0.001***	-0.5%
Goal: Being very well off financially		
Social and other-regarding values	-0.011***	-7.9%
Goal: Helping others who are in difficulty		
Goal: Influencing social values		
Future act: Participate in volunteer/service work		
Professional goals in the arts and sciences	-0.009***	-6.5%
Goal: Becoming accomplished in the perf. arts		
Goal: Creating artistic work		
Goal: Making a theoretical contribution to science		
Family considerations	0.000	0.0%
Goal: Raising a family		
Parental occupations	0.000***	0.2%
Father's aggregate career category		
Mother's aggregate career category		
Number of college applications	-0.000***	-0.1%
Total	-0.038***	-27.1%

Notes: reduces gender gap from 14.0% to 10.2%. Survey variables are entered linearly. Includes year and college type fixed effects, and use student weights. * p<0.1, ** p<0.05, *** p<0.01

sional goals. The former comprise 7.9% of the gap. Coefficients in Table 3.7 suggest that engineering deters students who are socially-oriented and other-regarding, exemplified by those who place greater importance on helping others in difficulty and influencing social values. Women are over-represented in this group, as 27% of female respondents believe that helping those in difficulty is essential, compared to

17% of males. Their other-regarding preferences underscore the perception of engineering as a field that is less compatible with contributing to the social good. Cross-gender differences in professional goals explain an additional 6.5%. Individuals who aim to make their mark on the creative and performing arts select away from engineering. Not surprisingly, those intent on making a theoretical contribution to science are significantly more likely to become an engineer. Hence, engineering attracts theoretically-minded individuals over those who prioritize community action and artistic pursuits. These variables capture pre-college differences in tastes for future careers that manifest in preferred major choice.

Findings on the influential roles of confidence and preferences contribute to what can best be described as mixed evidence in the literature. Among the few gender gap studies that jointly focus on preferences and ability beliefs, Zafar (2013) found that the majority of the gap is explained by gender differences in preferences and expected enjoyment of studying in different fields. He rejected a meaningful role for self-confidence, in contrast to the prominent role occupied by ability beliefs in other studies (Valian, 1998; Antecol and Cobb-Clark, 2013; Leslie et al., 2015).²⁰ This paper establishes that beliefs in academic ability, in addition to individual preferences and tastes, matter for initial entry into engineering.

Effect heterogeneity

Results from the full sample of college freshmen may disguise heterogeneous responses across individual attributes. This section examines the variation in factor contributions by ethnicity and math ability to inform whether ability beliefs and preferences explain a greater share of disparity among particular demographic or achievement

²⁰ In Zafar (2013), the cumulative contribution of academic ability beliefs, reconciling work and family, and beliefs about future earnings is less than 5% of the aggregate engineering gap and statistically insignificant, compared to 27% explained by beliefs about coursework enjoyment and 60% by other preferences.

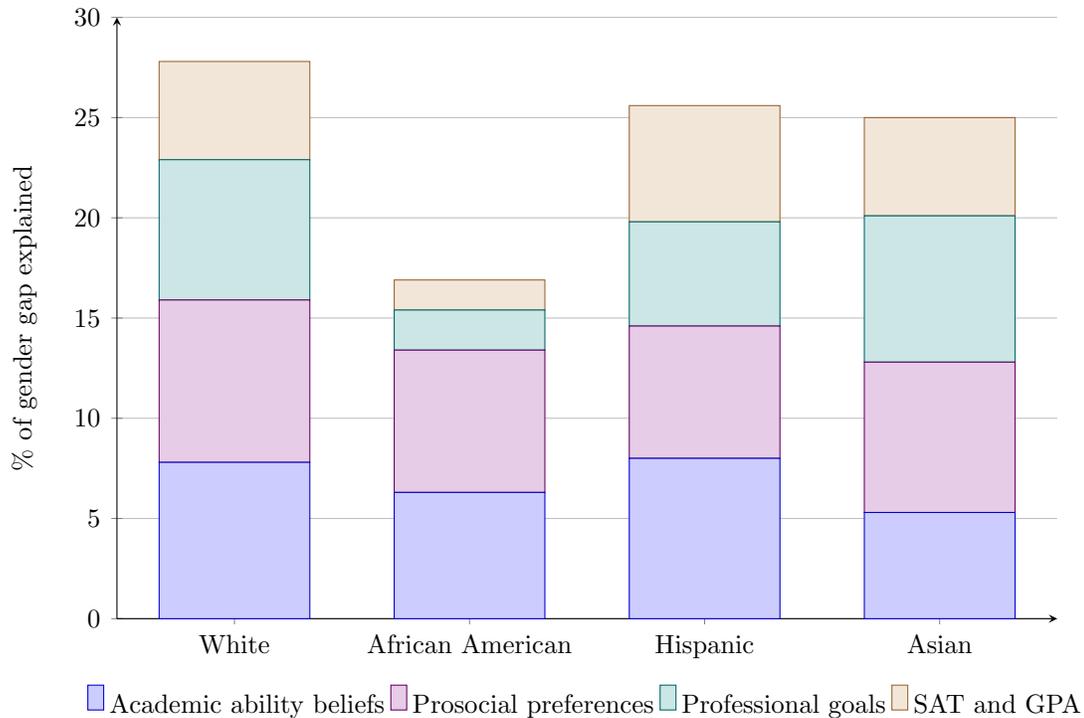


FIGURE 3.1: Decomposition by race

groups. The interaction between these factors and individual characteristics can arise for multiple reasons. For instance, cultural backgrounds can differentially shape how females perceive their academic mastery such that the contribution of academic ability beliefs vary significantly across ethnic groups. The female math self-confidence deficit can also play an outsized role among high achievers whose STEM entry decisions are especially sensitive to their self-perceived ability. Cross-gender differences in the importance of making a scientific contribution may become relevant only among high-scoring students who disproportionately aspire to these careers.

Figure 3.1 segments the sample into four groups: white, African American, Hispanic, and Asian. The aggregate explanatory power of academic ability beliefs, prosocial preferences, professional goals, and SAT scores and GPA is the highest for white students at 28%. Hispanic and Asian students lag slightly behind, while the model accounts for only 17% of the African American gender disparity. Professional

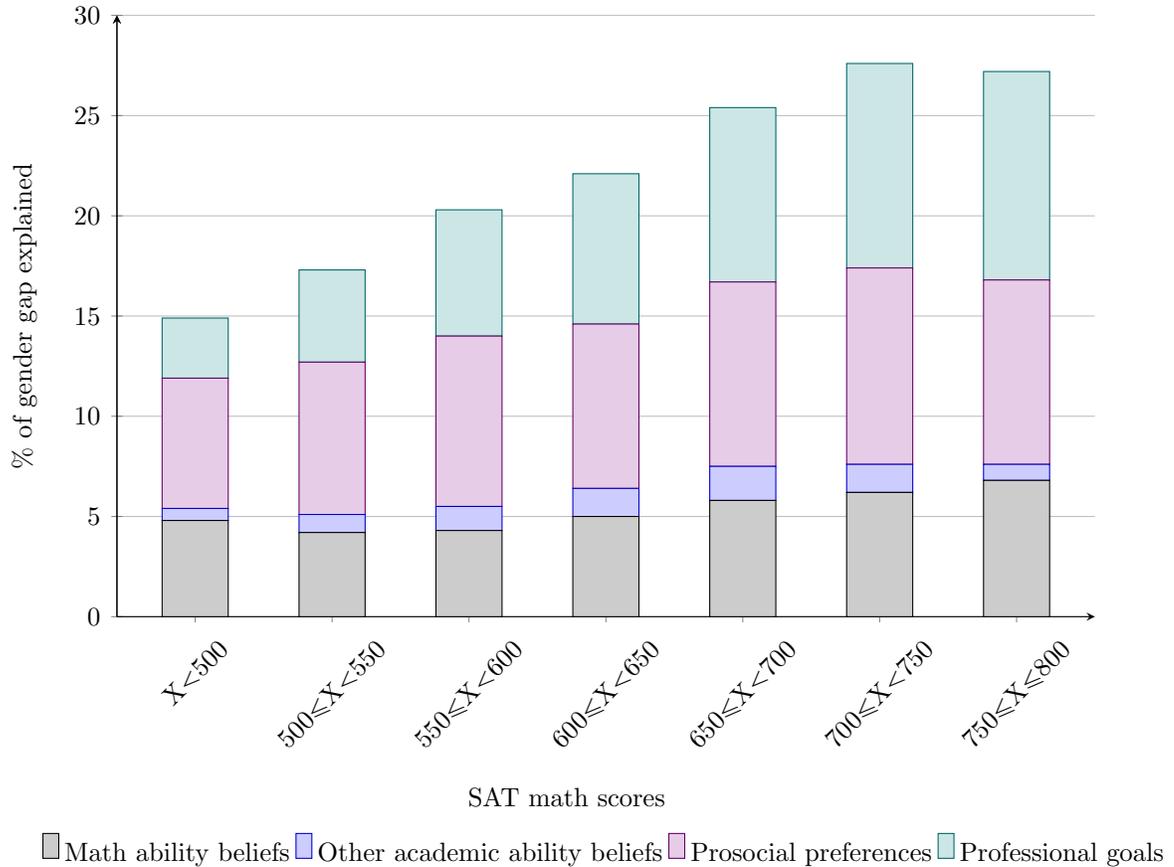


FIGURE 3.2: Decomposition by SAT math scores

goals and academic performance play a relatively marginal role among black students because black females face smaller deficits in academic performance and theoretical interests than females from other ethnic groups. Of the four groups, academic ability beliefs play the smallest part in explaining the engineering gap among Asian students.

Figure 3.2 segments the sample by SAT math scores instead of ethnicity. The share of explained variation increases in SAT math performance. Among low scorers, beliefs about mathematical ability account for 4-5% of cross-gender differences in engineering intentions, compared to over 7% among the highest SAT math scorers. Moreover, math ability beliefs comprise the majority of explanatory power

contributed by all academic ability beliefs. This suggests that math anxiety and confidence is a particularly salient feature of the decision to enter engineering. Similar to ability beliefs, the magnitude of the professional goals factor is steadily increasing in ability. 3% of the gap among low math achievers are explained by gender differences in professional goals, compared to over 10% among the highest achievers. As individuals move up in SAT scores, differences in men and women's stated importance of making a scientific or artistic contribution widen. The correlation between engineering intentions and goals such as making theoretical contributions in science also strengthens.

3.5.3 Family background and structure

Yet another source of variation is parental influence and family context. Parental expectations for academic achievement and occupational choices may differ by the child's gender. For instance, parents on average have lower academic expectations for daughters (Fryer and Levitt, 2010). The extent of this differential can depend on, among other factors, the nature of each parent's professional experiences. Gender-based differences may be inconsequential in families where mothers work in math-related occupations and serve as professional role models for daughters. To evaluate this possibility, I condition the models on mothers' and fathers' occupational categories in the CIRP Freshmen Survey. Table 3.8 shows that accounting for parental occupations actually increases the engineering gender gap by 0.2%. Results point to the limited scope that professional experiences have for bridging differences in engineering orientation among sons and daughters in the previous two decades.

Parental occupation belongs to a much broader set of mechanisms that can shape gender-based norms with consequences for STEM orientation. In addition to parents, teachers and peers can also establish and reinforce gender-based expectations of success in math- and science-oriented subjects, which in turn can lead boys and girls

to diverge in their human capital investments. While it is difficult to isolate each individual influence, one means of gauging their cumulative effect is to compare the academic trajectories of children from different family compositions. I use a sample of same- and opposite-sex twins in North Carolina, under the plausible assumption that the gender composition is as good as random across families with twins. What differentiates these families is the salience of gender norms in the context of opposite-sex twins. Divergent engineering orientation in this context, then, can be explained by differential gender role-based socialization during childhood and adolescence.

Table C4 traces the twins' academic outcomes from elementary through secondary schooling. The top panel juxtaposes engineering orientation and mean test scores of female (or male) students from opposite sex twins with that of same-sex twins, conditional on having taken the SATs. The bottom panel uses the full sample of 9569 twin pairs. The share of college-aspiring females from both types of family structures aiming to major in engineering is statistically indistinguishable at 2%. On the other hand, males from opposite-sex twin pairs are significantly more likely to indicate an interest in engineering. 19% named engineering as their preferred major compared to 15% of same-sex male twins. Expressed in regression form, the unadjusted gap in Table 3.9 is 3.8 percentage points.²¹

The first candidate explanation for these patterns is academic preparation. Males in opposite-sex pairs may prefer engineering because they are better qualified in math than male-male pairs due to differential selection into the SAT sample or investment over time in this skill. In fact, the opposite is true - males in opposite-sex twin pairs are more interested in engineering despite having significantly lower SAT math scores. When conditioning on SAT scores, males in opposite-sex twin pairs are now

²¹ A specification of the following form is run separately for males and females:

$$Y_{ih} = \gamma + \delta OppSex_h + \rho X_{ih} + \epsilon_{ih}$$

Engineering orientation for individual i in household h depends on family structure, in this case whether the individual is part of an opposite-sex twin pair, and a vector of individual attributes such as standardized test scores (X_{ih}).

Table 3.9: Engineering orientation among twin sample

	Females		Males	
Opposite-sex twin	-0.006 (0.005)	-0.006 (0.005)	0.038*** (0.013)	0.042*** (0.013)
Race and cohort FE		Yes		Yes
SAT scores		Yes		Yes
Observations	4421	4421	3571	3571
R^2	0.000	0.064	0.003	0.059

Notes: Female sample comprises both females in opposite-sex twins and same-sex female twins. Male sample comprises both males in opposite-sex twins and same-sex male twins. Models are augmented with indicators for SAT math and verbal scores for each 10-point bin, grade at the latest SAT test administration, and cohort fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2 percentage points more likely to choose an engineering major (Table 3.9). The second possibility is that males in opposite-sex pairs are acting on their relative math advantage, since they have higher math scores than their female twins. In results not shown, I revise the control group to include only the twin in male-male pairs with higher math performance. Males from opposite-sex pairs are still 2.6 percentage points more likely to choose engineering, despite a larger SAT math deficit than before. Relative advantage, then, also falls short as an explanation.

Summary statistics from Table C4 chronicle individuals' academic achievement, attrition, SES, and computer use during elementary and middle school for the full sample. Twins from same-sex and opposite-sex pairs are statistically indistinguishable across almost all measures, except for different likelihoods of choosing engineering and family structure. This suggests that the salience of gender and accompanying gender-based roles and expectations play an important role in the development of engineering orientation. One interesting pattern in the summary statistics is that males in opposite-sex pairs are more likely to use their computer at home at least once or twice per week in middle school. Assuming that families of these twins offer similar home access to computers, the evidence is consistent with males in opposite-

sex pairs investing more heavily in computer skills. As shown previously, computer activities and course-taking in high school are then predictive of future engineering orientation.

3.6 Conclusion

Female under-representation in STEM fields such as engineering is a long-standing phenomenon. With the ascendance of women in postsecondary enrollment and completion, researchers are focusing on differential take-up and attrition in college and beyond as a key explanation for the STEM gender gap (Preston, 2004; Hunt, 2016). This paper argues that in the case of engineering, doing so obscures the sizable gender gap already apparent by 9th grade. Administrative and survey datasets document a broadening of the gap during high school. Meanwhile, women do not exit engineering at disproportionately higher rates in North Carolina's public universities.

Once established, an engineering gender gap can be difficult to narrow. Engineering majors require pre-college exposure to advanced STEM curricula and a highly regimented sequence of courses during college, so it is especially costly to switch onto this track. Since a large disparity is already apparent by grade 9, programs that promote general STEM interest and exposure to engineering and technology-related skills during middle or even elementary school should be an important focus of efforts to bridge existing disparities. Doing so requires a better understanding of attributes that lead men and women to pursue and persist in engineering. I decompose the gender gap into several explanatory factors, including differences in academic performance, beliefs in ability, other-regarding values, and professional goals in the arts and sciences. SAT scores and high school GPA explain between 5 to 9% of the gap across three different datasets. High school sorting plays a negligible role, although high school curriculum choice, credits earned, and participation of extracurricular activities do matter for future academic tracks.

