The Dissertation Committee for Chester William Polson II
certifies that this is the approved version of the following dissertation:

Essays on the Economics of Education and Human Capital

Committee:

Sandra Black Youngblood, Supervisor

Leigh L. Linden

Stephen J. Trejo

Brendan A. Kline

Jane Arnold Lincove

Essays on the Economics of Education and Human Capital

by

Chester William Polson II, B.A.; M.S. Econ.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2015

Acknowledgments

First and foremost I wish to acknowledge the aid and supervision of my Adviser, Sandra Black, along with the rest of my dissertation committee. I have grown a lot as an economist and personally in the five years undergoing the rigors of an economics Ph.D. program. This is in no small part due to the individual attention I received during the slow process of learning what makes for good economic research. I hope I can be counted as one of the department's success stories.

Next, I wish to acknowledge the support of my family and friends, without whom I know I would not be where I am today. Two loving parents, Lee and Melissa, have supported me since birth in countless ways, including letting me move back in with them for two years at the start of my graduate studies. So many of the nurturing things you have done throughout the 27 years of my life can only be described by a Labor Economist as helping to "accumulate human capital." Thank you as well to all of my friends who have kept me an interesting person during graduate school.

Finally, I wish to acknowledge the advice and friendship of my officemates, Matthew Farber and Jeffrey Denning. We spent a lot of time together, bouncing a countably infinite number of ideas off each other (some of which ended up in this dissertation), over our five years in a windowless office. Good luck to the both of you.

Essays on the Economics of Education and Human Capital

Publication No. _____

Chester William Polson II, Ph.D.
The University of Texas at Austin, 2015

Supervisor: Sandra Black Youngblood

This dissertation examines several facets of the current educational landscape in the United States and the impacts these characteristics have on individual outcomes.[1] The first chapter examines high stakes exit exams, which are pervasive in the American education system and have the ability to impact students far beyond their earned scores. This chapter considers how exit exams in Texas impact student behavior and human capital formation before the end of high school. Employing a regression discontinuity framework, I examine the impact of failing the exam the first time it is administered for students within a small window of scores around the passing threshold. Considering behavioral responses to the administration of the Texas Assessment of Knowledge and Skills (TAKS), I study the impact on students' courses taken, attendance, and a set of disciplinary actions after the exam in the final year of high school. I find that, in line with a model of motivation with heterogeneous effects, students who fail do respond through an increase in the

[1]The research presented here utilizes confidential data from the State of Texas supplied by the Texas Education Research Center (ERC) at The University of Texas at Austin. The author gratefully acknowledges the use of these data. The conclusions of this research do not necessarily reflect the opinion or official position of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas. Any errors are attributable to the author.

number of courses taken in their senior year, and find a smaller increase in disciplinary infractions. I then consider heterogeneity among student subgroups to discern whether the TAKS exam has differential impacts across different portions of the high school population in Texas.

The second chapter quantifies the extent to which test scores and demographic variables account for the differing high school graduation rates between minority and white students in Texas. There are persistent, well documented gaps in both achievement on standardized tests and high school diploma receipt between minority students and their white peers. I employ a set of linear probability models to estimate the graduation gap for students who were eighth graders in Texas from 2003-2009 and examine specific sub-populations to try to disentangle some of the factors that could be contributing to these gaps. I find that student observable characteristics, especially test scores, can account for a substantial portion of this gap, which supports estimates in the previous literature.

The third chapter of my dissertation examines how merit-based scholarships, instead of need-based financial aid, impact the decisions students make when enrolling in post-secondary education. Using the 2000 US Census data and American Community Survey data from 2001-2010, I evaluate the effect of merit scholarships in Tennessee on current college enrollment using difference-in-difference estimation. In contrast to the estimated effects of merit scholarships in Georgia, the Tennessee Education Lottery Scholarship does not seem to impact student behavior; estimates are mildly negative but not statistically different from zero considering the whole population of youth ages 16-26, traditional college enrollees ages 18-19, or older students aged 20-22. I argue these estimates are in line with many more recent findings examining merit scholarship programs. Finally, I employ a synthetic control method to compare these estimates with more traditional estimation strategies.

Table of Contents

List of Tables

List of Figures

Chapter 1

TAKS-ing Students? Texas Exit Exam Effects on Human Capital Formation

1.1 Introduction

The United States has a long history of incorporating testing as a way to evaluate student knowledge. One visible aspect of this trend in educational assessment is the implementation of "exit exams," which are an additional testing requirement compelling students to meet some minimum threshold score in addition to completing traditional high school coursework in order to be awarded a high school diploma. Because these exams impose rewards for proficiency and penalties for sub-par achievement on both the school district and individual level, these are considered "high stakes" exams. In 2012, 25 states had exit exams in place, placing 69% of students nation-wide under some sort of testing regime (Center on Education Policy [2012]). While, on paper, exit exams only impact a student's ability to obtain a high school diploma, they conceivably have the ability to impact students in far reaching ways during high school and transitioning into the labor market.