Beliefs in academic ability explain 8% of gender disparities in engineering, conditional on objective measures of academic ability. The majority of the result derives from differential beliefs in mathematical ability. Expectations about academic environment may be moderating this effect, as young women often hold themselves to higher standards in male-dominated fields like mathematics and engineering (Hill, Corbett, and St. Rose, 2010). The belief that they must be exceptionally good to succeed in engineering can reinforce the confidence gap and further exacerbate female under-representation. One means of closing the engineering gap is to bridge gender differences in ability beliefs that are not justified by actual performance. These confidence deficits have diverse social origins ranging from teachers' stereotypical biases to parental evaluations of competency (Herbert and Stipek, 2005; Gunderson et al., 2011). Professional development programs that raise awareness of implicit biases among educators and devise communications designed to ease math-related anxieties among females may increase self-confidence and participation. A related approach is shifting away from the view that math ability is an innate fixed trait to one that underscores the malleability of this skill via cultivation and effort (Dweck, 2008).

Remaining factors capture dimensions of individual preferences. Women are more likely to assign greater importance to values that correlate with lower engineering participation, including helping others in difficulty, influencing social values, and participating in community service work, such that these other-regarding values collectively explain 8% of the gap. Professional goals in the arts and science explain a further 7%, with men disproportionately aiming to make a theoretical contribution to science. Since female students foresee engineering careers as contributing little to the social good, programs that emphasize the prosocial over abstract dimensions of engineering can be another avenue for closing the gap.

Even when all observable influences are tallied, over two-thirds of the gender

participation gap in engineering remain. This residual is often conveniently characterized as “other preferences,” but it is worth dwelling on its contents and subsequent implications for policy. If male and female students are given equal opportunities to pursue an engineering career throughout their life course, then we can attribute divergent choices to differential tastes that may well fall outside the purview of policy interventions. However, empirical evidence in this paper suggests that major and occupational orientation is shaped by social forces working to instill gender-stereotypical preferences. More salient gender roles in opposite-sex relative to same-sex twin pairs manifest in differential engineering orientation. The mechanism can involve teacher and parental expectations and other social cues which in turn shape men and women’s experiences in math- and science-intensive curriculum, preferences for engineering coursework, and labor market expectations. The dynamic nature of preference formation suggests that earlier interventions aimed at tempering gender norms and equalizing opportunities for both boys and girls can pay dividends in future female STEM participation.

A Class of Their Own: Explaining High School Attainment Gaps by Income, Race, and Gender

Disparities in educational attainment by gender, race, or socioeconomic background carry significant implications for future economic well-being. Groups that lag in high school graduation fail to capture substantive college wage premiums and follow substantially different labor market trajectories (Grogger and Eide, 1995; Card, 1999). As such, a thorough understanding of factors underlying attainment decisions is of eminent concern to researchers and policymakers.

This paper enriches existing research on educational attainment by documenting gaps at the three-way intersection of socioeconomic status, race, and gender. In this way we are able to explore gaps in educational outcomes along the socioeconomic gradient for a given ethnicity and gender pairing, for example distinguishing between the challenges faced by disadvantaged white males from disadvantaged African American males. Nuanced interactions between these attributes underscore the need for more tailored interventions for each subgroup.

Our models examine several possible explanations for cross-group disparities: high school characteristics, cognitive skills, and non-cognitive skills. Existing research offers mixed evidence on the contribution of school quality to equity gaps (Cook and Evans, 2000; Fryer and Levitt, 2004; Autor et al., 2016).¹ Proxies for cognitive skills, on the other hand, have been established as relevant determinants for both educational attainment and major choice (Paglin and Rufolo, 1990b; Weinberger, 1999; Heckman and Krueger, 2005). In recent years there has been a shift from an unidimensional focus on cognitive abilities towards the contributions of both cognitive and non-cognitive factors in ameliorating disadvantage (Heckman, Stixrud, and Urzua, 2006a; Cunha, Heckman, and Schennach, 2010).

Operationalizing this type of analysis is challenging because it requires detailed educational attainment records, as well as data from high school and earlier that can be used to identify school characteristics, cognitive skills, and non-cognitive skills. We also require a sufficiently large sample that enables the exploration of outcomes for more granular classifications of students. We overcome these challenges by using administrative data from the North Carolina Education Research Data Center (NCERDC). The sociodemographic information contained in the NCERDC allows us to segment students by gender, ethnicity, and socioeconomic status (SES).² Rich data on student standardized test scores, course-level grades, absences, homework effort, and high school attended enable the construction of covariates corresponding to each

¹ The majority of evidence focuses on racial gaps in achievement. The contribution of school quality to the narrowing black-white performance gap is less than 12% for math in one study (Cook and Evans, 2000). While it plays a more prominent role in explaining racial gaps for students entering kindergarten and first grade, the role of high school quality diminishes in a re-evaluation of the same students several years later (Fryer and Levitt, 2004, 2006). Evidence on the role of school quality in determining gender gaps is more limited, although what exists suggests the genders respond differentially to school inputs. Autor et al. (2016) finds varying gradients between school quality and educational and behavioral outcomes by gender. Boys benefit more from high quality middle schools, such that the usual female advantage in outcomes is declining in school quality.

² We focus on 8 student subgroups, representing the interactions between gender, white or African American, and free and reduced lunch eligibility.

of the three explanatory accounts. Data on school exit, graduation, and dropouts complete the dataset. While NCERDC forms the core sample, we supplement our analyses using the nationally representative National Longitudinal Survey of Youth 1997 (NLSY97) to ensure that documented educational gaps are valid outside of the North Carolina context.

A challenge in analyzing the contribution of cognitive and non-cognitive factors to academic outcomes is that observed measures (e.g. test scores) are noisy proxies for underlying skills. Furthermore, they are themselves the products of a dynamic process in which familial background and school inputs such as teachers exert their influences on the student. To address measurement error and endogeneity concerns, we specify a sequential model of educational attainment that is consistent with the factor analysis literature, in particular Cameron and Heckman (2001) and Aucejo (2015). We identify two latent skills that are mutually independent, and fixed by the time students make decisions about their transition to higher levels of secondary education and college.³ These latent cognitive and non-cognitive skills are allowed to be correlated with observed demographic characteristics, which enables us to analyze the effects of each factor by gender, race, and SES.⁴ The measurement system also allows observed cognitive and non-cognitive proxies to vary by other influences such

³ Restricting each factor to be constant after students enter school is a common assumption in the literature, as documented in Aucejo (2015).

⁴ The literature finds that non-cognitive factors such as social skills and effort are important determinants of graduation from high school and college, and a student's ultimate level of educational attainment after high school. Heckman, Stixrud, and Urzua (2006b) show that the effect of non-cognitive skills on the probability of graduating from high school and continuing to college are comparable in magnitude to that of cognitive skills. Factor models have also demonstrated the role that differences in non-cognitive abilities play in educational attainment across gender and race. Urzúa (2008) uses The National Longitudinal Survey of Youth, 1979 (NLSY79) to show that unobserved non-cognitive ability is a significant determinant of black-white gaps in educational attainment. Aucejo (2015) extends this analysis using data from the NLSY97 to show both factors play a significant role in explaining the gender gap in educational attainment both within and across races. In particular, Aucejo finds that while cognitive factors are a significant determinant of transitions between education levels, non-cognitive factors primarily explain the gender gap in educational attainment, particularly for African Americans.

as schooling inputs. This has the benefit of addressing measurement error. We are also able to control for the dynamic selection process that drives students' sequential educational choices. Doing so enables us to separately identify both factors, their distributions across demographic groups, and the importance of each factor at each stage of the model.

Descriptive statistics reveal striking and persistent differences in educational attainment across and within socioeconomic groups. 4-year graduation rates conditional on non-missing schooling covariates are 90-94% across all high SES groups, relative to 72-82% among low SES students.⁵ Economically disadvantaged black females have the highest graduation rates among low SES groups, despite large standardized test score deficits relative to their white female peers. This phenomenon is concentrated among the most economically deprived, and cannot be explained by income differences.⁶ As such, the narrative of black-white achievement gaps do not easily translate to similar disparities in high school attainment. These results also underscore the scale of challenges faced by poorer white students of both genders, since their test score advantages are not converting to higher attainment.

The complex landscape of graduation gaps underscores the first significant contribution of this paper. Categorizing students at the intersection of gender, race, and economic disadvantage permits a more tailored understanding of academic success and challenges than previous US-based research. While studies on racial or gender gaps are prevalent, socioeconomic status is usually incorporated as an additional co-

⁵ Our outcome variable is graduating within 4 years of first enrolling in 9th grade while being "on-track." We define on-track as on-time persistence - being observed in grade 9/10/11/12 during year 1/2/3/4, respectively. The distinction between low and high SES comes initially from free and reduced lunch eligibility, although we also run robustness checks using parental education as an alternative proxy for SES and patterns of intergroup disparities are similar.

⁶ Poor white females in North Carolina fall behind in graduation rates despite residing in neighborhoods with higher median incomes. The same trends are borne out using nationally representative longitudinal surveys, where poorer white females also hold an income advantage over their black peers.

variate. Studies on the black-white achievement gap find that socioeconomic factors account for anywhere between a quarter up to 85% of disparities at various points of the educational experience (Fryer and Levitt, 2004, 2006; Murnane et al., 2006; Reardon and Robinson, 2008; Clotfelter, Ladd, and Vigdor, 2009).⁷ A small selection of papers that explore the full interaction effects of race, gender, and SES acknowledges the importance of distinguishing between race and SES (Strand, 2014). Using race to confer disadvantage obscures the narratives of underprivileged white students and more advantaged black students, as the socioeconomic spectrum can be broad for a given racial group.

To explain these striking patterns, we consider a rich set of measures capturing high school characteristics, cognitive, and non-cognitive skills. Reduced form results accounting for non-cognitive skills alone fully explain attainment gaps between advantaged white females and all black student subgroups with the exception of poorer black males. However, the sum of these explanatory accounts is far from adequate in explaining the graduation rate gap between advantaged and disadvantaged white student subgroups. The role of schooling and skill inputs thus bifurcates by race - raising non-cognitive skills among black students to the same levels as advantaged white peers overcomes the majority of high school attainment disparities. Meanwhile, economic deprivation among white students materialize in ways beyond schooling and skill inputs to put downward pressure on graduation rates.

Another contribution of this paper is the specification of a structural model that

⁷ Fryer and Levitt (2004) is relatively unique in the literature in finding that the vast majority of the black-white achievement gap can be accounted for using a set of socioeconomic controls. While they observe this result for kindergarteners, the explanatory power of SES diminishes significantly by the time students reach third grade (Fryer and Levitt, 2006). Murnane et al. (2006) documents that one-third of the math achievement gap during kindergarten and third grade is eliminated after accounting for family covariates such as SES background. Clotfelter, Ladd, and Vigdor (2009) find that a vector of socioeconomic factors including free and reduced lunch eligibility, parental education, gender, age, and district and region type explain approximately one-third of black-white gaps in math and reading from grades 3-8.

permits us to distinguish the effects of each factor at every educational stage. This implies identifying the relative contributions of cognitive and non-cognitive skills. We find evidence that non-cognitive skills exert greater influence on persistence for every year of schooling. While both factors account for almost all on-track persistence disparities in the initial three years, their roles diminish during the last year of high school. Moreover, we use estimates of each latent factor’s distribution to simulate how student choices vary when their cognitive and non-cognitive abilities are equalized to peers.

The remainder of the paper proceeds as follows. Section 4.1 outlines the sample construction using North Carolina administrative data, as well as our use of nationally representative longitudinal surveys. Section 4.2 summarizes the extent of disparities across student subgroups and presents a series of linear probability models to provide suggestive evidence on the contribution of each account. We shift from OLS evidence to a description of the factor model in Section 4.3. Section 4.4 presents parameter estimates from the structural model for comparison with previous results. We conclude with a discussion of the findings’ implications in Section 4.5.

4.1 Data

We use administrative data from the state of North Carolina as the basis of our estimation sample. The resulting student population is significantly larger than samples constructed from longitudinal surveys such as the National Educational Longitudinal Study (NELS) (Jacob, 2002) and the National Longitudinal Survey of Youth of 1997 (NLSY97) (Aucejo, 2015). To ensure the external validity of documented attainment gaps, we construct the same student subgroups by race, gender, and income using the nationally representative NLSY97.

4.1.1 NCERDC

The primary data source is statewide administrative data for public and charter school students from the North Carolina Education Research Data Center (NCERDC). Data elements include students' standardized test scores, attendance and delinquency, high school transcripts, and school exit files. Our main dataset tracks 9th grade students in 2008 - 2011 for four years to the presumed end of their high school careers. Covariates broadly fall into three categories: student socio-demographic information, high school characteristics, and measures of cognitive and non-cognitive skills. We discuss each of these covariates and outcomes in turn.

Time-invariant demographic data on race and gender are taken from student roster files from 2008 - 2011.⁸ Non-Hispanic white and African American students comprise nearly 90% of 9th grade public school students in North Carolina during the timespan of our study. We focus on the two largest racial groups in this study, although the conceptual framework can be readily extended to examine educational outcomes for the Hispanic, American Indian, and Asian student populations. Socioeconomic status is captured via the dichotomous variable of free and reduced lunch (FRL) eligibility. Children qualify for free lunch under the federally subsidized National School Lunch program if they reside in a household with incomes at or below 130% of the federal poverty level, and they qualify for reduced lunch if household income is between 130% and 185% of the stated poverty level.⁹ Notably, SES status can fluctuate across time as parental incomes change or due to administrative errors. To ensure we are distinguishing families that are persistently advantaged or disadvantaged from families experiencing income shocks, we take a longitudinal view

⁸ We use demographic information from Masterbuild files and supplement with socioeconomic data from End-of-Grade and End-of-Course files. An encrypted student ID links all NCERDC files to the base sample.

⁹ In North Carolina, children from households that are eligible for receiving Supplemental Nutrition Assistance Program (SNAP) or Cash Assistance (CA) are automatically eligible for free meals.

of SES status and compile free and reduced lunch eligibility for students going as far back as third grade.¹⁰ We categorize students who were never tagged as FRL eligible as high SES, and those who were ever tagged as low SES. The latter is further distinguished between those who were consistently eligible for the National School Lunch program over the course of their public elementary and middle school studies and those who were ineligible at least once.

We interact the three binary categories of female, black, and SES status to produce eight mutually exclusive student subgroups that reflect nuanced relationships between race and family background. Economically disadvantaged black students may face systemically different sets of challenges and constraints than their white peers from equally under-privileged households. Our analyses examines the extent to which these differences can manifest in varying educational decisions at the end of secondary schooling.

The second set of variables comprises high school attributes. School-level data in the NCERDC provide information on the type, locale, and student composition. We include a variable on magnet school status, and use locale coding that categorizes a school's mailing address as urban, suburban, or rural. We designate urban schools as those located in cities with a population of at least 250,000, and rural schools as those in locations with fewer than 2,500 people or ZIP codes designated as rural by the Census Bureau. Lastly, we construct the shares of 9th grade students in the school who belong to each of eight categories in the final analytic sample.

Third, we compile cognitive and non-cognitive measures from multiple source files. Cognitive skills are approximated using students' End-of-Grade (EOG) math and reading scores from 8th grade, the last year for which statewide standardized testing occurs at the grade level, and End-of-Course (EOC) reading grades from 9th

¹⁰ Table D1 shows the source of FRL information for each grade-year combination.

grade.¹¹ To ensure that our sample captures the results of both retained and non-retained students, we take scores from the first time the student took the 8th grade EOG and 9th grade EOC exams across the 3-year period of t , $t - 1$ and $t - 2$. We normalize scores using the statewide population of test-taking students in a given grade and year to ensure comparability across schools and time.¹²

Non-cognitive measures include GPA, absences in 9th grade, and time spent on homework during 8th grade. We construct grade-specific weighted GPAs using credits earned and course-level grades from student transcripts.¹³ Along the disciplinary dimension, we use the number of days absent during 9th grade.¹⁴ Lastly, we proxy for student effort using an 8th grade data element summarizing the number of hours students spend weekly on homework. We create indicators for whether the student is below, at, or above the class median, as well as a separate variable for students who refuse to complete their assigned homework.¹⁵

Our outcome measures capture students' progression through the end of secondary schooling. The dependent variable derives from three sets of files: graduation, drop out, and school exit.¹⁶ We allow a 4-year window from the time of 9th grade enrollment to exit. Specifically, a 9th grader in 2008 is matched to 2011 graduation

¹¹ We restrict to 9th grade reading scores only because the same English standardized test is taken by all 9th grade students, while there is no corresponding prescribed course for 9th grade math.

¹² Test scores are normalized to mean 0 with a standard deviation of 1.

¹³ Honors and AP courses, for instance, are allocated more credits than regular coursework. As a result the 9th grade weighted GPA reflects both course levels and performance.

¹⁴ We forgo the use of in- and out-of-school suspensions, since in the absence of additional behavioral measures we cannot assess whether differences in disciplinary sanctions by race or gender are due to actual infractions or the mediating influence of bias (Gregory, Skiba, and Noguera, 2010; Kinsler, 2011).