But do these high stakes standardized tests impact students only at the end of high school, influencing decisions on post-secondary education and labor force outcomes, or does a high profile exam encompassing four days of academic instruction each spring have more immediate consequences? Do the results of the first administration of the exit exam change coursework in ways that affect human capital accumulation? Or do the exams elicit other behavioral responses that

would hamper learning before the end of high school, perhaps by acting out in class or reducing the number of days of instruction? It is possible these high stakes standardized tests have direct effects on things beyond just the receipt of a high school diploma.

By examining the effect the first administration of exit exams during the spring of their junior year of high school to students in Texas, this paper attempts to answer whether these exams impact student behavior before the end of high school. I consider whether students respond to the exam by the number of courses they take, the difficulty of those courses enrolled in, and behavioral responses encompassing the number of days absent and a set of disciplinary actions. This research adds to our understanding of how these exams impact human capital investment in the short run by considering new dimensions of student response before the end of high school. I find that, in line with a model of motivation with heterogeneous effects, students who initially fail do respond through a 0.015-0.141 increase in the number of courses taken in their senior year. I also find a smaller increase in disciplinary infractions. I then address how high stakes standardized testing in Texas varies among different dimensions of students. This research furthers our current understanding of how these exams change student behavior between the first administration of the exam and the expected end of high school by examining student behavorial responses other than high school dropout and diploma receipt.

Exit exams are currently a cornerstone of the existing education system in the United States. High stakes standardized testing continues to be a policy that garners public interest and debate but is one that is hard to study. If there is some minimum standard of student knowledge necessary to ensure that individuals are prepared for post-secondary education or the labor market after high school, then these high stakes exams can serve as a test of meeting this criteria. Conversely, if the results of the exam, which happens before the end of high school, elicit changes in student behavior

before they earn a high school credential these tests could have large unintended consequences.

But students make more decisions in high school than simply to continue to attend or drop out. Additionally, while there are several retakes available for students to attempt to pass any failed portion of the exit exam before the end of high school, focusing attention on the initial administration allows for cleaner study of any behavioral response associated with the negative shock of failure. This paper uses administrative data from Texas to improve understanding of the effects of implementing exit exams on student behavior in high school under a classical human capital framework.

It is difficult to estimate the causal effect of exit exams on student outcomes because the groups of students who pass and fail are inherently different, so simply comparing means for the two groups could wrongly attribute the effect of some unmeasured difference to the exit exam regime. To address this, I use a fuzzy regression discontinuity approach to take advantage of institutional features of the TAKS exit exams as a way to examine the impact of just barely failing on student behavior in high school and human capital formation. I am able to address whether the first administration of an exit exam differentially impacts students of relatively equal ability in the courses they take and certain measures of their behavior in the classroom. I look at the number and subject of courses completed or failed, absences, number and type of disciplinary infractions reported, the number of days served for disciplinary infractions, and where the student was assigned to serve them. These variables capture new dimensions of student response both academically and behaviorally, I find that while the results vary by the specific exam segment, just passing an exam can change coursework by up to 14% of the mean for students within the bandwidth. While the impacts vary by the specific outcome measured, I also find statistically significant evidence that students at the threshold for passing the exam are varying their expended

effort in a way consistent with models of heterogeneous response to performance standards like those discussed in Lindo et al. [2010].

The paper proceeds as follows: Section 2 offers a review of the relevant literature, Section 3 explains the Texas Assessment of Knowledge and Skills in more detail; Section 4 offers a conceptual framework behind how the exit exam affects students. Section 5 describes the research design, Section 6 presents internal validity of the empirical strategy, and Section 7 offers the results. Section 8 considers heterogeneity and addresses the robustness of my findings; the final section offers possible policy considerations and conclusions.

1.2 Review of Current Literature

There is compelling research showing that high stakes standardized tests have an impact on obtaining a high school credential.[1] Ou [2010] employs a regression discontinuity research design to assess the impacts of exit exams in New Jersey on leaving high school before completion, and finds that students who just fail the exam are more likely to exit high school early, with even larger effects for minority students. Papay et al. [2010] look at students in Massachusetts under a testing regime that allows retakes of the exit exam. Also implementing a regression discontinuity design, they consider the exam's impact on repeating a grade and dropout rates as well as high school completion. The authors find that the majority of students are not impacted by exit exams, but do

[1] A large body of economic research studies the impacts of obtaining a high school diploma more generally: Heckman et al. [2008] examine the marginal returns to years of schooling, finding a higher rate of return for high school diplomas than for a post-secondary credential and an increasing rate of return over time. Obtaining a high school diploma also impacts more than just future wages; there are well documented effects of dropping out of high school on a wide range of outcomes. For example, Marcotte [2013] uses the presence of exit exams in a state along with compulsory schooling laws as instruments for examining the impact of dropping out of high school on teen pregnancy and finds a positive, statistically significant elasticity on teen childbearing.

find low-income, urban students are more likely to drop out or fail to graduate. Baker and Lang [2013] use a difference-in-difference approach to evaluate the effects of high school exit exams on high school graduation, incarceration, employment, and wages. They exploit the staggered timing of the implementation of exit exams across states using the Current Population Survey (CPS), and find that more challenging standards for exit exams reduce graduation and increase incarceration rates coinciding with an increase in GED receipt. Overall, this research paints a picture of exit exams having a small but measurable impact on high school completion.

Other economic research tries to assess the importance of exit exams in an individual's long term well-being, including post-secondary educational attainment and labor force outcomes. Martorell [2004] considers the impact of exit exams in Texas on students both in high school and after graduation. He focuses on students who, after multiple retakes, must take the exam one last time the spring of their senior year of high school in order to graduate. Martorell, too, finds that students who barely fail the exam are less likely to earn a high school diploma. Furthermore, while dropouts are more likely to earn a GED, they do not make up the decrease in high school diploma receipt, are less likely to attend post-secondary institutions, and —of the individuals that are strongly attached to the labor force —earn less initially, although the earnings effect lessens over time. Considering just the impact of exit exams on earnings, Clark and Martorell [2014] use the impact exit exams have on high school diploma receipt to test a signaling model of human capital. Employing a fuzzy regression discontinuity design, they look at the impact that just passing the exams during the final attempt senior year on labor force outcomes, finding little evidence of signaling effects of a diploma. Given these results, it is plausible to think additional unstudied impacts of high school exit exams will occur before the end of high school and not after.

Some research has been done considering the effects of these exams on student outcomes

other than high school graduation. Zieleniak [2013] uses a change in the exit exam administered in the second largest school district in California to estimate some positive motivational effects on test scores for students of all abilities. Employing a triple-difference strategy, Zieleniak considers the entire distribution of students. He finds exit exams increase the dropout rate for lower ability students but increase math test scores for students in courses with substantial overlap with the material covered on the exam. I instead focus on students right around the threshold who would be most affected by the exit exam; I also consider a different dimension of high school achievement. Rather than looking at how the imposition of a standardized test in a course affects performance on both test and in the classroom, I look at subsequent coursework in subject areas after the results of the first exam are announced. I am additionally able to observe students in the entire state of Texas, a large and important state for educational standards, estimating a more widely applicable local average treatment effect.

Reardon et al. [2010] also use data from four large school districts in California to estimate the effect of exit exams on academic achievement, subsequent coursework, persistence to the 12th grade, and graduation. They consider the first administration of an exit exam in the tenth grade and examine English language arts scores one year later, as well as types of math courses taken the year after the exit exam, but find very little effect on behavior. I am able to consider the number of courses students take rather than year-over-year improvement in the subject area, providing a clearer picture of the material students encounter before the end of high school. I additionally consider other types of course choices, including Advanced Placement and International Baccalaureate credentials and additional subject areas as ways students respond to the outcome of the exams. Furthermore, I am able to look at changes in disciplinary outcomes. This allows for study of an additional dimension of student behavioral response to the results of an exit exam and

to test for the presence of discouragement effects. This research continues to expand our knowledge of how students respond academically to exit exams by looking at their course-taking patterns more generally, and also examines an additional behavioral response through disciplinary actions. I find that students are responding through an increase in the courses they take their senior year, and this continues to sharpen our understanding of the role exit exams play in human capital formation during high school.

1.3 Texas Assessment of Knowledge and Skills

Since 1980 Texas has had some form of standardized exam in place (Amrein and Berliner [2002]). This research considers the Texas Assessment of Knowledge and Skills (TAKS), a statewide standardized test with an exit exam component that was initially administered in 2003.[2] The exams are administered statewide on four consecutive days, one day for each subject: English, mathematics, social studies, and science. The guidelines for administering the test are created in such a way to minimize distractions for test takers and to remove any unfair advantages for students.[3] The tests are untimed, and students are allowed as much time to respond to every question as is necessary.[4] There are 57 questions in the English language arts section, 60 questions on the mathematics section, 55 questions on the science portion of the test, and 55 social studies

[2]It replaced the Texas Assessment of Academic Skills (TAAS), which also incorporated an exit exam component, with a substantive change being the addition of a subject test in social studies as well as changes in standards. The TAKS tests were then replaced in the spring of 2012 by the State of Texas Assessment of Academic Readiness (STAAR), which is not a directly comparable regime because it involves end of course examinations in specific high school courses rather than exit exams (Texas Education Agency [2014c]).

[3]Guidelines require, among other concerns, no talking, no cell phones, and covering up any information around the classroom that could offer aid on a test.

[4]While not a requirement that administration of the exam continue beyond school hours, districts are allowed to offer students even that additional time.

questions.[5] This means students are capable of scoring integer values between 0-73 for ELA, 0-60 for math, 0-55 for science, and 0-55 for social studies. Because the difficulty of the specific exam administered varies from year to year even trying to keep the standards constant, raw scores are then converted to a scaled score, which is directly comparable between years. The minimum standard for passing is 2100 each year, and students who score 2400 or above achieve "commended performance."

The exit level TAKS are first administered the spring of a student's junior year of high school.[6] Because students are required to pass all four subject sections of the exit level exam, several retakes are offered for students who do not pass all four during the first administration. Including the initial administration the spring of junior year, students have five opportunities to take the exam before their "on time" graduation date.[7] It is compulsory for any enrolled student to retake any portion they have not already passed for every administration they are present for.[8]

The guidelines for administration set forth very clear procedures stating what test administrators can and cannot do or say with regards to the testing materials, student questions, and the answer sheets. Tests and answer sheets are kept in a locked storage locker except when actively

[5]Students face the additional requirement of scoring at least a 2 on the essay portion, which is factored into the overall ELA score by $((4 * EssayScore) + MCScore)$, which is out of 73.