¹⁵ Our choice of non-cognitive measures are consistent with variables used in previous studies on college enrollment. Jacob (2002) proxies for non-cognitive skills with middle school grades, hours spent on homework during 9th grade, a composite measure of disciplinary incidents, and an indicator for elementary school retention. Aucejo (2015) uses 8th grade GPA, retentions and suspensions between grades 1 and 8, involvement in fights, and precocious sex.

¹⁶ School exit files provide the rationale for leaving, which includes graduation and dropping out, along with other reasons such as transfers to home schooling and private schools.

files and 2008-2011 dropout and school exit data to ensure a comprehensive view of their manner of exit.¹⁷ We exclude those who exit for reasons that are exempt from graduation rate calculations.¹⁸ Students are designated as on-track in year 2 if they have not dropped out or exited for other reasons, and are observed in grade 10 by year 2. The final outcome, on-track by year 4, requires students to be observed in and graduate by grade 12.

The final sample spans 2008 - 2011 9th grade cohorts. We retain unique individual observations by only including the first time an individual was observed in a 9th grade cohort. Moreover, we keep student records with non-missing high school characteristics, 8th grade EOG scores, 9th grade EOC reading scores, hours spent on homework in 8th grade, 9th grade absences and GPA. The final sample comprises 274,555 individuals.

4.1.2 NLSY97

We supplement NCERDC-based analyses with data from the National Longitudinal Survey of Youth 1997 (NLSY97) to ensure that documented academic disparities are not specific to North Carolina or the particular period covered. 8,984 individuals between the ages of 12 to 18 were interviewed in 1997. NLSY97 tracks them through the 2013-2014 year, when respondents were 28 to 34 years old. We construct 8 student groups using interactions between race, gender, and socioeconomic status. Although no data element exists on free and reduced lunch status, we can approximate eligibility by denoting respondents with family incomes at or below 185% of

¹⁷ We cross-check graduation and dropout outcomes with reasons given in school exit data. For instance, we code students as having graduated or dropped out if they are marked as such in either the graduation/dropout or exit files.

¹⁸ These primarily comprise students who exit from the public school system and therefore can no longer be observed via graduation files. NCERDC exempts students who exit for the following reasons: leaving for another school in the system, leaving to another system, leaving for a different state, death, visiting student status, transfers to within-state private schools and home schools, unconfirmed transfers including detention center, and students who do not in fact belong in the cohort. All dropped observations comprise slightly less than 20% of the sample.

the federal poverty guideline. The dependent variable takes on a value of 1 if the student indicated 12th grade or above as their highest level of schooling attained. The final sample of 4,936 respondents comprises all students who completed at least middle school.

4.2 Reduced form evidence

We begin by presenting reduced form evidence on the observed gaps in educational outcomes across racial, gender, and socioeconomic groups. Evidence on the magnitude of disparities in the North Carolina context is corroborated with a nationally representative sample from the NLSY97. We then shift from unconditional means to a series of linear probability models that successively account for the roles of high school characteristics, cognitive skills, and non-cognitive skills.

4.2.1 Attainment disparities by subgroup

Table 4.1 summarizes outcomes and covariates across the eight subgroups. The first four columns show summary statistics for high SES students, defined as those who were never observed as FRL-eligible, while the final four show the same for more disadvantaged populations who were eligible at least once. Across all outcomes, we observe distinct differences along socioeconomic lines. 90-94% of high SES students exhibit on-track persistence and graduate high school in 4 years relative to 72-82% among low SES students.

Several observations emerge from these sample statistics. The first is the importance of examining patterns at the intersection of race and SES. Our approach explicitly studies the academic trajectories of poor white students and advantaged black students of both genders, rather than using minority status to confer economic disadvantage. Within each racial group, outcomes bifurcate by SES. Poor white students are at least 17 percentage points less likely to graduate compared to their

Table 4.1: NCERDC - Summary statistics

	High SES				Low SES			
	Female		Male		Female		Male	
	White	Black	White	Black	White	Black	White	Black
On track by Year 2	0.984	0.977	0.970	0.944	0.900	0.888	0.854	0.823
On track by Year 3	0.966	0.962	0.951	0.930	0.836	0.850	0.791	0.772
On track by Year 4 and graduated in 4	0.935	0.937	0.915	0.900	0.767	0.818	0.721	0.735
Skill proxies: Cognitive								
8th grade math EOG z-scores	0.555	0.040	0.542	-0.096	-0.027	-0.439	-0.040	-0.535
8th grade reading EOG z-scores	0.585	0.131	0.488	-0.060	0.057	-0.417	-0.024	-0.561
9th grade reading EOC z-scores	0.610	0.248	0.411	0.018	0.138	-0.207	-0.066	-0.398
Skill proxies: Non-cognitive								
9th grade GPA	3.511	3.058	3.195	2.611	2.750	2.382	2.392	2.005
Total absences in 9th grade	5.346	4.083	5.347	4.114	9.402	7.960	9.552	8.124
HW time below 8th grade class median	0.203	0.219	0.271	0.255	0.236	0.263	0.313	0.305
HW time at 8th grade class median	0.462	0.512	0.452	0.507	0.490	0.509	0.459	0.485
HW time above 8th grade class median	0.335	0.269	0.277	0.238	0.274	0.229	0.228	0.209
Has homework, but refuses to do it	0.004	0.005	0.015	0.010	0.011	0.011	0.038	0.023
High school characteristics								
Magnet high school	0.045	0.145	0.045	0.127	0.013	0.097	0.013	0.091
High school in city with 250K+ residents	0.173	0.304	0.173	0.295	0.058	0.200	0.061	0.200
High school in rural area	0.546	0.372	0.544	0.378	0.633	0.404	0.635	0.411
Share of high SES white females in 9th grade	0.233	0.152	0.225	0.154	0.187	0.118	0.185	0.117
Share of high SES black females in 9th grade	0.022	0.054	0.022	0.046	0.016	0.038	0.016	0.036
Share of high SES white males in 9th grade	0.234	0.157	0.238	0.162	0.194	0.123	0.195	0.125
Share of high SES black males in 9th grade	0.023	0.046	0.023	0.052	0.015	0.037	0.016	0.038
Share of low SES white females in 9th grade	0.100	0.057	0.099	0.055	0.150	0.066	0.143	0.067
Share of low SES black females in 9th grade	0.074	0.161	0.075	0.154	0.078	0.201	0.078	0.189
Share of low SES white males in 9th grade	0.106	0.060	0.106	0.060	0.152	0.070	0.159	0.072
Share of low SES black males in 9th grade	0.072	0.148	0.074	0.152	0.077	0.186	0.079	0.197
Observations	61,824	8,611	63,179	8,173	31,407	37,498	30,756	33,107

Note: The share of students by race are taken from the characteristics of the student's school in 9th grade.

more privileged white peers, with similar or smaller differences by SES among black students. Some of these findings are surprising in light of observable test scores. For one, black females have comparable or higher graduation rates than their white female peers across the socioeconomic spectrum, despite sizable standardized test score deficits.¹⁹ The presence of a black-white achievement gap, then, does not translate readily to a racial attainment gap once conditioning on SES.

¹⁹ Black students demonstrate significantly lower achievement scores in both math and reading. Relative to high SES white students, advantaged black students score 0.6 standard deviations lower in 8th grade math, while disadvantaged black students score approximately 1 standard deviation lower. The magnitudes of these black-white achievement gaps are consistent with the 0.8 standard deviation gap found by Clotfelter, Ladd, and Vigdor (2009) among eighth graders. On non-cognitive measures, black students exhibit a similar disparity in GPA but are often less likely to accumulate absences.

Second, gender gaps are larger among African American students than whites, and this racial differential is increasing in economic adversity. This provides a more nuanced view than the general finding that females out-perform males in educational attainment. Among well-to-do black students, women have a 4 percentage point graduation advantage while they are ahead by nearly 9 percentage points among disadvantaged populations. The corresponding figures for high and low SES white females are 2 and 5 percentage points. These empirical regularities are traceable back to middle school, when black females are developing an academic edge in EOG and EOC test scores over their male peers that are not as salient among white females. These findings prompt questions on how academic experiences, expectations, and challenges may differ across gender lines for black students.

The final observation involves the under-performance of poorer white students across outcomes. Poorer white males rival black males in low high school graduation rates. This is particularly striking given that disadvantaged white females score 0.4 standard deviations higher than disadvantaged black females in both 8th grade reading and math, while the corresponding advantage is closer to 0.5 standard deviations among low-SES male students. Under-privileged white students also perform better in 9th grade GPA and are more likely to spend at least 3 hours per week on homework relative to African American students in the same socioeconomic bracket. Yet these advantages in academic track records are not translating to superior attainment outcomes.

4.2.2 Alternative explanations

To ensure that these attainment patterns are robust across contexts, we investigate the feasibility of several alternative explanations. For example, is the under-performance of disadvantaged white students the result of ability-based selection into the sample that differs substantially by race? Is operationalizing socioeconomic sta-

tus using free and reduced lunch eligibility crucial for the results, and how much do they change when including more detailed family income controls or using parental education as a proxy for socioeconomic status? Are these results only valid for North Carolina because they depend on features specific to the state?

Selection into sample

The analytic sample used in Table 4.1 only includes observations with non-missing high school, cognitive, and non-cognitive measures. To ensure that results are not driven by differential selection into the sample, we juxtapose them with unconditional graduation rates. Table D2 begins with an unconditional sample of all 9th grade students in 2008-2011 and exclude students with missing data in select variables. Nearly 20% of all students are dropped from the full sample due to missing data. The raw on-track persistence 4-year graduation rate for the analytic sample is 79%. There is some evidence of positive selection by ability into EOG and EOC test-taking, as the rate increases to 85% among test-takers..

Importantly, key patterns in cross-group attainment disparities remain. Disadvantaged white females are also 5 percentage points less likely to graduate relative to disadvantaged black females in the unconditional sample. Among economically deprived males, both races share a graduation rate of 64%, which increases to 72% among whites and 74% among blacks in the analytic sample. Further evidence documents changes in the sample's socio-demographic composition when dropping students with missing data. Table D3 shows the share of each student subgroup under the unconditional and final analytic samples. Shares are very stable for low SES females. Low SES males across racial groups are less likely to be represented in the final sample, with black students exhibiting the highest rates of attrition. While selection can cause adjusted graduation rates of disadvantaged black males to become overstated relative to unconditional rates, these black students are still performing

no worse than their white peers in the unconditional sample. As such, the under-performance of poorer white students are not the consequence of selection into the sample.

Conclusions so far derive from results aggregated across time. Figure D1 plots separate attainment rates for the four graduation cohorts of 2011-2014. Black-white disparities among low SES students are notably consistent across cohorts for both the unconditional and analytic samples. Disadvantaged black females maintain a stable 5 percentage point attainment advantage over their white peers, and similar outcomes are shown across cohorts among disadvantaged males.

Alternative SES measures

The present categorization of low SES includes students who were ever eligible for free and reduced lunch between 8th and 10th grades. This groups students experiencing transitory negative income shocks around the 185% threshold with those experiencing more permanent shocks and students misclassified as low income. The question then arises of whether the empirical regularities we observe are maintained at more fine-grained or alternative definitions of socioeconomic status. We divide the ‘Ever FRL’ population further into those who were consistently FRL eligible during their elementary and middle school careers (Bottom SES) versus those who had mixed eligibility records over time (Middle SES). In addition, we consider facets of socioeconomic status beyond household income. In the absence of parental occupation data, we use a student-reported variable on parental education over the same time span. The corresponding high, middle, and bottom SES categories correspond to at least some college coursework, high school diploma, and high school dropout. Table 4.2 shows the respective shares and average 4-year graduation rates across 12 subgroups for these two alternative definitions.

The segmentation of disadvantaged students into middle and bottom SES cat-

egories across both SES measures illuminates the types of families for whom the white-black attainment gap is starkest. Attainment levels for black females are comparable in both middle and bottom SES categories using free and reduced lunch eligibility, suggesting that this variable is too coarse to distinguish between families near the eligibility threshold from those who are severely disadvantaged.²⁰ Moreover, disadvantaged white females in both middle and bottom SES categories have lower attainment of 4-8 percentage points relative to black peers. Switching to parental education as an alternative SES measures reveals a larger attainment advantage among black females in the bottom SES group. 62% of white females whose parent is a high school dropout graduate in 4 years, compared to 67% of black females. The analogous statistics for middle SES white and black students are 81% and 82%, respectively. The under-performance of low SES white students in Table 4.1 is therefore most pronounced among the lowest socioeconomic groups. White families with little formal education are more likely to be economically deprived, and this manifests in particularly low attainment records among females.

Analyses using alternative SES measures still leave the possibility that poor white students under-perform because they come from families that are more economically impoverished than their black peers in the same SES category. Table 4.2 provides evidence to the contrary. Within each SES group, white students reside in Census block groups with higher median incomes. The findings are even more striking because poor white students fare poorly in graduation rates in spite of their relative economic advantage.

²⁰ We do not separate FRL further into eligibility for free vs. reduced lunch because this level of granularity is missing from administrative data in select years.

Table 4.2: NCERDC - Summary statistics for alternative SES definitions

	FRL^a			Parental education^b		
	Share	Grad rate	BG income ^c	Share	Grad rate	BG income
High SES						
White female	0.23	0.93	63780	0.21	0.93	63452
Black female	0.03	0.94	55776	0.06	0.92	48410
White male	0.23	0.92	63925	0.21	0.91	63702
Black male	0.03	0.90	56391	0.05	0.87	48984
Middle SES						
White female	0.07	0.78	46836	0.12	0.81	47078
Black female	0.05	0.82	42031	0.10	0.82	39137
White male	0.07	0.73	47393	0.12	0.76	47338
Black male	0.05	0.73	41841	0.09	0.73	39459
Bottom SES						
White female	0.05	0.74	44457	0.01	0.62	42851
Black female	0.08	0.82	38219	0.01	0.67	35881
White male	0.05	0.70	44295	0.01	0.58	42864
Black male	0.07	0.74	38666	0.01	0.57	35474
Observations	274555	274555	274555	252044	252044	252044

Note: Sample includes all students observed in 9th grade between 2008-2011. Indicators for race/gender/SES categories are defined relative to high-SES white females.

^a High SES is defined as students who were never FRL eligible. Middle SES is defined as students who were both eligible and ineligible during the period they were observed in the dataset. Bottom SES is defined as students who were always eligible for FRL.

^b High SES corresponds to students with parents attaining some college coursework or above. Middle SES refers to parents who are high school graduates, while bottom SES corresponds to high school dropouts.

^c Block group median income come from the American Community Survey 2013 5-year sample.

Nationally representative samples

To ensure that these patterns are not specific to the North Carolina context, we turn to the national sample available in NLSY97. Table 4.3 distinguishes between the same 8 student groups using gender, race, and predicted National School Lunch Program eligibility constructed from household income relative to the federal poverty line. The graduation rate, defined by whether the student ever completed 12th grade, is 82% for the full sample. Subgroups maintain consistent achievement patterns relative to each other as in North Carolina, with the exception of poor black males. Only 56% graduate from high school in the national sample, compared to 74% in North Carolina. The higher persistence of the latter may be partially explained by

the positively selected sample containing non-missing data. A notable observation is that poorer white females exhibit lower graduation rates than poorer black females, despite their relative economic advantage.

Table 4.3: NLSY97 - Summary statistics

	High SES				Low SES			
	Female		Male		Female		Male	
	White	Black	White	Black	White	Black	White	Black
Completing 12th grade	90.1%	87.3%	86.0%	82.0%	73.3%	75.0%	66.8%	56.4%
Average percent of poverty level	444.5%	355.8%	443.2%	359.4%	108.0%	82.6%	111.4%	79.6%
Observations	444	356	443	359	108	83	111	80

Note: Base sample comprises 4,936 respondents across the 8 categories who completed at least 8th grade. A respondent has a high school degree if he or she reports completing the 12th grade. Outcomes are weights-adjusted.

4.2.3 OLS

In this section we examine the extent to which high school characteristics, cognitive, and non-cognitive skills can explain cross-subgroup disparities in educational attainment. Table 4.4 present results using linear probability models. We increasingly augment the base model with additional covariates. The left-most column in Table 4.4 regresses high school graduation on the eight subgroups and cohort fixed effects, omitting the reference group of high SES white females. Disadvantaged white males are shown to have the largest graduation deficit at 21.4 percentage points, while under-privileged black males follow closely behind with 20.0 percentage points. The addition of high school characteristics spanning magnet school status, urbanicity, and racial composition of classrooms in the second column has little impact across coefficients.