[6]Students begin taking TAKS exams in the third grade, and grade level exams continue for every grade through 10th grade.

[7]These retakes are administered in July after the student's junior year, October of senior year, February or March of senior year (depending on the year), and April of senior year.

[8]In addition to the state mandated curriculum tested on the Texas Assessment of Knowledge Skills (TAKS) exam, two other formats–accommodated and modified–exist. The TAKS Accommodated form is for special education students who meet additional eligibility criteria and assesses the same curriculum standards but offers changes in formatting and permits accommodations the student may require. The TAKS Modified is for special education students that satisfy standard eligibility requirements and assesses modified academic standards. While the TAKS-M covers the same grade-level content, the format and design of the test has been changed. Because students who take either the TAKS-A or TAKS-M and score right around the cutoff will not compare to the marginal student on other observable characteristics, individuals taking the TAKS-A or TAKS-M are omitted from subsequent analysis.

administered. At the end of each testing administration (daily) all materials must be collected by the campus administrator, boxed up and mailed to the district coordinator, who then forwards them on to Pearson, a third-party corporation that specializes in high-stakes tests. They are not scored by the students' teachers (TEA 2014). As a result, it is unlikely that the TAKS test is subject to systematic manipulation by administrators.

1.4 Conceptual Framework

Exit exams could have important intermediate effects on human capital investment beyond granting high school credentials to students. One possible mechanism is through discouragement effects on students, whether negative feedback can induce students to stop trying, quit, or drop out. Papay et al. [2010] consider discouragement effects on poor performers on exit exams through dropping out of high school. But student discouragement could surface in ways less dramatic than dropping out of high school completely. Students could take less challenging classes, or fewer classes in general, especially in the subject matter covered on the exit exam. Discouragement could also manifest itself in an increase in the number of absences during the school year. All of these are testable predictions. Examining whether the number of advanced, AP and IB courses taken increases upon passing the TAKS test, or whether the number of courses in more challenging subjects (like math and science) decreases after failing a portion of the exam could display student discouragement. Discouragement effects could also be identifiable through a change in the attitude of the student in the classroom. While this is not directly observable (no surveys of student sentiment or teacher reviews are available) one measurable way this could surface is through an increase in disruptive behavior in the classroom, resulting in an increase in disciplinary infractions or suspension.

Another possible effect would be if the students just below the cutoff were motivated to work harder in order to pass at the next administration of the exam. Bénabou and Tirole [2000] offer a more formalized model of individual response to the imposition of a performance standard. Some agents respond by working harder while some other individuals stop exerting any effort. Lindo et al. [2010] apply this in an educational setting by considering students near a GPA cutoff for academic probation in college in Canada. This model of behavior is equally applicable for students in high school facing a firm threshold for passing the exit exam requirement. Consider utility maximizing students, whose cost of effort is decreasing in student quality, that have three options: they can expend zero effort (and get nothing in return), they can expend a small amount of effort for a small cost in order to achieve a small gain, or they can exert a large amount of effort for a larger gain. In this framework, the lowest quality students will decide to exert no effort and drop out of school, students above a certain quality threshold will reap the highest gains by exerting the high effort level, and the intermediate quality students will exert the low level of effort.

The threshold for passing the exit exam can be considered leveraging an additional cost for the low threshold of effort because without expending additional effort, students in this category will be unable to obtain a high school diploma, unlike their peers just above the threshold. Students who are above the threshold for failing the exam still face the same three options; they can exert no effort and drop out, exert a high level of effort and achieve the academic gains that come with it, or exert a low level of effort and achieve a lower level of academic success. Students who are below the threshold, on the other hand, now face a low effort option that is too costly to be a viable option. They will be unable to obtain a high school diploma with their current level of effort; their choice set is in essence restricted to exerting no effort by dropping out of school, or exerting a higher level of effort and achieving higher academic gains required to secure their high school credential.

There are several testable implications for student behavioral responses coming from this model. Because students below the passing threshold face a choice set that encourages either no effort or high effort, we could expect to see change in effort at both the margin of dropping out of school and within the classroom. However, for the students who remain, exerting more effort to compensate for failing the exam might surface empirically in the number or quality of courses students take. This change in effort levels might also be reflected into disruptive behavior in the classroom caused by failing an exam. If students are experiencing discouragement effects, we would expect to see an increase in disruptive behavior, both in the frequency and severity of punishment. On the other hand, if students who fail but decide to remain in school are more focused in their studies it might be the students who have passed the standard and are exerting a lower level of effort that act out more. It is not readily apparent whether failing an exit exam should increase or decrease the frequency of any specific type of disciplinary infraction. Do students who fail become more aggressive and are more likely to be disciplined for violent behavior, like assault, fighting, or bringing a weapon to campus? Are they more or less likely to use drugs or alcohol? None of the models referenced above offer a clear prediction, making this an empirical question.