The third specification incorporates measures for non-cognitive skills, including 9th grade GPA, homework effort exerted in 8th grade, and absences. Their inclusion erases the gap between high SES white females and high-SES black subgroups. The difference relative to disadvantaged black males is reduced by over 90%, while the gap with disadvantaged black females is more than accounted for. When we include

cognitive measures in lieu of non-cognitive proxies, more of the racial graduation gap with high SES black students and less of the gap with low SES black students are explained. In contrast to the substantially reduced racial gaps, both skill sets exhibit lower explanatory power when applied to the population of white students. This reflects the relatively higher performance of under-privileged white students on academic measures such as GPA and test scores (Table 4.1).

When both cognitive and non-cognitive measures are included in Column 5, the racial gap is more than entirely accounted for in all black subgroups with the exception of disadvantaged black males. The combination of high school attributes and skill measures reduce gaps across the SES gradient for white students by approximately two-thirds or more. Thus emerges an educational attainment narrative that bifurcates by race. If efforts to raise the cognitive and non-cognitive measures of black students on par with their advantaged white female peers are successful, we would expect the majority to have greater levels of completion. However, deficits in yet unobserved measures are setting back the progress of disadvantaged whites in particular.

Additional analyses verify the stability of these results over time and across regions of varying population density. Table D4 shows that adjusted coefficients on student subgroups are of similar magnitude from 2011 - 2014. Gaps are recently narrowing relative to advantaged white females, such that accounting for high school characteristics, cognitive, and non-cognitive skills explains all racial gaps in the three most recent cohorts. Table D5 documents the prevalence of these attainment disparities across geographies, with similar patterns observed in both rural and urban neighborhoods.²¹

²¹ In analyses not show, we verify that these gaps are consistent across the wester, central, and eastern parts of North Carolina.

Table 4.4: Linear probability model: Complete 12th grade

	(1)	(2)	(3)	(4)	(5)
High SES					
Black female	0.002 (0.004)	0.012*** (0.004)	0.028*** (0.003)	0.049*** (0.004)	0.027*** (0.003)
White male	-0.020*** (0.002)	-0.021*** (0.002)	0.000 (0.002)	-0.012*** (0.002)	0.000 (0.002)
Black male	-0.035*** (0.005)	-0.029*** (0.004)	0.029*** (0.004)	0.031*** (0.004)	0.028*** (0.004)
Low SES					
White female	-0.168*** (0.004)	-0.158*** (0.003)	-0.063*** (0.003)	-0.115*** (0.003)	-0.064*** (0.003)
Black female	-0.116*** (0.005)	-0.100*** (0.005)	0.015*** (0.003)	-0.005 (0.004)	0.016*** (0.003)
White male	-0.214*** (0.005)	-0.210*** (0.004)	-0.071*** (0.003)	-0.149*** (0.004)	-0.070*** (0.003)
Black male	-0.200*** (0.006)	-0.184*** (0.006)	-0.013*** (0.004)	-0.063*** (0.004)	-0.010*** (0.004)
Additional controls					
HS characteristics ^a		Y	Y	Y	Y
Skill proxies: Non-cognitive ^b			Y		Y
Skill proxies: Cognitive ^c				Y	Y
Observations	274555	274555	274555	274555	274555
R^2	0.057	0.061	0.280	0.136	0.284

Note: Sample includes all students observed in 9th grade between 2008-2011. Indicators for race/gender/SES categories are defined relative to high-SES white females. All specifications include cohort fixed effects.

^a High school characteristics are taken from a student's 9th grade institution, and include indicators for a magnet high school and being located in a city with 250K+ residents or a rural area, and continuous measures for the share of 9th grade students in each of the 8 categories.

^b Non-cognitive skill proxies include cumulative 9th grade GPA, number of absences in 9th grade, and homework completion relative to classroom peers. GPA enters non-parametrically as a set of 20 bins.

^c Cognitive skill proxies include a student's 8th grade End-of-Grade scores in math and reading and 9th grade End-of-Course scores in reading, normalized relative to statewide test scores for the year a student took a given test. All test scores enter non-parametrically as sets of 20 bins.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.3 Factor model

The differences we observe between students from different demographic and socioeconomic groups, both in descriptive statistics and reduced form evidence, suggest skills may play a role in explaining different attainment outcomes. However, the analysis we have presented thus far suffers from two shortcomings, which are key to explaining the gaps in outcomes we observe by race, gender, and socioeconomic

status. First, these proxies measure the underlying skills of interest with error, and hence the parameters we report are likely biased estimates of the effects of skills on student outcomes. Second, our measures of non-cognitive skills, particularly in the case of GPA, are jointly determined by both cognitive and non-cognitive skills. Hence we are unable, by construction, to interpret what the relative importance of either skill is in determining student outcomes, or how the role skills play varies by race, gender, or socioeconomic status.

In this section we outline a structural factor model that overcomes these issues, and allows us to identify students' underlying cognitive and non-cognitive skills from our noisy proxies. We are then able to use these underlying skills to examine the relative roles they play in determining how students progress through and graduate from high school. Finally, we allow students' underlying skill distributions to vary by student race, gender, and socioeconomic status. This heterogeneity in skills allows us, in turn, to leverage detailed information from the NCERDC data on student demographics to explore how the effects of these skills vary for each group.

We start by first developing the measurement system for these cognitive and non-cognitive factors and how they vary by demographic characteristics. Second, we take these factor measures and incorporate them into a sequential model of secondary educational attainment. Finally, we specify the likelihood for this model and detail how we modify it for estimation.

4.3.1 Measurement system

Each student has a given vector of latent abilities $\boldsymbol{\theta} = \{\theta^C, \theta^{NC}\}$, where the elements denote cognitive and non-cognitive factors, respectively. This vector of abilities is unobserved to the econometrician. However, we assume that individuals know their levels of each factor and both latent traits are fixed by the time individuals make the schooling choices relevant to our study. Identifying the distribution of these factors

is key because they are relevant for students' choices in each stage of the model: from choosing whether to continue to the next grade in high school to graduating. We also make the simplifying assumption that these latent abilities are independent random variables with the following distribution:²²

$$\begin{pmatrix} \theta^C \\ \theta^{NC} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \quad (4.1)$$

One of the important features of our approach is that we allow the distribution of students' latent skills to vary by their race, gender, and socioeconomic status. We implement this in the model by allowing students' vector of skills to depend both on the above individual-specific independent random variables, and a linear interaction with indicators for their demographic characteristics Z_i . In particular, a student's cognitive and non-cognitive factors, $\{F_i^C, F_i^{NC}\}$ are specified relative to those of high-SES, white women:²³

$$\begin{aligned} F_i^C &= \theta_i^C + \alpha_1^C \mathbf{Z}_i + \alpha_2^C \mathbf{Z}_i * \theta_i^C \\ F_i^{NC} &= \theta_i^{NC} + \alpha_1^{NC} \mathbf{Z}_i + \alpha_2^{NC} \mathbf{Z}_i * \theta_i^{NC} \end{aligned} \quad (4.2)$$

$$\begin{aligned} Z_i = \{ & male_i, black_i, lowSES_i, male_i * black_i, male_i * lowSES, \\ & black_i * lowSES_i, male_i * black_i * lowSES_i \} \end{aligned}$$

Since we are unable to directly observe any of the components of Equation 4.2, other than Z_i , we use a measurement system that allows us to identify $\{F_i^C, F_i^{NC}\}$ and θ_i from our set of proxy measures for cognitive and non-cognitive skills. We also

²² Carneiro, Hansen, and Heckman (2003) and Cunha and Heckman (2008) extend a similar factor model to allow for correlated factors. However, we maintain the assumption that students' cognitive and non-cognitive skills are independent, given empirical evidence shows that score correlations within cognitive or non-cognitive tests are stronger than across-test correlations.

²³ In practice we define a student as "high-SES" if they do not qualify for either the free, or reduced-price lunch program.

follow Aucejo (2015) in allowing for our proxy measures for non-cognitive skills to be jointly determined by both the latent cognitive and non-cognitive skills F_i^C and F_i^{NC} , but constraining our proxies for cognitive skills to only be a function of latent cognitive skills. With this specification, there are G and H measures of cognitive and non-cognitive skills, such that G_j denotes the j th cognitive test taken and H_k denotes the k th non-cognitive measure for individual i :

$$\begin{aligned} G_{ij} &= \gamma_{1j}^C + \gamma_{2j}^C F_i^C + \gamma_{3j}^C \mathbf{X}_i + \epsilon_{ij}^G \\ H_{ik} &= \gamma_{1k}^{CNC} + \gamma_{2k}^{CNC} F_i^C + \gamma_{3k}^{CNC} F_i^{NC} + \gamma_{4k}^{CNC} \mathbf{X}_i + \epsilon_{ik}^H \end{aligned} \tag{4.3}$$

Equation 4.3 is asymmetric in the sense that the latent cognitive factor is a determinant of all chosen measures of non-cognitive ability, but the latent non-cognitive factor is not an input into cognitive measures. Since we assume the proxy measures for non-cognitive skills H_k are jointly determined by both latent cognitive and non-cognitive skills, we allow for measures like high school GPA to be determined both by a student’s academic ability and features like their ability to complete tasks on time, or personality traits like conscientiousness. However, this added flexibility in the model, along with our lack of “pure” measures of non-cognitive skills like a Rotter scale score, implies that “non-cognitive” skills are orthogonal to cognitive skills. In other words, “non-cognitive” skills are all other skills aside from cognitive skills.

We rely on cognitive and non-cognitive measures taken from 8th and 9th grade years, since we assume that students’ skills are fixed by the time they make high school persistence and completion decisions. Our three cognitive measures G_j include standardized test scores on the North Carolina 8th grade End-of-Grade (EOG) math and reading exams, and 9th grade End-of-Course (EOC) reading exam.²⁴ The three non-cognitive measures H_k include 9th grade cumulative grade point average (GPA),

²⁴ The 8th grade EOG math and reading and 9th grade reading exams are independently standardized relative to the statewide distribution of test scores for a given exam year.

the total number of absences in 9th grade, and indicators for whether a student exerted homework effort that was below, at, or above the 8th grade class median.

These measures may also be driven in part by other covariates. For example, a school’s racial composition can affect how teachers and school administrators address absenteeism among students of different demographic groups. Hence, measures that don’t control for such school-level differences - for example, the proportion of black students in a student’ high school - would incorrectly estimate the levels of non-cognitive skills among black and white students. We capture these features by including in \mathbf{X} a vector of covariates that affect the values of these measures, such as high school characteristics. Lastly, we assume that once we condition on these covariates, the idiosyncratic error terms are mutually independent and independent of the factor variables and \mathbf{X} , such that $\epsilon^{C,NC} \perp \{\theta^{C,NC}, \mathbf{X}\}$, $\epsilon_{ja}^C \perp \epsilon_{jb}^C$, and $\epsilon_{ka}^{NC} \perp \epsilon_{kb}^{NC}$ for all $a \neq b$. Furthermore, all ϵ are assumed to be normally distributed with mean 0 for continuous variables, and Logit distributed for binary variables.

Anderson and Rubin (1956), and Carneiro, Hansen, and Heckman (2003) establish that this sort of measurement system is non-parametrically identified under three conditions.²⁵ First, factors must be orthogonal to each other such that $\theta^C \perp \theta^{NC}$, and as we described earlier, our specification of the measurement system implies the non-cognitive factor is orthogonal to the cognitive factor by construction. Second, if we have M factors we must have at least $2M + 1$ measures, $G + H$, in order to identify the factors. Since we use two factors in this model, $M = 2$, we require a combination of at least five cognitive and non-cognitive measures. Third, the underlying factor structure must exhibit the following pattern: first, least two measurements must depend exclusively on $\theta_1, \dots, \theta_{M-1}$, and second, at least three measurements must depend on $\theta_1, \dots, \theta_M$. Applied to our model, two or more measurements must depend

²⁵ This only holds up to one normalization, or when one of the loadings in each measurement equation is set to 1.

solely on θ^C and three or more on $\{\theta^C, \theta^{NC}\}$. The intuition for this restriction is that, in order to identify the non-cognitive factor separately from the cognitive factor, we need at least one additional measure that depends on both factors H , instead of just the cognitive factor G . Otherwise, we would be unable to distinguish the effect of the unobserved non-cognitive factor from the intercept parameters in the measures that only depend on the cognitive measures.

4.3.2 *Student outcomes*

Once we have established our measurement system for students' unobserved cognitive and non-cognitive factors, we can incorporate how these skills influence their decisions in the transitions through secondary education. In this section we model the sequential decision to continue in, and graduate from, high school.

Student i begins at the end of their 9th grade year and chooses whether to continue in high school by completing 10th grade. If the student chooses to drop out at any point, they enter the labor market, which is an absorbing state.²⁶ However, if a student chooses to continue, they decide at the end of each grade s whether to complete that year of education, up to the point of graduating from high school S .²⁷ Students only choose to continue in school at level s if the associated latent utility is sufficiently high:

²⁶ Some papers like Arcidiacono et al. (2014) allow students to drop- or stop-out of schooling, particularly at the college level. However we follow papers like Stange (2012) in restricting reentry for two reasons. First, modeling reentry is not the primary focus of this paper, as fewer than 8% of our sample drops out of high school at any point, and an even smaller fraction chooses to reenter. Second, allowing students to reenter schooling greatly increases the complexity and computational burden of the model.

²⁷ We also assume that all students that choose to complete 12th grade, $S = 12$ graduate, and hence do not model the stochastic process that determines whether students meet graduation requirements and are allowed to graduate. Instead, these individuals are modeled as choosing to drop out at the end of 11th grade. This may be an issue if a significant number of students are promoted through 11th grade, but not able to ultimately graduate. However, Table 4.1 shows this “bunching” of students completing 11th grade but not 12th does not appear to occur in our sample. Instead, we see proportional drops in the percentage of the original 9th grade sample that complete 10th, 11th, and 12th grade.

$$V_{is}^{HS} = \beta_{0s}^{HS} + \beta_{1s}^{HS} F_i^C + \beta_{2s}^{HS} F_i^{NC} + \beta_{3s}^{HS} \mathbf{Z}_i + \beta_{4s}^{HS} \mathbf{X}_i + \eta_{is}^{HS} \quad (4.4)$$

$$D_{is}^{HS} = \begin{cases} 1, & \text{if } V_{is}^{HS} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Equation 4.4 illustrates another assumption and benefit of using a factor model to examine students' choices. Note that the value term for pursuing a particular grade s in high school does not explicitly contain a future value term that captures the option value of remaining in school for another academic year. This is because, as is detailed in Cunha, Heckman, and Navarro (2007), factor models are able to capture the characteristics of a dynamic discrete models like this one, where students make a sequence of educational choices. Essentially, the unobserved factors $\{F_i^C, F_i^{NC}\}$ drive the dynamic selection process, whereby students drop out of high school in successive grades, or choose not to apply for and attend college.²⁸

4.3.3 Estimation

Although our model contains a number of components corresponding to the factor measurement system and student outcomes, we can express the likelihood of any sequence of schooling and major intentions for our sample as a simple expression:

$$\mathcal{L} = \prod_{i=1}^N \Pr(\mathbf{D}_i, \mathbf{G}_i, \mathbf{H}_i | \mathbf{Z}_i, \mathbf{X}_i, \mathbf{D}_{i,-1}) \quad (4.5)$$

We can then expand this function to express the likelihood as a function of three components: first, the probability of choosing to progress through high school D_{is} ;

²⁸ Cunha, Heckman, and Navarro (2007) find that a factor model must meet two requirements to accurately model a dynamic discrete choice model. First, as we explicitly state in this section, the unobservable factors $\{F_i^C, F_i^{NC}\}$ must be separable from observables like X_i and Z_i . Second, we assume that the local optimum students find by comparing choices across proximal states is also the global optimum.

second, observing our measures of cognitive and non-cognitive skills G_{ij} and H_{ik} ; and third, the probability density functions (PDFs) for cognitive and non-cognitive skills $f(\theta_i^C)$ and $f(\theta_i^{NC})$.

$$\prod_{i=1}^N \iint_{\theta_i} \left[\prod_{s=1}^S \Pr(D_{is} | F_i^C, F_i^{NC}, \mathbf{Z}_i, \mathbf{X}_i, D_{i,s-1}) \right] \times$$

$$\left[\prod_{j=1}^J \Pr(G_{ij} | F_i^C, \mathbf{X}_i) \right] \left[\prod_{k=1}^K \Pr(H_{ik} | F_i^C, F_i^{NC}, \mathbf{X}_i) \right] \times$$

$$f(\theta_i^C) f(\theta_i^{NC}) d\theta_i^C d\theta_i^{NC}$$

Since the vector of latent abilities $\boldsymbol{\theta}$ is unobserved, we estimate our parameters by integrating over the individual component of the factors, $\{\theta_i^C, \theta_i^{NC}\}$, which follow the independent standard normal distributions we specify in Equation 4.1. However, since integrating over each unobserved factor is computationally burdensome, we use the Legendre-Gauss Quadrature to discretize the integral into a weighted sum of ten points for each of the unobserved latent factors.²⁹

4.4 Results

Table 4.5 reports the parameter estimates corresponding to Logits for progressing through secondary education. The factor model specified in Section 4.3 explains away gaps between high-SES white females and black student subgroups more readily than other white student subgroups. After flexibly controlling for high school characteristics and nonlinear functions of cognitive and non-cognitive skill proxies, there is still a statistically significant deficit in the probability of high school completion among white students across the SES gradient.