1.5 Research Design

1.5.1 Empirical Motivation

The motivating concern for this empirical strategy is that the group of students who pass or fail the exam are fundamentally different, and so we cannot simply compare the sample means of the two groups. Comparing the whole group of students who pass to the whole group of students who fail would attribute any observable differences to the effect of the TAKS test when many other variables could be causing this gap. However, by selecting students right around the cutoff,

I identify a group of students who are observably similar, some of whom passed and some of whom did not. One could argue assignment of passing or failing the exams is not due to any underlying bias for this group of students who otherwise look the same. As a result, the difference in means between the two groups is an estimate of the impact of the exams on all of the outcomes of interest described above: dropout, courses taken, and disciplinary infractions. In order for this regression discontinuity framework to be valid, there needs to be a clear cutoff for treatment status, an inability to manipulate treatment status, and smooth densities around the cutoff. Section 1.6 on Internal Validity addresses these concerns at length and shows that students do in fact seem to be similar on observable characteristics and unable to manipulate treatment.

While students who fail any portion of the TAKS exit exam are offered up to four retakes before the spring of their senior year, I focus on the first administration of the exam to see the effect of failing initially on human capital decisions during the senior year of high school.[9] I also consider each of the four sections of the exam individually rather than create some sufficient statistic for whether the student met the requirement for all sections in a given administration. Considering the four sections independently captures another important dimension of variation for outcomes before the end of high school, especially when considering specific classes students take after the results of the first exam. It is reasonable to expect failure of different subject tests would differentially affect the courses taken.

[9]Other research (Clark and Martorell [2014], Martorell [2004]) focuses on the group of students that find themselves retaking the exam the spring of their senior year as a "last chance" administration before their expected high school diploma receipt date because the results of this administration are most binding for high school diploma receipt. I can and do replicate this trend of an increasing importance of the results of the TAKS exam on high school diploma receipt. Because all students in public school must take the exam at least once, looking at the first administration rather than subsequent ones yields local average treatment effects for a larger segment of the population than for later test group. The initial results may also come as a larger shock to estimate behavioral responses than later administrations.

1.5.2 Empirical Specification

To look at the effect of failing a single section of the TAKS test, I employ fuzzy regression discontinuity methods to exploit the break in passing status. I employ a fuzzy regression discontinuity technique because while there is a discontinuity in the probability of a student being classified as "meeting the standard" at the minimum cutoff score, it does not predict this perfectly. The percentage of students being classified as meeting the standard for each possible test score can be seen in Figure 1.1. For three subjects, math, science, and social studies, once students score a minimum score 100% of them are classified as meeting the standard. However, even among the group of students who do not take an alternate form of the exam or have a waiver for special education status, some are still being classified as meeting the standard with scores below the minimum score threshold. For the English portion of the exam, some students score above the minimum but do not also achieve the required 2 on the writing section and so still do not meet the standard. For all subjects of the exam, the discontinuous changes in student status are some evidence towards the argument that a fuzzy regression discontinuity approach is justified.

More concretely, I use the minimum passing score threshold as an instrumental variable for meeting the standard controlling for functions of the exam score. I do so by estimating:

$$Y_i = \alpha_0 + \alpha_1 MET_{ti} + \alpha_2 Score_{ti} * MET_{ti} + \alpha_3 Score_{ti} * (1 - MET_{ti}) + \epsilon_{ti} \tag{1.1}$$

$$MET_{it} = \beta_0 + \beta_1 FAIL_{ti} + \beta_2 Score_{ti} * FAIL_{ti} + \beta_3 Score_{ti} * (1 - FAIL_{ti}) + \gamma_{ti} \tag{1.2}$$

where MET_{ti} in Equation 1.1 is an indicator for whether individual *i* did not meet the TAKS exit exam standard as reported by the TEA for subject *t*, and includes an interaction between score and an indicator for met standard, MET_{ti}, on both sides of the threshold.[10] Here α_1 is the parameter of interest, the effect of meeting the exit exam minimum standard in a given subject *t*: English, mathematics, science, and social studies. For all empirical estimates I report robust standard errors clustered by test score.

In Equation 1.2, the first stage, $FAIL_{ti}$ is an indicator for whether the student's score was below the minimum threshold for meeting the standard as reported in the TEA documentation, which is also interacted with the running variable of the student's score on section *t*. *FAIL* must be orthogonal to γ_{ti} and also satisfy the exclusion restriction in order to be a valid instrument. This will be true assuming student test scores near the cutoff are uncorrelated with the error term. Moreover, the passing status based on student score around the cutoff must only affect the likelihood of meeting the TEA standard and not have a direct impact on any of the outcomes of interest. This seems to be the case in practice as students are not reassigned to classes based solely on their TAKS score on a given section, nor do they suddenly face different codes of conduct or rules regarding attendance and absences from school. I will present evidence that both of these assumptions are satisfied later.