²⁹ We also specified a model where we follow Aguirregabiria and Mira (2007) by discretizing the distributions of the individual-specific components of the latent factors as Normal(0, 1) distributions with $T = 21$ points of support. Our results are robust to either numerical integration method.

We also find, like Aucejo (2015), that the importance of both sets of skills for completing a given level of education declines as students progress through higher levels of secondary education. The parameter estimates across years 2-4 show that the size of the attainment gap increases among poorer students of both races. Coefficients on both factors confirm that the explanatory functions of cognitive and non-cognitive skills are decreasing over the schooling period. By the time students are choosing whether to complete their high school education, both skills have diminished considerably in magnitude. This implies that selection on these skills operates more heavily in earlier grades. By the time students get to 12th grade, those with lower stocks of cognitive and non-cognitive skills have already chosen to exit secondary education, making these skills less crucial inputs in educational attainment decisions.

Table 4.5: Structural high school continuation parameter estimates

	On track by Year 2	On track by Year 3	On track by Year 4 and graduate
Cognitive factor	1.665*** (0.0173)	0.566*** (0.0161)	-0.192*** (0.0145)
Non-cognitive factor	2.785*** (0.0288)	1.038*** (0.0225)	0.426*** (0.0203)
Demographic subgroups^a			
High SES			
White Male	0.672*** (0.0462)	0.717*** (0.0409)	-0.067** (0.0314)
Black Female	1.049*** (0.0912)	0.532*** (0.0788)	-0.171** (0.0685)
Black Male	0.993*** (0.0725)	1.531*** (0.1063)	0.138* (0.0794)
Low SES			
White Female	0.224*** (0.0457)	-0.302*** (0.0380)	-0.789*** (0.0340)
White Male	0.369*** (0.0449)	-0.056 (0.0395)	-0.750*** (0.0354)
Black Female	1.014*** (0.0470)	0.668*** (0.0453)	-0.289*** (0.0427)
Black Male	0.957*** (0.0473)	0.574*** (0.0453)	-0.645*** (0.0422)

Note: Logits for each grade are estimated on a sample of students that completed the prior grade.

^a Indicators for race/gender/SES categories are defined relative to high-SES white females.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We see in Table 4.6 that all 7 listed groups have levels of cognitive skills that are significantly lower than the reference group of high SES white females. Magnitudes correspond closely to race and SES pairings. Disadvantaged black students have the lowest average cognitive factors, while the cognitive achievement of advantaged

black students is comparable to disadvantaged white students. A differential story emerges for non-cognitive skills, with patterns tracking by gender and SES pairings. Underprivileged females, for example, have similar levels of non-cognitive skills that rise above that of underprivileged male students. These results suggest that the relative attainment advantage among poorer black female students observed in Table 4.4 is not attributable to higher non-cognitive skills as measured by school-based outcomes such as GPA, absences, and homework effort.

Table 4.6: Structural factor parameter estimates

	Cognitive	Non-cognitive
High SES		
White Male	-0.169*** (0.0059)	-0.057*** (0.0052)
Black Female	-0.543*** (0.0133)	0.007 (0.0117)
Black Male	-0.789*** (0.0132)	-0.093*** (0.0144)
Low SES		
White Female	-0.546*** (0.0076)	-0.247*** (0.0069)
White Male	-0.739*** (0.0081)	-0.415*** (0.0074)
Black Female	-1.150*** (0.0083)	-0.258*** (0.0077)
Black Male	-1.415*** (0.0083)	-0.483*** (0.0084)

Note: Indicators for race/gender/SES categories are defined relative to high-SES white females.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: Structural cognitive measure parameter estimates

	8th Grade EOG Scores ^a		9th Grade EOG Scores	
	Math	Reading	Reading	
Cognitive factor ^b	1	0.595*** (0.0009)	0.687***	(0.0013)
σ^2	0.332	(1.0034)	0.184	(1.0034)
			0.345	(1.0033)

^a End-of-Grade scores in math and reading are normalized relative to statewide test scores for the year a student took a given EOG test.

^b In order to identify the cognitive factor, we must normalize the parameter for the cognitive factor to 1 on at least one of the measures.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: Structural non-cognitive measure parameter estimates

	Continuous		Binary - Homework effort		
	9th Grade GPA	Days Absent	Below	Above	Refuse
Cognitive	0.543*** (0.0021)	-0.752*** (0.0184)	-0.101*** (0.0054)	0.263*** (0.0054)	-0.166*** (0.0181)
Non-cognitive ^a	1	-4.873*** (0.0215)	-0.138*** (0.0069)	0.071*** (0.0072)	-0.961*** (0.0198)
σ^2	0.229 (1.0042)	52.353*** (1.0032)			

^a In order to identify the non-cognitive factor, we must normalize the parameter for the non-cognitive factor to 1 on at least one of the measures.

^b High school characteristics are taken from a student's 9th grade institution.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5 Conclusion

An individual's educational decisions throughout secondary school have important implications for postsecondary and labor market outcomes. We examine this crucial period for student groups at the three-way intersection of race, gender, and socioeconomic status, using statewide administrative data that enables a richer characterization of cross-group disparities than previous US-based work. The detailed segmentation makes clear that collapsing across socioeconomic status masks significant heterogeneity. Attainment bifurcates by SES, with more economically advantaged students graduating at higher rates regardless of race or gender. Strikingly, low SES white females consistently graduate at lower rates than their disadvantaged black female peers, despite having higher standardized test scores. As such, the black-white achievement gap does not convert to an analogous high school graduation gap.

The characterization of family socioeconomic status using a binary indicator of free and reduced lunch eligibility obscures the nuances of family background. Alternative definitions of SES using the history of free and reduced lunch eligibility and parental income reinforce the findings on the relative under-performance of poorer white students. The patterns are especially pronounced among students in the lowest SES category. These empirical regularities are not explained by higher household incomes for disadvantaged white students, since they fall behind in attainment despite residing in neighborhoods with higher median incomes. Nationally representative

data in the NLSY97 confirm the attainment advantage of poorer black females, even though they are worse off financially than their white peers.

To account for the determinants of these disparities, we examine the relative contributions of high school inputs, cognitive, and non-cognitive skills.³⁰ We begin with OLS regressions and progress to a sequential factor model of educational attainment. Conditioning on high school characteristics and the same levels of cognitive and non-cognitive inputs leads all black student subgroups to exceed the graduation rates of high SES white females with the exception of poorer black males. In contrast, these skills fall far short of explaining the low graduation rates of disadvantaged white students. We supplement reduced form results with structural analyses to address concerns with identification. Variables such as test scores measure “pure” abilities with error, and can be themselves affected by schooling attainment. Furthermore, some schooling measures are jointly determined by underlying cognitive and non-cognitive abilities. Shifting to a structural model of schooling attainment permits us to separately identify the contributions of these two factors at every educational stage. We find that non-cognitive skills exert greater influence on graduation outcomes relative to cognitive skills. However, we do not find evidence that higher non-cognitive skills among poorer black students contribute to their relative graduation advantage compared to poorer white peers.

Taken together, economic deprivation among black students translates to lower cognitive and non-cognitive skills that account for their reduced attainment. In contrast, economic disadvantage among white students materializes in different ways to put downward pressure on graduation rates. For one, their markedly higher cognitive scores do not translate into attainment success. This suggests a need to tailor

³⁰ The characterization of observed ability measures as explanatory accounts should not obscure the fact that they are shaped over time by familial, neighborhood, and schooling attributes. To the extent that skills are determined by contextual factors not considered in the model, the contributions of these skills to educational disparities reflect the collective impact of these influences.

policies aimed at bridging graduation gaps by the race of the economically disadvantaged student. While closing test score and absenteeism gaps may go a long way for poor black students, improving academic outcomes, absenteeism, or homework effort will likely fall short for poor white students. Future research can investigate labor market conditions or public health trends that differentially impact white and black communities, and the effects they may have on students' schooling attainment.

Appendix A

CHAPTER 1 - TABLES

Table A1: Cutoff definition 2

CD	Application year				Total
	2009	2010	2011	2012	
1	0.98	0.85	0.97	0.85	0.91
2	0.95	0.84	0.86	0.90	0.89
3	0.85	0.88	0.91	0.95	0.90
4	0.86	0.89	0.96	0.94	0.91
5	0.96	0.70	0.87	0.94	0.87
6	0.98	0.92	0.94	0.93	0.95
7	0.99	0.82	0.90	0.87	0.90
8	0.93	0.88	0.99	0.91	0.93
9	0.89	0.89	0.96	0.90	0.91
10	0.94	0.93	0.95	0.95	0.94
11	1.00	0.96	0.97	0.84	0.94
12	0.93	0.88	0.92	0.85	0.90
13	0.96	0.82	0.96	0.87	0.89
Total	0.92	0.87	0.94	0.91	0.91

Table A2: Cutoff definition 3

CD	Application year				Total
	2009	2010	2011	2012	
1	0.94	0.83	0.85	0.86	0.87
2	0.94	0.75	0.87	0.76	0.82
3	0.89	0.85	0.90	0.94	0.90
4	0.51	0.61	0.83	0.71	0.66
5	0.91	0.69	0.95	0.92	0.87
6	0.96	0.90	0.60	0.92	0.86
7	0.94	0.79	0.93	0.82	0.87
8	0.90	0.85	0.95	0.90	0.90
9	0.73	0.87	0.95	0.67	0.80
10	0.90	0.88	0.96	0.93	0.92
11	0.94	0.91	0.94	0.80	0.89
12	0.84	0.79	0.85	0.83	0.83
13	0.95	0.73	0.95	0.80	0.84
Total	0.81	0.77	0.88	0.81	0.82

Table A3: Balance of black applicants across cutoff

CD	All	5 points at or below cutoff	5 points above cutoff	
1	0.36		0.31	0.42
2	0.12		0.02	0.13
3	0.07		0.02	0.04
4	0.09		0.01	0.01
5	0.04		0.00	0.07
6	0.09		0.04	0.13
7	0.15		0.10	0.27
8	0.15		0.09	0.21
9	0.12		0.02	0.12
10	0.05		0.02	0.07
11	0.02		0.00	0.05
12	0.24		0.10	0.50
13	0.13		0.11	0.19
Total	0.12		0.05	0.14

Notes: table shows the distribution of black applicants in the neighborhood around the cutoff, distinguishing between applicants who are at or ≤ 5 points below the cutoff and applicants who are ≤ 5 points above the cutoff.

Table A4: Baseline statistics

	Gender		Race		SES		ZIP urbanicity		HS quality		HS adv. curriculum		
	All (1)	Female (2)	Male (3)	Black (4)	Non-black (5)	Low (6)	High (7)	Rural (8)	Urban (9)	Low (10)	High (11)	Low (12)	High (13)
SAT math percentile	82.90	80.28	86.40	62.57	84.07	75.01	83.98	78.38	87.50	78.14	87.58	77.85	88.26
SAT verbal percentile	80.62	79.91	81.56	62.93	81.63	73.38	81.60	75.25	86.02	75.01	85.91	74.39	87.04
Consis STEM intention	0.29	0.20	0.42	0.19	0.30	0.31	0.29	0.28	0.30	0.27	0.31	0.27	0.31
Total institutions	5.86	5.65	6.14	5.83	5.86	6.20	5.81	5.21	6.55	5.22	6.47	5.15	6.64
Share of selective colleges	0.83	0.81	0.87	0.61	0.85	0.72	0.85	0.78	0.89	0.78	0.89	0.77	0.90
Avg share STEM BA	0.22	0.20	0.25	0.18	0.22	0.21	0.22	0.21	0.23	0.20	0.23	0.20	0.24

Notes: table shows baseline characteristics for applicants who are at or \leq 5 points below the cutoff.

Table A5: Heterogeneous treatment effects - nonparametric estimates

	Gender		Race		SES		ZIP urbanicity		HS quality		HS adv. curriculum	
	Female	Male	Black	Non-black	Low	High	Rural	Urban	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
SAT math percentile	1.566* (0.813)	-0.041 (0.681)	1.833 (1.782)	0.758 (0.606)	4.099*** (1.500)	0.380 (0.604)	2.405*** (0.704)	-0.599 (0.715)	2.483*** (0.774)	-0.813 (0.680)	2.422*** (0.777)	-0.862 (0.666)
SAT verbal percentile	1.385 (1.187)	1.623 (1.497)	1.184 (2.527)	1.523 (1.121)	1.992 (2.193)	1.414 (1.160)	3.144** (1.376)	0.193 (1.152)	3.789*** (1.427)	-0.844 (1.118)	2.615* (1.449)	0.081 (1.079)
Consistent STEM intention	0.039 (0.036)	-0.002 (0.043)	0.023 (0.071)	0.020 (0.035)	0.041 (0.071)	0.018 (0.035)	0.044 (0.040)	0.003 (0.040)	0.032 (0.046)	0.002 (0.037)	0.013 (0.044)	0.027 (0.042)
Total no. of colleges	0.282 (0.307)	0.164 (0.327)	0.055 (0.517)	0.252 (0.286)	-0.394 (0.428)	0.327 (0.283)	-0.026 (0.307)	0.468 (0.314)	0.126 (0.282)	0.309 (0.329)	-0.049 (0.337)	0.463 (0.302)
Share of selective colleges	0.041*** (0.016)	0.022 (0.015)	0.071* (0.036)	0.028** (0.013)	0.071*** (0.026)	0.027** (0.013)	0.055*** (0.018)	0.011 (0.012)	0.057*** (0.017)	0.003 (0.014)	0.050*** (0.018)	0.014 (0.012)
Average share of STEM BAs	0.018*** (0.006)	0.018*** (0.007)	0.003 (0.010)	0.020*** (0.005)	0.015 (0.010)	0.018*** (0.005)	0.018*** (0.006)	0.020*** (0.006)	0.017** (0.007)	0.016*** (0.006)	0.012* (0.006)	0.024*** (0.006)

Notes: one specification is run for every category (e.g. gender). Reduced form coefficients are the interactions between the individual or HS attributes (e.g. female) and an indicator for exceeding the score cutoff. All specifications include a vector of individual characteristics including gender, race, SES, 10th grade SAT math, and CD-cohort fixed effects. Reported estimates are nonparametric estimates using optimal bandwidths and a linear control of the distance to the cutoff on each side of the threshold. Robust standard errors are clustered at the cohort, CD, and gender level. * p<0.1, ** p<0.05, *** p<0.01

Table A6: Heterogeneous treatment effects - nonparametric estimates using race-based cutoffs

	Gender		Race		SES		ZIP urbanicity		HS quality		HS adv. curriculum	
	Female (1)	Male (2)	Black (3)	Non-black (4)	Low (5)	High (6)	Rural (7)	Urban (8)	Low (9)	High (10)	Low (11)	High (12)
SAT math percentile	1.441* (0.858)	0.259 (0.773)	2.289 (1.723)	0.861 (0.732)	4.976*** (1.557)	0.487 (0.722)	2.425*** (0.775)	-0.450 (0.849)	2.775*** (0.837)	-0.726 (0.801)	2.794*** (0.804)	-1.179 (0.791)
SAT verbal percentile	0.769 (1.032)	1.343 (1.144)	2.692 (2.293)	0.880 (0.959)	2.494 (1.894)	0.831 (0.972)	2.465** (1.223)	-0.043 (0.987)	3.081** (1.291)	-0.502 (0.935)	2.239** (1.135)	0.096 (0.973)
Consistent STEM intention	0.002 (0.031)	-0.029 (0.046)	-0.001 (0.083)	-0.011 (0.031)	0.004 (0.072)	-0.012 (0.031)	-0.022 (0.037)	0.006 (0.040)	0.017 (0.040)	-0.040 (0.038)	-0.015 (0.037)	-0.018 (0.040)
Total no. of colleges	0.039 (0.257)	0.163 (0.244)	-0.285 (0.477)	0.123 (0.214)	-0.714* (0.394)	0.198 (0.216)	-0.111 (0.255)	0.256 (0.249)	-0.152 (0.238)	0.244 (0.250)	-0.086 (0.266)	0.275 (0.248)
Share of selective colleges	0.037** (0.014)	0.021 (0.015)	0.098*** (0.036)	0.025** (0.012)	0.044 (0.030)	0.028** (0.012)	0.055*** (0.016)	0.007 (0.012)	0.058*** (0.014)	0.001 (0.013)	0.053*** (0.016)	0.013 (0.012)
Average share of STEM BAs	0.015** (0.006)	0.020*** (0.007)	0.006 (0.010)	0.018*** (0.006)	0.011 (0.012)	0.018*** (0.005)	0.017** (0.007)	0.020*** (0.006)	0.018*** (0.007)	0.015*** (0.006)	0.011 (0.007)	0.025*** (0.006)

Notes: one specification is run for every category (e.g. gender). Reduced form coefficients are the interactions between the individual or HS attributes (e.g. female) and an indicator for exceeding the score cutoff. All specifications include a vector of individual characteristics including gender, race, SES, 10th grade SAT math, and CD-cohort fixed effects. Reported estimates are nonparametric estimates using optimal bandwidths and a linear control of the distance to the cutoff on each side of the threshold. Robust standard errors are clustered at the cohort, CD, and gender level. * p<0.1, ** p<0.05, *** p<0.01

Appendix B

CHAPTER 1 - ROBUSTNESS CHECKS

Identifying assumptions must hold for an unbiased causal interpretation of estimated coefficients. This section provides additional analyses on the validity of the RD design and examines alternative explanations that can confound inference.

Challenges to local random assignment

The main challenge to local random assignment is sorting and selection around cut-offs. A direct method of testing for strategic behavior around the cutoff derives from the local linear regression approach outlined in McCrary (2008). The density test computes frequency counts within equally-spaced bins and smooths the histogram separately on either side of the cutoff using local linear regression. In the presence of sorting or selective attrition, more observations would cluster around the cutoff than would otherwise be expected, resulting in a discontinuity of assignment variable density at the threshold. Figure B1 shows a small increase in density to the left of the threshold. Further investigation shows that this pattern is exacerbated among white and Asian students, suggesting the presence of some selection by race. This echoes the graphs on covariate smoothness: baseline characteristics are smooth

across the threshold for all observables with the exception of race. These patterns are consistent with two explanations: either minority students exercise more control in exceeding cutoffs relative to their peers, or the admissions committee intervened to ensure diversity in the student population.

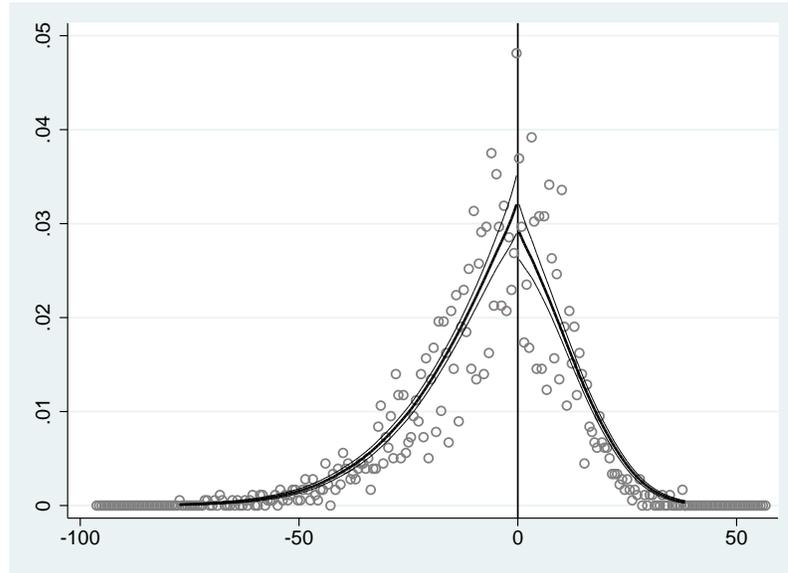


FIGURE B1: McCrary density around the cutoff

Notes: X-axis shows the distance from cutoff in the pooled sample. Full sample includes unique matches with non-missing admission scores and CD information.

The design of the admission process renders score manipulation on behalf of students and parents a difficult task. This level of control requires at least two sources of information. The first is the content of the admission rubric that translates application materials into a summary score. Presumably students know that inputs such as SAT math scores matter, but they do not know the exact weight given to each part of the application, and as such cannot precisely influence their admission scores. The second source of knowledge requires applicants to know the score distribution of other applicants within their cohort-district-gender group. Only then can students forecast a reliable cutoff score around which to manipulate their application materials. The magnitude of information required to meet the second condition in

particular implies that applicants do not have precise control over their scores relative to the threshold and renders strategic self-selection around this cutoff unlikely. In contrast, selection by the admissions reviewers is feasible because they are explicitly tasked with the goal of ensuring student body diversity. This permits them to admit students on the basis of socio-demographic characteristics such as race.

Race is included across all specifications to account for selection on this dimension of applicant background and correlated observables. Its inclusion serves the secondary purpose of increasing estimate precision, particularly when a baseline covariate is highly correlated with the outcome of interest. It is possible to investigate the role of race by exploring the sensitivity of results when it is omitted. Table B1 shows that RD estimates without race indicators are qualitatively similar to the main results. All outcomes with the exception of postsecondary STEM-intensity are insignificant as before. Enrollment in the selective high school induces a slightly larger effect on the average STEM-intensity of prospective colleges.

Table B1: Robustness - inclusion of race

Dependent variable	With race	Without race
SAT math percentile	0.745 (0.819)	-0.381 (0.851)
SAT verbal percentile	1.638 (1.459)	0.507 (1.454)
Consistent STEM intention	0.064 (0.047)	0.069 (0.046)
Total no. of colleges	0.021 (0.421)	0.163 (0.433)
Share of selective colleges	0.018 (0.020)	0.005 (0.019)
Average share of STEM BAs	0.024*** (0.008)	0.026*** (0.008)

Notes: Indicators include African American, Hispanic, Asian, American Indian, and Other.

A second means of addressing the selection on race is to replace existing cutoffs

with CD-, cohort-, and race-specific thresholds. Each applicant’s distance to the cutoff then depends on within-group peer performance, and the cutoffs are permitted to vary across racial and ethnic groups in a given district. Table B2 shows that this hypothetical admissions procedure produces similar shares of black students on both sides of the cutoff, in contrast to Table A3. Next I pool across all samples and re-estimate heterogeneous treatment effects under a nonparametric model. The results shown in Table A6 are very similar in magnitude and significance to both parametric and nonparametric estimates in Table A5 under the original cutoff definition. Economically disadvantaged students are still registering the highest SAT math gains of 4-5 percentile points.

Table B2: Robustness - balance under race-based cutoffs

CD	All	5 points at or below cutoff	5 points above cutoff
1	0.38	0.39	0.40
2	0.12	0.07	0.03
3	0.07	0.02	0.03
4	0.09	0.03	0.03
5	0.04	0.02	0.03
6	0.09	0.10	0.10
7	0.14	0.12	0.10
8	0.16	0.25	0.11
9	0.13	0.08	0.15
10	0.05	0.03	0.03
11	0.01	0.00	0.00
12	0.24	0.18	0.15
13	0.15	0.14	0.16
Total	0.12	0.09	0.08

Notes: table shows the distribution of black applicants in the neighborhood around the cutoff, distinguishing between applicants who are at or \leq 5 points below the cutoff and applicants who are \leq 5 points above the cutoff.

Another challenge to local random assignment is differential attrition. Non-admitted students from public sending schools may be more likely to switch to private high schools or move out of state, leaving the control group negatively selected.

Counterfactual analyses on compliers suggest that the vast majority remain in the state's public school system. While I cannot rule out the possibility of mismatching students who move out of North Carolina or attend private school, this form of selective attrition is unlikely given that switches must take place during an academic and socially crucial period for students. Junior year of high school marks the beginning of college preparations, and therefore the gains from a school switch are less likely to offset the disruptive costs associated with such a move.

Match quality

Another dimension that affects inference is match quality. It is well known that measurement error introduced into explanatory variables attenuates estimates. Merging on demographic and sending school attributes can result in non-random measurement error in dependent variables. Biases arise if match quality and consequently the extent of measurement error in outcomes introduce correlation between explanatory variables and the error term.

I undertake two strategies to gauge the results' sensitivity to the merging process. The first re-estimates the models using only applicants with an unique match from North Carolina public and charter high schools. Applicants with unique matches may differ systematically from those with multiple in ways that are difficult to observe, for instance if under-privileged applicants achieve more precise matches while also experiencing greater benefits from attending the selective school. The second relies on bootstrap sampling and estimation. Within each applicant, I randomly sample with replacement one observation from the set of matches, such that the same individual is always sampled for applicants with one unique match and the probability of being sampled decreases in the number of total matches. Bootstrapped coefficients and standard errors are computed for 1000 repetitions.

Results from the sample of unique matches are close in magnitude to reduced

form estimates using the full sample (Table B3). The average STEM-intensity of colleges the students apply to decreases slightly to 2.3 percentage points from 2.4 percentage points. Figure B2 shows the distribution of bootstrapped estimates for STEM-intensity. The mean of the distribution approximates the reduced form estimates using the full sample. Coefficients using other outcomes are also consistent with previous estimates. In short, the evidence suggests that results are not sensitive to the merging process.

Table B3: Robustness - unique matches

Dependent variable	Unique matches only
SAT math percentile	0.380 (0.856)
SAT verbal percentile	1.124 (1.543)
Consistent STEM intention	0.060 (0.052)
Total no. of colleges	0.073 (0.440)
Share of selective colleges	0.006 (0.022)
Average share of STEM BAs	0.023** (0.009)

Notes: Results exclude applicants linked to two or more observations. Robust standard errors are clustered at the cohort, CD, and gender level.

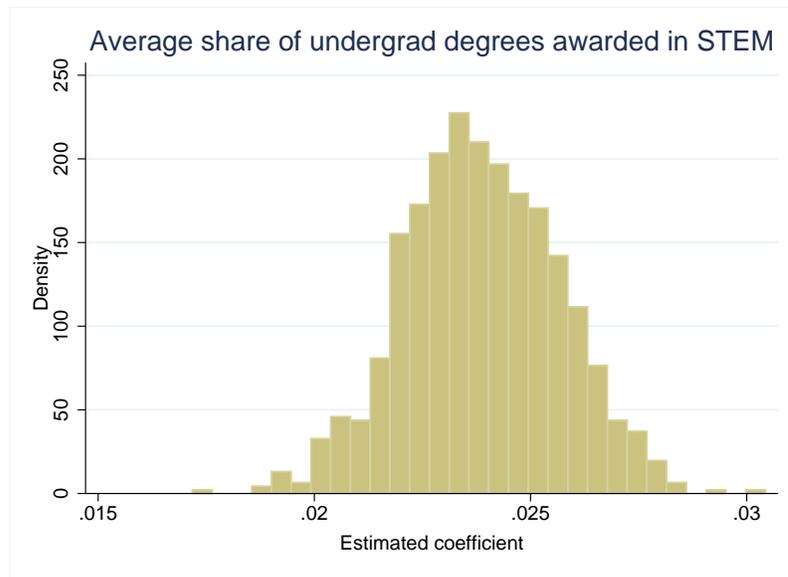


FIGURE B2: Distribution of STEM-intensity coefficient in bootleg sampling

Appendix C

CHAPTER 2 - SUMMARY STATISTICS

Table C1: Engineering orientation in second half of high school

	Female		Male	
	Engineer	Not engineer	Engineer	Not engineer
SAT math	568.10	493.50***	553.97	517.09***
SAT verbal	533.42	487.90***	501.82	491.94***
Cumulative GPA in Grades 9 and 10	3.82	3.41***	3.44	3.24***
Earned reading credits in Grade 9	1.18	1.18	1.14	1.16***
Earned math credits in Grade 9	1.22	1.24**	1.23	1.24***
Earned physical science credits in Grade 9	0.08	0.07**	0.07	0.07**
Earned comp. prog. credits in Grade 9	0.01	0.00***	0.01	0.00***
Earned reading credits in Grade 10	1.11	1.15***	1.07	1.09***
Earned math credits in Grade 10	1.26	1.17***	1.20	1.16***
Earned physical science credits in Grade 10	0.53	0.34***	0.39	0.33***
Earned comp. prog. credits in Grade 10	0.01	0.00***	0.02	0.01***
GRD 9 COMPUTER ACTIVITY	0.15	0.09***	0.16	0.12***
GRD 10 COMPUTER ACTIVITY	0.15	0.09***	0.16	0.12***
GRD 9 MUSIC/VOCAL ACTIVITY	0.08	0.11***	0.03	0.04***
GRD 10 MUSIC/VOCAL ACTIVITY	0.08	0.11***	0.03	0.04***
GRD 9 THEATER ACTIVITY	0.08	0.09**	0.03	0.04***
GRD 10 THEATER ACTIVITY	0.08	0.10***	0.03	0.05***
GRD 9 JUNIOR ROTC	0.04	0.02***	0.05	0.04***
GRD 10 JUNIOR ROTC	0.04	0.02***	0.05	0.04***
GRD 9 DANCE ACTIVITY	0.11	0.12	0.01	0.01***
GRD 10 DANCE ACTIVITY	0.11	0.12**	0.01	0.01***
GRD 9 GOVT/POLITICAL ACTIVITY	0.08	0.05***	0.03	0.03***
GRD 10 GOVT/POLITICAL ACTIVITY	0.11	0.07***	0.04	0.05***
GRD 9 JOURNALISM/LITERARY	0.03	0.03	0.01	0.01***
GRD 10 JOURNALISM/LITERARY	0.05	0.06**	0.01	0.02***
Observations	147292		119618	

Notes: analytic sample comprises students who took the SAT exam in 2009 - 2014 at least once as juniors or seniors. Sample excludes observations with missing SAT scores and those who cannot be linked to 9th or 10th grade transcript data. Stars denote statistically significant difference in means relative to students of same gender group who chose engineering as preferred major. * p<0.1, ** p<0.05, *** p<0.01

Table C2: Majoring in engineering at UNC

	Female		Male	
	Engineer	Not engineer	Engineer	Not engineer
SAT math	593.20	520.31***	603.18	556.80***
SAT verbal	549.77	514.11***	537.82	525.79***
Cumulative GPA in HS	4.19	3.75***	3.95	3.66***
Earned reading credits in Grade 9	1.15	1.14	1.10	1.11
Earned math credits in Grade 9	1.16	1.15	1.18	1.17
Earned physical science credits in Grade 9	0.11	0.12	0.13	0.12
Earned comp. prog. credits in Grade 9	0.01	0.00***	0.01	0.01**
Earned reading credits in Grade 10	1.13	1.14	1.06	1.08
Earned math credits in Grade 10	1.14	1.10	1.17	1.10***
Earned physical science credits in Grade 10	0.51	0.35***	0.42	0.36***
Earned comp. prog. credits in Grade 10	0.01	0.00	0.03	0.02*
Earned reading credits in Grade 11	1.13	1.15	1.02	1.05***
Earned math credits in Grade 11	1.15	1.06***	1.17	1.07***
Earned physical science credits in Grade 11	0.61	0.64	0.78	0.69***
Earned comp. prog. credits in Grade 11	0.01	0.00	0.03	0.02***
Earned reading credits in Grade 12	1.23	1.22	1.07	1.13***
Earned math credits in Grade 12	1.23	0.99***	1.25	1.05***
Earned physical science credits in Grade 12	0.49	0.25***	0.63	0.35***
Earned comp. prog. credits in Grade 12	0.01	0.00	0.03	0.02***
Observations	12564		9610	

Notes: UNC sample comprises 2009-2010 enrollees who declared a major by the time of attaining 30 credit hours at their home institution, with non-missing high school GPAs, SAT scores, and transcript data. Stars denote statistically significant difference in means relative to students of same gender group who chose engineering as preferred major. * p<0.1, ** p<0.05, *** p<0.01