In estimating the effects of exit exams on student behavior, I tried several different func-

[10]This polynomial was estimated up to a quartic in student score with an interaction for meeting the standard. Estimated effects remain fairly stable for higher-order polynomials. Moreover, Gelman and Imbens [2014] encourage researchers not to use higher order polynomials and instead use local linear or other smooth functions. I also calculate Akaike's Information Criterion (AIC) for each model and use this test to select the preferred specification. Jacob et al. [2012] point out that the AIC can be thought of as a measure of the relative goodness of fit between models, one based on the trade off between variance and bias. For the majority of outcome variables of interest among all subjects of the TAKS, the simple linear model is the model that minimizes the AIC, which is logical because of the relatively narrow bandwidth and smooth underlying trend for the scores.

tional forms and estimation methods. In choosing my primary specification, I draw on the work of Lee and Card [2008], who suggest that for data of a discrete nature such as test scores a parametric model is more appropriate than nonlinear methods. Estimates are fairly stable up to a quartic polynomial term in test score. My preferred specification, and all results in the following tables are using a linear regression discontinuity with a bandwidth of five points. This five point bandwidth is roughly 13% of possible English score range, 16.6% of possible math scores, and 18% each of possible science and social studies scores. I also present results showing the sensitivity of my conclusions to these choices.

1.5.3 Dataset

Data for this project is from the Texas Education Resource Center (ERC), which houses administrative data from the Texas Education Agency (TEA), the Texas Higher Education Coordinating Board (THECB), and the Texas Workforce Commission (TWC). The main sample for the analysis consists of ten cohorts of individuals who took at least one TAKS exit exam the spring of their junior year during 2003-2012. The TEA keeps a separate file for each administration of the exam during the school year by exam level and whether it was a retake or not. I combine these files across years at the individual level to have the student's complete test-taking history. In addition to reporting the raw and scaled scores for each section, the TAKS files also contain information on whether the student took an alternative form of the test, had a special accommodation, or had a waiver for any other official reason.[11,12] As noted earlier, I remove students with a waiver or

[11] This includes taking the TAKS in braille form, in Spanish, on a computer, or with a different accommodation for documented learning ability.

[12] For some students, the records indicate that the student failed the exam but that their score was re-coded to zero. Because these are not true zeros, these observations are omitted from the sample.

special accommodation. A final variable used for analysis is the state-reported "Met Standard" indicator, which is 1 if the state records the student as not meeting the standard set for their year and administration. While it is possible for students to come back and attempt any remaining exit exams after the end of their senior year of high school, because my sample of interest is students before their expected high school completion I only keep the first five administrations of the TAKS exam for a given student to focus on the time period up to expected graduation date.

The TEA records two student scores for each subsection of the test for each administration of the exam: a raw score (the number of questions a student got correct) and a scaled score (an affine transformation of the raw score that makes tests easily comparable across dates). Because the TAKS test can vary slightly in difficulty and length, the scaled score reflects a consistent cutoff across subsections that is pegged to a number of correct answers on the test section itself.[13] These corresponding cutoffs for the raw scores vary slightly over time and between subjects, but give the impression that the score cutoff was in fact very consistent over time.[14] Because the density of the scaled score is non-smooth, I instead use the raw score to determine whether a student passed or not, adjusting scores on each section up or down for a given test wave so each section has a single passing threshold for all administrations of the test.[15] This has the added benefit of being able to discuss student performance in the number of questions answered correctly. The TAKS score constitutes the running variable for the empirical strategy described above.

The Texas Education Agency also curates very detailed information on students during

[13]For all administrations and sections of the TAKS test during the period of observation, the cutoff score was 2100.

[14]The raw cutoff scores were either 29 or 30 for science, ranged between 41-44 for English language arts, were between 32 and 34 for mathematics, and were either 27 or 28 for social studies.

[15]My adjusted raw score cutoff points are: 30 for science, 44 for English language arts, 33 for mathematics, and 28 for social studies.

their high school career. Student demographic characteristics from the first time the individual is observed in high school are incorporated into my sample.[16] These include campus and district attended, gender, age, race, if they are an English language learner (ESL) or limited English proficiency (LEP), whether the student is eligible for free and reduced lunch (FRL), and whether the student is a special education or gifted and talented student. A separate file containing graduation information and type of diploma received is also incorporated for each individual. Throughout the period of observation, the state of Texas offered three types of diploma of differing prestige, all requiring the successful completion of the exit exams. As diploma receipt is not a main outcome of interest, this research treats all these types of high school diploma as equivalent and pools them into one diploma receipt variable. It also offered specific degrees for students who were exempt from the TAKS exam; any student who obtained one of these alternate diplomas is omitted from the sample. Students who show up in either the enrollment and demographic files or the graduation files but do not have any TAKS scores associated with them are omitted from the sample.

I also merge in data on high school coursework; while the actual course grade is not available to the researcher, the file does contain information on which courses a student enrolled in each semester, whether it carried a special status such as advanced or AP or IB course, and whether the student passed or failed the course. I use records provided by the TEA to categorize each course as a mathematics class, an English language arts class, science, or social studies. I consider whether a student successfully completed a given course in a semester by interacting the course with whether the student passed, then sum the number of successfully completed courses by subject across an individual's senior year of high school. Because failing a course in high school could be considered

[16]This is freshman year for the vast majority of students, but for students who transfer into public school or move into the state later in high school this information is from the first year they are a high school student.

another dimension of student effort, I also consider the total number of failed semester courses as an outcome. As students who fail the TAKS and drop out of high school before the end of their senior year take fewer courses, these observations are left in even though an observed decrease in subsequent coursework is partially mechanical in this instance.

Another possible response would be some form of discouragement effect observable in instruction time; perhaps failing the TAKS test the first time makes students less invested in school and more likely to skip class more frequently. In order to examine the impact of failing an exit exam on high school attendance I sum the number of days absent across the student's senior year, which are recorded yearly.