Table C3: Freshmen engineering intentions: CIRP Survey

	Female		Male	
	Engineer	Not engineer	Engineer	Not engineer
SAT math	630.20	555.14***	631.80	581.88***
SAT verbal	579.35	546.65***	565.59	549.94***
Self Rating: Academic ability	4.28	3.91***	4.24	4.00***
Self Rating: Mathematical ability	4.15	3.35***	4.18	3.61***
Self Rating: Writing ability	3.47	3.56***	3.36	3.52***
Goal: Being very well off financially	3.02	2.96***	3.16	3.09***
Goal: Helping others who are in difficulty	2.79	2.97***	2.55	2.70***
Goal: Influencing social values	2.18	2.42***	2.03	2.29***
Future Act: Participate in volunteer/service work	3.11	3.11***	2.61	2.67***
Goal: Becoming accomplished in the performing arts	1.46	1.60***	1.36	1.52***
Goal: Creating artistic work	1.42	1.57***	1.37	1.51***
Goal: Making a theoretical contribution to science	2.24	1.67***	2.18	1.77***
Goal: Raising a family	2.94	3.07***	3.04	3.07***
Observations	1083838		958994	

Notes: Freshmen Survey sample spans 1990-1999, 2001, 2004, 2006, 2008, and 2010 academic years. Means reported using sample weights. Self-rating variables are reported on a scale of 1-5: 1) lowest 10%, 2) below average, 3) average, 4) above average, and 5) highest 10%. Goals and Future Acts are reported on a 1-4 scale. The corresponding categories for goals are: 1) not important, 2) somewhat important, 3) very important, and 4) essential. The scale for future acts inquire about the probability of undertaking a future action: 1) no chance, 2) very little chance, 3) some chance, and 4) very good chance. Stars denote statistically significant difference in means relative to students of same gender group who chose engineering as preferred major. * p<0.1, ** p<0.05, *** p<0.01

Table C4: Engineering orientation: twin sample

	Female				Male			
	Mean		N		Mean		N	
	Same-sex	Opp-sex	Same-sex	Opp-sex	Same-sex	Opp-sex	Same-sex	Opp-sex
Conditional on taking SATs								
Engineering orientation	0.02	0.02	2828	1593	0.15	0.19***	2256	1315
SAT math	491.55	491.75	2828	1593	525.07	515.66**	2256	1315
SAT verbal	484.77	485.76	2828	1593	493.47	485.68**	2256	1315
Full sample								
Engineering orientation	0.01	0.01	6174	3385	0.05	0.07***	6194	3385
Took SATs	0.46	0.47	6174	3385	0.36	0.39**	6194	3385
Math EOG in Grade 8	0.15	0.12	5131	2837	0.11	0.09	4943	2745
Math EOG in Grade 7	0.11	0.09	5291	2902	0.10	0.08	5088	2819
Math EOG in Grade 6	0.09	0.09	5301	2945	0.07	0.06	5153	2889
Reading EOG in Grade 8	0.15	0.12	5131	2838	0.05	0.03	4911	2742
Reading EOG in Grade 7	0.17	0.16	5287	2901	0.02	0.01	5059	2810
Reading EOG in Grade 6	0.14	0.14	5291	2942	0.00	-0.01	5119	2884
Attrited in Gr. 7	0.05	0.04*	6174	3385	0.05	0.04	6194	3385
Attrited in Gr. 6	0.03	0.03	6174	3385	0.03	0.03	6194	3385
Attrited in Gr. 5	0.02	0.02	6174	3385	0.03	0.03	6194	3385
Free or reduced lunch in Grade 8	0.42	0.41	5135	2842	0.41	0.40	4994	2753
Free or reduced lunch in Grade 7	0.42	0.40	5080	2781	0.40	0.38	4947	2709
Free or reduced lunch in Grade 6	0.43	0.42	5178	2855	0.41	0.41	5087	2831
Use computer \geq 1-2 times/wk in Gr 8	0.35	0.34	5113	2824	0.28	0.31***	4919	2730
Use computer \geq 1-2 times/wk in Gr 7	0.27	0.29*	5267	2886	0.22	0.25***	5077	2800
Use computer \geq 1-2 times/wk in Gr 6	0.23	0.24	5300	2939	0.19	0.21**	5154	2875

Notes: The full sample comprises all twin pairs appearing in grade 3 during 2000 - 2005, up to grade 8 in 2005 - 2010. Twins are flagged in NCERDC on the basis of identifying information such as birth date and home address. There are 3385 opposite-sex pairs, 3087 female same-sex pairs, and 3097 male same-sex pairs. The top panel retains all individuals who took the SATs and have non-missing score information between 2009 - 2014. Stars denote statistically significant difference in means of students in a given gender group who belong to a same-sex versus opposite-sex pair. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix D

CHAPTER 3 - TABLES AND FIGURES

Table D1: Grade progression by year using the NCERDC

	2002	03	04	05	06	07	08	09	10	11	12	13	14
Grade	3	4	5	6	7	8	9	10	11	12			
Grade		3	4	5	6	7	8	9	10	11	12		
Grade			3	4	5	6	7	8	9	10	11	12	
Grade				3	4	5	6	7	8	9	10	11	12

Note: Red text refers to the four 9th grade cohorts used in the construction of our master dataset. Shading of cells determine the source files for information on socioeconomic status. Blue-shaded cells derive from EOG files, while gray-shaded cells derive from Masterbuild files.

Table D2: Sample attrition and 12th grade graduation outcomes

	(1)	(2)	(3)	(4)	(5)
F + W + Hi SES	0.92	0.93	0.93	0.93	0.93
F + B + Hi SES	0.88	0.94	0.94	0.94	0.94
M + W + Hi SES	0.89	0.91	0.91	0.91	0.92
M + B + Hi SES	0.81	0.90	0.90	0.90	0.90
F + W + Lo SES	0.71	0.76	0.76	0.76	0.77
F + B + Lo SES	0.76	0.81	0.81	0.81	0.82
M + W + Lo SES	0.64	0.71	0.72	0.72	0.72
M + B + Lo SES	0.64	0.72	0.73	0.73	0.74
Total	0.79	0.84	0.84	0.84	0.85
Non-missing cognitive skills		Y	Y	Y	Y
Non-missing 9th grade GPA			Y	Y	Y
Non-missing 9th grade absence data				Y	Y
Non-missing HS attributes, homework effort					Y
<i>N</i>	342245	286363	280298	280281	274555

Note:

Table D3: Sample attrition and share of each student group

	(1)	(2)	(3)	(4)	(5)
F + W + Hi SES	21.12	22.41	22.42	22.42	22.52
F + B + Hi SES	3.23	3.14	3.13	3.13	3.14
M + W + Hi SES	22.00	22.93	22.93	22.93	23.01
M + B + Hi SES	3.29	3.00	2.98	2.98	2.98
F + W + Lo SES	11.02	11.28	11.41	11.41	11.44
F + B + Lo SES	13.78	13.82	13.71	13.71	13.66
M + W + Lo SES	11.83	11.11	11.21	11.21	11.20
M + B + Lo SES	13.73	12.31	12.21	12.21	12.06
Non-missing cognitive skills		Y	Y	Y	Y
Non-missing 9th grade GPA			Y	Y	Y
Non-missing 9th grade absence data				Y	Y
Non-missing HS attributes, homework effort					Y
<i>N</i>	342245	286363	280298	280281	274555

Note:

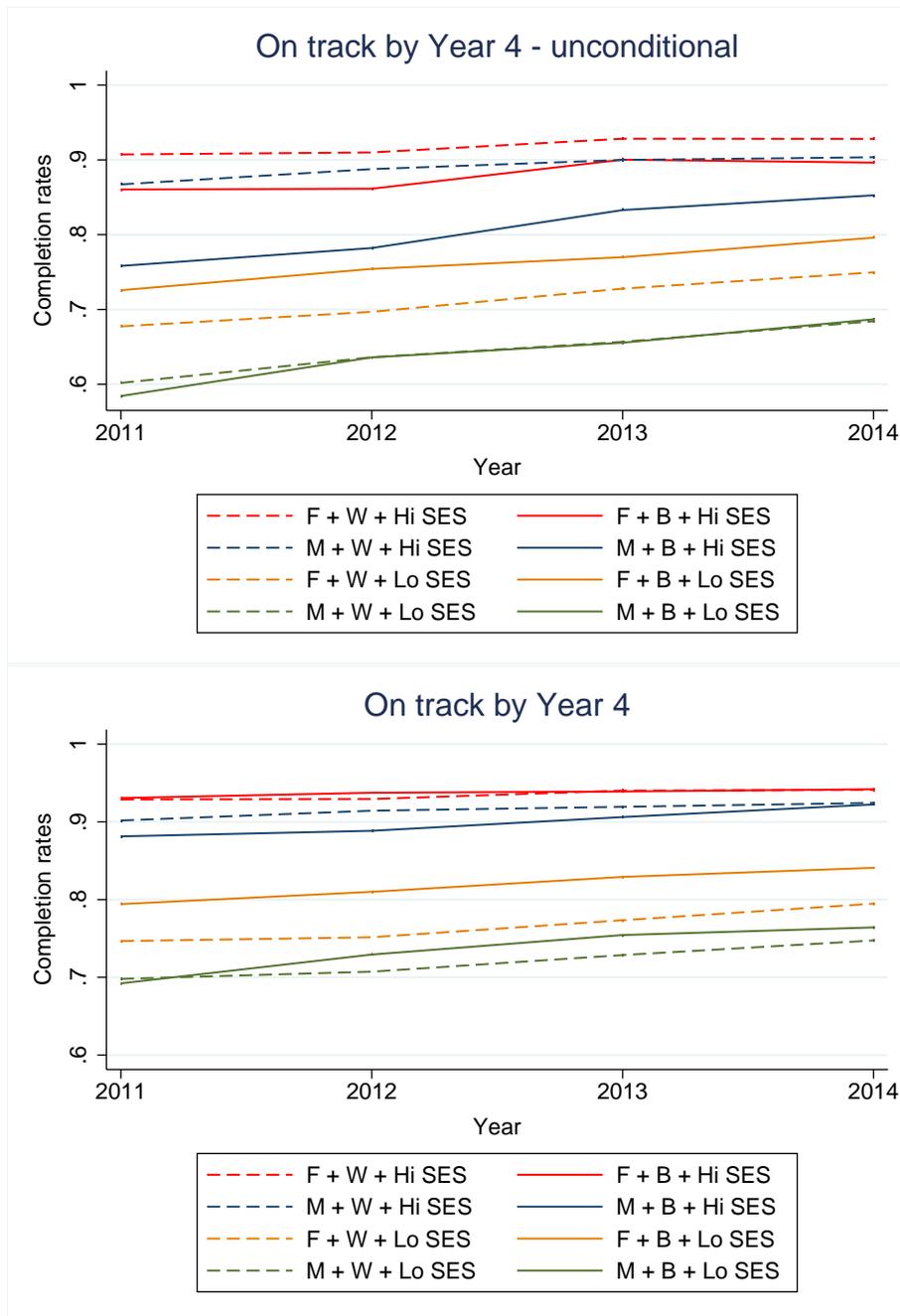


FIGURE D1: On track and graduating by Year 4

Table D4: LPM by year: Complete 12th grade

	2011	2012	2013	2014
High SES				
Black female	0.027*** (0.006)	0.034*** (0.005)	0.024*** (0.006)	0.025*** (0.005)
White male	-0.004 (0.003)	0.005 (0.003)	0.001 (0.003)	0.002 (0.003)
Black male	0.027*** (0.007)	0.025*** (0.007)	0.025*** (0.007)	0.037*** (0.006)
Low SES				
White female	-0.068*** (0.005)	-0.067*** (0.005)	-0.065*** (0.005)	-0.054*** (0.005)
Black female	0.016*** (0.005)	0.019*** (0.005)	0.012** (0.005)	0.019*** (0.005)
White male	-0.071*** (0.006)	-0.069*** (0.006)	-0.075*** (0.006)	-0.062*** (0.005)
Black male	-0.022*** (0.006)	-0.006 (0.006)	-0.008 (0.006)	0.000 (0.006)
Additional controls				
HS characteristics ^a	Y	Y	Y	Y
Skill proxies: Non-cognitive ^b	Y	Y	Y	Y
Skill proxies: Cognitive ^c	Y	Y	Y	Y
Observations	67514	67622	68715	70704
R^2	0.311	0.289	0.277	0.263

Note: Annual samples correspond to students observed in 9th grade from 2008 to 2011. Indicators for race/gender/SES categories are defined relative to high-SES white females.

^a High school characteristics are taken from a student's 9th grade institution, and include indicators for a magnet high school and being located in a city with 250K+ residents or a rural area, and continuous measures for the share of 9th grade students in each of the 8 categories.

^b Non-cognitive skill proxies include cumulative 9th grade GPA, number of absences in 9th grade, and homework completion relative to classroom peers. GPA enters non-parametrically as a set of 20 bins.

^c Cognitive skill proxies include a student's 8th grade End-of-Grade scores in math and reading and 9th grade End-of-Course scores in reading, normalized relative to statewide test scores for the year a student took a given test. All test scores enter non-parametrically as sets of 20 bins.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D5: Results by block group characteristic

	All	All	Urban ^a	Rural ^b
High SES				
Black female	0.025*** (0.003)	0.026*** (0.003)	0.028*** (0.005)	0.017* (0.009)
White male	0.000 (0.002)	-0.000 (0.002)	0.001 (0.003)	-0.001 (0.004)
Black male	0.026*** (0.004)	0.026*** (0.003)	0.028*** (0.004)	0.024** (0.010)
Low SES				
White female	-0.063*** (0.003)	-0.063*** (0.003)	-0.050*** (0.005)	-0.067*** (0.005)
Black female	0.015*** (0.003)	0.017*** (0.003)	0.023*** (0.004)	0.016** (0.007)
White male	-0.070*** (0.003)	-0.070*** (0.003)	-0.063*** (0.005)	-0.071*** (0.005)
Black male	-0.011*** (0.004)	-0.010*** (0.003)	-0.009* (0.005)	0.004 (0.008)
Additional controls				
HS characteristics	Y	Y	Y	Y
Skill proxies: Non-cognitive	Y	Y	Y	Y
Skill proxies: Cognitive	Y	Y	Y	Y
Block group median income ^c		Y	Y	Y
Observations	256292	256292	102169	57949
R^2	0.282	0.282	0.310	0.262

Note: Annual samples correspond to students observed in 9th grade from 2008 to 2011. Indicators for race/gender/SES categories are defined relative to high-SES white females.

^a All housing units in block group are classified as urban in 2010 Census.

^b All housing units in block group are classified as rural in 2010 Census.

^c Block group median income and racial composition come from the American Community Survey 2013 5-year sample.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

- Abadie, Alberto. 2003. “Semiparametric instrumental variable estimation of treatment response models.” *Journal of econometrics* 113 (2):231–263.
- Abdulkadiroglu, Atila, Joshua D Angrist, and Parag Pathak. 2014. “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools.” *Econometrica* 82 (1):137–196.
- Aguirregabiria, Victor and Pedro Mira. 2007. “Sequential Estimation of Dynamic Discrete Games.” *Econometrica* 75 (1):1–53.
- Alon, Sigal and Marta Tienda. 2005. “Assessing the “Mismatch” Hypothesis: Differences in College Graduation Rates by Institutional Selectivity.” *Sociology of Education* 78 (4):294–315.
- Altonji, Joseph G., Erica Blom, and Costas Meghir. 2012. “Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers.” *Annual Review of Economics* 4 (1):185–223.
- Ammermueller, Andreas and Jörn-Steffen Pischke. 2009. “Peer effects in European primary schools: Evidence from the progress in international reading literacy study.” *Journal of Labor Economics* 27 (3):315–348.
- Anderson, Theodore W and Herman Rubin. 1956. “Statistical Inference in Factor Analysis.” In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5. 1.
- Anelli, Massimo and Giovanni Peri. 2015a. “Gender of Siblings and Choice of College Major.” *CESifo Economic Studies* 61 (1):53–71.
- . 2015b. “Peers’ Composition Effects in the Short and in the Long Run: College Major, College Performance and Income.” .
- Antecol, Heather and Deborah A. Cobb-Clark. 2013. “Do psychosocial traits help explain gender segregation in young people’s occupations?” *Labour Economics* 21:59–73.

- Arcidiacono, Peter. 2005. "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica* 73 (5):1477–1524.
- Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. 2014. "College Attrition and the Dynamics of Information Revelation." In *2014 Meeting Papers*, 529. Society for Economic Dynamics.
- Arcidiacono, Peter, Shakeeb Khan, and Jacob L. Vigdor. 2011. "Representation versus assimilation: How do preferences in college admissions affect social interactions?" *Journal of Public Economics* 95 (1–2):1–15.
- Arcidiacono, Peter and Michael Lovenheim. 2016. "Affirmative Action and the Quality–Fit Trade-off." *Journal of Economic Literature* 54 (1):3–51.
- Aucejo, Esteban. 2015. "Explaining Cross-Racial Differences in the Educational Gender Gap." .
- Autor, David H., David N. Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2016. "School Quality and the Gender Gap in Educational Achievement." Working Paper 21908, National Bureau of Economic Research.
- Benbow, C. P. and J. C. Stanley. 1983. "Sex differences in mathematical reasoning ability: more facts." *Science* 222 (4627):1029–1031.
- Bertanha, Marinho. 2015. "Regression discontinuity design with many thresholds." *Available at SSRN* .
- Bettinger, Eric P. and Bridget Terry Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *The American Economic Review* 95 (2):152–157.
- Beyer, Sylvia and Edward M. Bowden. 1997. "Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias." *Personality and Social Psychology Bulletin* 23 (2):157–172.
- Billger, Sherrilyn M. 2002. "Admitting men into a women's college: A natural experiment." *Applied Economics Letters* 9 (7):479–483.
- Blau, Francine D. and Lawrence M. Kahn. 2000. "Gender Differences in Pay." NBER Working Paper 7732, National Bureau of Economic Research, Inc.
- Booij, Adam, Ferry Haan, and Erik Plug. 2015. "Enriching excellent students: Evidence from a gifted and talented program in secondary education." .
- Bowen, William G. and Derek Bok. 2016. *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton University Press.

- Brown, Charles and Mary Corcoran. 1997. "Sex-Based Differences in School Content and the Male-Female Wage Gap." *Journal of Labor Economics* 15 (3):431–465.
- Bui, Sa A., Steven G. Craig, and Scott A. Imberman. 2014. "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students." *American Economic Journal: Economic Policy* 6 (3):30–62.
- Cameron, S and JJ Heckman. 2001. "The Dynamics of Education Attainment for Blacks, Whites and Hispanics." *Journal of Political Economy* 109 (3):455–499.
- Canes, Brandice J. and Harvey S. Rosen. 1995. "Following in Her Footsteps? Faculty Gender Composition and Women's Choices of College Majors." *Industrial and Labor Relations Review* 48 (3):486–504.
- Card, David. 1999. "The causal effect of education on earnings." *Review of Educational Research* 73 (4):441–449.
- Card, David and Laura Giuliano. 2014. "Does Gifted Education Work? For Which Students?" Tech. rep., National Bureau of Economic Research.
- . 2016. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" Tech. rep., National Bureau of Economic Research.
- Card, David and Alan B. Krueger. 2005. "Would the elimination of affirmative action affect highly qualified minority applicants? Evidence from California and Texas." *Industrial & Labor Relations Review* 58 (3):416–434.
- Carneiro, P, KT Hansen, and JJ Heckman. 2003. "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44 (2):361–422.
- Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *The Quarterly Journal of Economics* 125 (3):1101–1144.
- Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3):855–882.
- Cattaneo, Matias D., Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare. 2016. "Interpreting Regression Discontinuity Designs with Multiple Cutoffs." *The Journal of Politics* 78 (4):1229–1248.
- Clark, Damon. 2010. "Selective Schools and Academic Achievement." *The B.E. Journal of Economic Analysis & Policy* 10 (1).

- Clotfelter, Charles T., Helen F. Ladd, and Jacob Vigdor. 2009. "The Academic Achievement Gap in Grades 3 to 8." *The Review of Economics and Statistics* 91 (2):398–419.
- Cohodes, Sarah. 2015. "The Long-Run Impacts of Tracking High-Achieving Students: Evidence from Boston's Advanced Work Class."
- Cook, Michael D. and William N. Evans. 2000. "Families or Schools? Explaining the Convergence in White and Black Academic Performance." *Journal of Labor Economics* 18 (4):729–754.
- Corcoran, Sean Patrick and E. Christine Baker-Smith. 2016. "Pathways to an Elite Education: Application, Admission, and Matriculation to New York City's Specialized High Schools." *Education Finance and Policy* :1–55.
- Cullen, Julie Berry, Brian A Jacob, and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica* 74 (5):1191–1230.
- Cunha, Flavio, James Heckman, and Susanne Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3):883–931.
- Cunha, Flavio and James J Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Non-cognitive Skill Formation." *Journal of Human Resources* 43 (4):738–782.
- Cunha, Flavio, James J Heckman, and Salvador Navarro. 2007. "The Identification and Economic Content of Ordered Choice Models With Stochastic Thresholds." *International Economic Review* 48 (4):1273–1309.
- Dar-Nimrod, Ilan and Steven J. Heine. 2006. "Exposure to Scientific Theories Affects Women's Math Performance." *Science* 314 (5798):435–435. PMID: 17053140.
- Dee, Thomas S. 2007. "Teachers and the gender gaps in student achievement." *Journal of Human Resources* 42 (3):528–554.
- DiPrete, Thomas A. and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools*. Russell Sage Foundation.
- Dobbie, Will and Roland G. Fryer, Jr. 2014. "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools." *American Economic Journal: Applied Economics* 6 (3):58–75.

- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (5):1739–74.
- Dweck, Carol. 2008. *Mindsets and math/science achievement*. New York: Carnegie Corporation of New York, Institute for Advanced Study, Commission on Mathematics and Science Education.
- Dweck, Carol S. 2000. *Self-theories: Their Role in Motivation, Personality, and Development*. Psychology Press.
- Eccles, Jacquelynne S., Janis E. Jacobs, and Rena D. Harold. 1990. "Gender Role Stereotypes, Expectancy Effects, and Parents' Socialization of Gender Differences." *Journal of Social Issues* 46 (2):183–201.
- Ellison, Glenn and Ashley Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives* 24 (2):109–128.
- Elsner, Benjamin and Ingo E. Isphording. 2015. "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." IZA Discussion Paper 9121, Institute for the Study of Labor (IZA).
- Favara, Marta. 2012. "The cost of acting "girly": Gender stereotypes and educational choices." Tech. rep., Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.
- Fryer, Roland G. and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *Review of Economics and Statistics* 86 (2):447–464.
- . 2006. "The Black-White Test Score Gap Through Third Grade." *American Law and Economics Review* 8 (2):249–281.
- . 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics* 2 (2):210–240.
- Gelbach, Jonah B. 2016. "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics* 34 (2):509–543.
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics* 118 (3):1049–1074.
- Goldin, Claudia, Lawrence Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the Gender Gap in College." *Journal of Economic Perspectives* 20:133–156.

- Gregory, A., R. J. Skiba, and P. A. Noguera. 2010. "The Achievement Gap and the Discipline Gap: Two Sides of the Same Coin?" *Educational Researcher* 39 (1):59–68.
- Griffith, Amanda L. 2010. "Persistence of women and minorities in STEM field majors: Is it the school that matters?" *Economics of Education Review* 29 (6):911–922.
- Grogger, Jeff and Eric Eide. 1995. "Changes in College Skills and the Rise in the College Wage Premium." *The Journal of Human Resources* 30 (2):280–310.
- Gunderson, Elizabeth A., Gerardo Ramirez, Susan C. Levine, and Sian L. Beilock. 2011. "The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes." *Sex Roles* 66 (3-4):153–166.
- Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19 (2):141–164.
- . 2006. "Chapter 14 School Resources." In *Handbook of the Economics of Education*, vol. 2, edited by E. Hanushek and F. Welch. Elsevier, 865–908.
- Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does peer ability affect student achievement?" *Journal of Applied Econometrics* 18 (5):527–544.
- Heckman, James J. and Alan B. Krueger. 2005. *Inequality in America: What Role for Human Capital Policies?* Cambridge, Mass.; London: The MIT Press, revised ed. edition ed.
- Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006a. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3):411–482.
- Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006b. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." Tech. rep., National Bureau of Economic Research.
- Herbert, Jennifer and Deborah Stipek. 2005. "The emergence of gender differences in children's perceptions of their academic competence." *Journal of Applied Developmental Psychology* 26 (3):276–295.
- Heyman, Gail D. [Martyna. 2002. "Gender and Achievement-Related Beliefs among Engineering Students." *Journal of Women and Minorities in Science and Engineering* 8 (1):41–52.

- Hill, Catherine, Christianne Corbett, and Andresse St. Rose. 2010. *Why so few?: women in science, technology, engineering, and mathematics*. Washington, D.C.: American Association of University Women.
- Hoekstra, Mark, Pierre Mouganie, and Yaojing Wang. 2016. “Peer Quality and the Academic Benefits to Attending Better Schools.” Tech. rep., Social Science Research Network.
- Hoxby, C. M. and G. Weingarth. 2005. “Taking race out of the equation: School reassignment and the structure of peer effects.” *Working paper* .
- Hoxby, Caroline. 2000. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” Working Paper 7867, National Bureau of Economic Research.
- Hunt, Jennifer. 2016. “Why do Women Leave Science and Engineering?” *Industrial and Labor Relations Review* 69 (1):199–226.
- Hyde, Janet S., Sara M. Lindberg, Marcia C. Linn, Amy B. Ellis, and Caroline C. Williams. 2008. “Gender Similarities Characterize Math Performance.” *Science* 321 (5888):494–495.
- Hyde, Janet S. and Janet E. Mertz. 2009. “Gender, culture, and mathematics performance.” *Proceedings of the National Academy of Sciences* 106 (22):8801–8807.
- Imbens, G. and K. Kalyanaraman. 2012. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *The Review of Economic Studies* 79 (3):933–959.
- Imberman, Scott, Adriana D. Kugler, and Bruce Sacerdote. 2009. “Katrina’s Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees.” Working Paper 15291, National Bureau of Economic Research.
- Jackson, Kirabo C. 2010. “Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago.” *The Economic Journal* 120 (549):1399–1429.
- Jacob, Brian A. 2002. “Where the boys aren’t: non-cognitive skills, returns to school and the gender gap in higher education.” *Economics of Education Review* 21 (6):589–598.
- Kinsler, Josh. 2011. “Understanding the black–white school discipline gap.” *Economics of Education Review* 30 (6):1370–1383.
- Kuncel, Nathan R., Marcus Credé, and Lisa L. Thomas. 2005. “The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature.” *Review of Educational Research* 75 (1):63–82.

- Lazear, Edward P. 2001. "Educational Production." *The Quarterly Journal of Economics* 116 (3):777–803.
- Legewie, Joscha and Thomas A. DiPrete. 2014. "The High School Environment and the Gender Gap in Science and Engineering." *Sociology of Education* 87 (4):259–280.
- Leslie, S.-J., A. Cimpian, M. Meyer, and E. Freeland. 2015. "Expectations of brilliance underlie gender distributions across academic disciplines." *Science* 347 (6219):262–265.
- Long, Mark C. and Marta Tienda. 2008. "Winners and Losers: Changes in Texas University Admissions Post-Hopwood." *Educational Evaluation and Policy Analysis* 30 (3):255–280.
- Lucas, Adrienne M. and Isaac M. Mbiti. 2014. "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya." *American Economic Journal: Applied Economics* 6 (3):234–263.
- Marsh, Herbert W. 1987. "The big-fish-little-pond effect on academic self-concept." *Journal of Educational Psychology* 79 (3):280–295.
- McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142 (2):698–714.
- Murnane, Richard J., John B. Willett, Kristen L. Bub, Kathleen McCartney, Eric Hanushek, and Rebecca Maynard. 2006. "Understanding trends in the black-white achievement gaps during the first years of school [with comments]." *Brookings-Wharton papers on urban affairs* :97–135.
- Murphy, Richard and Felix Weinhardt. 2014. "Top of the Class: The Importance of Ordinal Rank." SSRN Scholarly Paper ID 2457864, Social Science Research Network, Rochester, NY.
- Niederle, Muriel and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics* 122 (3):1067–1101.
- . 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24 (2):129–144.
- Paglin, Morton and Anthony M. Rufolo. 1990a. "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." *Journal of Labor Economics* 8 (1):123–144.

- Paglin, Morton and Anthony M Rufolo. 1990b. "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." *Journal of Labor Economics* :123–144.
- Pop-Eleches, Cristian and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4):1289–1324.
- Preston, Anne E. 2004. *Leaving Science*. Russell Sage Foundation.
- Reardon, Sean F. and Joseph P. Robinson. 2008. "Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps." *Handbook of research in education finance and policy* :497–516.
- Rothstein, Donna S. 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial and Labor Relations Review* 48 (3):515–530.
- Sander, Richard and Stuart Taylor. 2012. *Mismatch: How Affirmative Action Hurts Students It's Intended to Help, and Why Universities Won't Admit It*. Basic Books.
- Sass, Tim R. 2015. "Understanding the STEM Pipeline. Working Paper 125." *National Center for Analysis of Longitudinal Data in Education Research* .
- Sax, Linda J. 1994. "Mathematical self-concept: How college reinforces the gender gap." *Research in Higher Education* 35 (2):141–166.
- Sax, Linda J., M. Allison Kanny, Jerry A. Jacobs, Hannah Whang, Dayna S. Weintraub, and Amber Hroch. 2016. "Understanding the Changing Dynamics of the Gender Gap in Undergraduate Engineering Majors: 1971–2011." *Research in Higher Education* 57 (5):570–600.
- Solnick, Sara J. 1995. "Changes in Women's Majors from Entrance to Graduation at Women's and Coeducational Colleges." *Industrial and Labor Relations Review* 48 (3):505–514.
- Stange, Kevin M. 2012. "An Empirical Investigation of the Option Value of College Enrollment." *American Economic Journal: Applied Economics* 4 (1):49–84.
- Strand, Steve. 2014. "School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11." *Oxford Review of Education* 40 (2):223–245.
- Tienda, Marta and Sunny Xinchun Niu. 2006. "Capitalizing on Segregation, Pretending Neutrality: College Admissions and the Texas Top 10% Law." *American Law and Economics Review* 8 (2):312–346.
- Turner, Sarah E. and William G. Bowen. 1999. "Choice of Major: The Changing (Unchanging) Gender Gap." *Industrial and Labor Relations Review* 52 (2):289–313.

- Urzúa, Sergio. 2008. "Racial Labor Market Gaps The Role of Abilities and Schooling Choices." *Journal of Human Resources* 43 (4):919–971.
- Valian, Virginia. 1998. *Why So Slow? : The Advancement of Women*. Cambridge, Mass: MIT Press.
- Valla, Jeffrey M. and Wendy M. Williams. 2012. "Increasing Achievement and Higher-Education Representation of Under-represented Groups in Science, Technology, Engineering, and Mathematics Fields: A Review of Current K-12 Intervention Programs." *Journal of Women and Minorities in Science and Engineering* 18 (1):21–53.
- Vigdor, J. and T. Nechyba. 2007. "Peer effects in North Carolina public schools." *Schools and the equal opportunity problem* :73–101.
- Vogt, Christina M., Dennis Hocesvar, and Linda Serra Hagedorn. 2007. "A Social Cognitive Construct Validation: Determining Women's and Men's Success in Engineering Programs." *The Journal of Higher Education* 78 (3):337–364.
- Weinberger, Catherine J. 1999. "Mathematical college majors and the gender gap in wages." *Industrial Relations: A Journal of Economy and Society* 38 (3):407–413.
- Wiswall, Matthew, Leanna Stiefel, Amy Ellen Schwartz, and Jessica Boccardo. 2014. "Does attending a STEM high school improve student performance? Evidence from New York City." *Economics of Education Review* 40:93–105.
- Xie, Yu and Kimberlee A. Shauman. 2003. *Women in science: career processes and outcomes*. Harvard University Press.
- Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources* 48 (3):545–595.
- Zhang, Hongliang. 2012. "The mirage of elite schools: evidence from lottery-based school admissions in China." *Chinese University of Hong Kong* .

Biography

Ying Shi was born on May 2, 1986 in Beijing, China. Her research interests in the economics of education and labor economics focuses on two main objectives. The first is to understand the causes of attainment and skill gaps by race, gender, income, and immigration status. The second is to evaluate a range of education policies for their effectiveness in improving academic outcomes and closing existing disparities. Her dissertation research is supported by the American Educational Research Association Minority Dissertation Fellowship and the Horowitz Foundation for Social Policy. She received her Ph.D. in Public Policy from the Sanford School of Public Policy at Duke University in 2017. She also earned a M.A. in Economics from Duke University in 2013, a MSc. in Economics and Philosophy from the London School of Economics in 2011, and a B.A. from Rice University in 2007.