I also create several possible behavioral outcomes from the TEA disciplinary files. The first outcome is the number of days of disciplinary action assigned to students. Each infraction is recorded separately by student by year, so in order to look at the aggregate effect I sum the number of days of disciplinary action served across instances by year, and then aggregate this number to record the total number of days assigned during their senior year.

Additionally, two of the more common punishments assigned are "in-school suspensions" (*ISS*) and "out of school suspensions" (*OSS*), which remove the student from the classroom and have a greater potential to impact human capital accumulation. I create a binary variable that coincides with each disciplinary infraction recorded to indicate whether the punishment was in-school or out of school suspension, and then sum the number of instances for each student to create a variable that records the total number of times a student was assigned to one of these punishment times during their senior year.

I am also able to observe in the TEA disciplinary files the reason for disciplinary actions,

and examine several different types of infraction for any systematic change in frequency in line with the first TAKS administration. The TEA files are fairly detailed in the type of infraction incurred, so I often aggregate related reasons for violation to gain a clearer overall picture of different types of behavior. I define *Drugs* infractions to be any instance of having controlled substances or drugs, alcohol, or tobacco, including any felonies for any of these substances. *Weapons* indicates whether a student incurred an infraction for having a knife, gun, club, or other prohibited weapon. The *Assault* variable aggregates instances of assault, aggravated assault, and sexual assault both of district employees and non-district employees. The TEA differentiates between several types of truant behavior: three days of truancy, ten days of truancy, failure to enroll in school, or whether the parent contributed to it—all of these are jointly considered *Truant*. The *Fight* variable records an instance of fighting on campus. Finally, a vast majority of instances are simply recorded as "violated local code of conduct," and because of its prevalence these are considered as well. Because students who fail the TAKS exam and drop out of high school will not be written up for any disciplinary infractions, these individuals are excluded from the disciplinary analysis.

Table 1.1 reports summary statistics of important observable characteristics for students in the sample. Column (1) reports the mean for every student with a valid TAKS score and no waiver or special accommodation in the ten year observation period. Sample means for the group of students within the five-point bandwidth with TAKS scores below (Column (2)) and above (Column (3)) the cutoff for meeting the standard are also reported. While there are small differences in the racial makeup of the sample on both sides of the cutoff, the two groups appear to be fairly similar. Column (4) reports the difference in means for the populations of students below and above the cutoff score. While the t-statistics signify statistical differences in the means, the mean differences themselves are very small. Due to the densities of test scores as shown in Figure 1.2, more stu-

dents are within the five-point bandwidth but above the threshold than below, as is reflected in the number of observations for each category.

1.6 Internal Validity

To verify that my regression discontinuity research design is valid, I first need a clear jump in treatment status. For each subject area and each administration of the test, there is a clearly published cutoff score below which a student is said not to meet the standards set (fails) and above which the student has met the standard and is able to graduate. Furthermore, these raw scores are scaled to a 2100 passing standard after each time the test is administered, so they are directly comparable.

Second, students do not seem to be able to manipulate their treatment status on the TAKS exit exam. While educators would hope that more effort exerted by the student will result in a higher score, students have no way of guaranteeing exactly what their score will be. Lee and Lemieux [2010] point out that even when individuals have some influence on treatment, if they are not able to precisely manipulate assignment then the variation observed around the cutoff is as good as random. Furthermore, while not private information, the number of questions answered correctly in order to pass the TAKS exam is not widely publicized, so it is difficult for students to know what target score to aim for. The fact that students are unable to perfectly achieve the score they would like satisfies the second assumption of inability to manipulate the treatment status for a regression discontinuity approach to be appropriate. Students do learn exactly how well they did on each portion of the exam after the fact. Appendix Figure A.1 displays an example of the score report the student receives. It clearly displays their score on each section and how close they were to the minimum threshold.

While the design of the TAKS exam appears to rule out the possibility of perfect manipulation of treatment status by the students, direct examination of the densities of both the treatment variable and any covariates allow for a partial test that these conditions are satisfied. A discontinuity on either side of the pass/fail cutoff could suggest that individuals do have some control in their treatment status. Figure 1.2 shows the distribution of scores for the entire possible range of scores for each exam. Again, the distribution of scores looks very smooth. This can be taken as partial evidence that the densities of the TAKS scores are smooth around the cutoff for each subject.[17]

In a similar manner, graphs of the densities of any observable covariates should also be smooth around the cutoff to support the idea that students are assigned to their treatment status with roughly equal probability and that there is not some other observable factor that is driving any estimated result. Control variables available from the Texas Education Agency are gender, free and reduced lunch receipt, being an English as a second language student, and a vector of race variables. For all of these, plotting the frequency by TAKS score looks smooth through the cutoff value. A selection of these can be seen in Figure 1.3 for the mathematics section of the exam. Control variables look similar and smooth for the other portions of the exam as well. Taken together, this can all be seen as suggestive evidence that it is not a change in some other observable characteristic or manipulation on the part of the student that is driving any effects from the TAKS exam and it is appropriate to employ a regression discontinuity in this situation.

Finally, I also check that the conditions of the regression discontinuity are satisfied by looking at the predicted values that result from regressing outcomes of interest on the vector of

[17] McCrary [2008] offers a formal test for discontinuities in the density of the running variable that are sometimes employed in regression discontinuity research designs. However, the discrete nature of the running variable in this case violates the central assumption of continuity in the McCrary test, producing invalid results; the test is not appropriate in this setting.

observable controls. Figure 1.4 does just this by calculating the predicted values for the regression of high school diploma receipt on the vector of covariates, averaging the predicted value by test score, and plotting them. Figure 1.4 plot (a) shows the average fitted value by test score for the social studies section, plot (b) displays average fitted values for math, plot (c) shows average fitted values for English, and plot (d) reports average fitted value by science score. Again, the graphs all look fairly smooth through the cutoff score, suggesting that there is not some jump in the observable characteristics that occurs around the cutoff that could instead be driving any estimated results.

1.7 Results

Table 1.2 reports the estimates of the impact of achieving a failing score have on the met standard status reported by the TEA, the first stage regression Equation 1.2. For all four subjects of the TAKS exam, the instrument is positive, statistically significant, and has a large F-statistic. However, simply achieving a score below the passing threshold does not affect students in a direct sense before the end of high school; it does not change a student's learning environment or in some other measurable way impact the outcomes of interest. Because the exclusion restriction seems satisfied, this appears to be a valid instrument.

Table 1.3 Panel A reports the probability of dropping out of high school and reports the impacts of failing the TAKS exam for a given subject on the number of subsequent courses in that same subject area during senior year in Panel B.[18] Column (1) reports the estimated impact of

[18]It is also worth considering how failing a segment of the TAKS exam impacts high school diploma receipt during subsequent administrations. Table A.1 reports the reduced form effect of the TAKS exam on high school diploma receipt for the initial administration and each of the four subsequent administrations through the expected graduation date. Each column reports a given administration, moving right from the initial administration that all high school

the English portion of the TAKS test, Column (2) the mathematics section, Column (3) the social studies section, and Column (4) the science section. Panel A reports the change in the probability of officially dropping out of high school before receiving a diploma. For both the math and science portions of the exam, consistent with the motivational framework, I find that students who just fail the TAKS exam are 1-3% more likely to drop out of high school. This is statistically significant at the 0.1% level. Just failing the English or social studies portion of the exam, however, is associated with a smaller but still statistically significant decrease in the probability of dropping out. Panel B shows consistent effects of exit exams on the number of courses taken; these results are graphed in Figure 1.5. A coefficient of 1 for any estimate reported would imply a one-semester increase in courses due to passing the exit exam. In all cases, the estimated effect of just failing a given subject of the exit exam increases the number of subsequent courses taken, which is statistically significant for English, mathematics and science. As is consistent with Table A.1, the two subjects that impact high school diploma receipt —science and math —also have the largest effect on coursework before the end of high school. The mean number of courses taken is reported in each panel of Table 1.3 and students are taking between one and three more semesters of coursework

juniors take across subsequent retakes to the "last chance" exam for seniors in Column (5). Three major conclusions stand out from the table. First, as also documented in previous research on older testing regimes, exit exams have a growing impact on high school diploma receipt over a student's career (Martorell [2004], Jepsen et al. [2012], etc.). Moving across each panel, it is clear that for all subjects in varying magnitudes, the results of the exam matter more for high school diploma over time. Second, by the end of high school these exams are having large effects on the portion of students who find themselves taking a last chance exam. Students who fail the mathematics portion of the exam are 20.9% less likely to obtain a high school diploma; students who fail the science portion are roughly 16% less likely to obtain a high school diploma. Third, some portions of the exam have a larger impact on students than others. Mathematics moves from having a small but statistically significant impact in the initial administration to the largest effect, at 20%, by the last chance administration; social studies displays no clear trend across the subsequent administrations and is not statistically significant for even the last chance exam. The simple percentage of degree receipt of the last chance test group is graphed for each section of the exam in Figure A.2. This final fact also suggests it is worth examining the portions of the exam independently to see if this general trend of some portions being more binding than others is apparent in other outcomes as well.

depending on the subject. Compared to those individuals who passed the TAKS exam, just failing the exam results in more courses in mathematics, science, social studies, and English, which is consistent with the motivational framework story of students just above the threshold being able to expend less effort than students below it.

While students seem to respond to failing the TAKS test by taking more courses in subjects tested on the exam, it is possible that the requirement induces a differential effect on effort through the difficulty level of the course. Table 1.4 considers the impact of exit exams on whether a student took courses designated as advanced (Column (1)) and Advanced Placement (AP) or International Baccalaureate (IB) (Column (2)). While the coefficients themselves are smaller, they also suggest a story of increased effort levels. English and social studies report very small, statistically insignificant results that both include zero in their confidence internal, suggesting students who fail the exam are not experiencing some negative feedback effect which motivates them to take less challenging material. There is a more pronounced inducement of extra effort for the science and mathematics portions of the exam; both sections report positive, statistically significant impacts on advanced courses taken the senior year of high school. Students in danger of failing the TAKS test are towards the lower end of the educational distribution, which is evident here in the low mean number of advanced courses the students take (students are taking less than one course in each subject area during the senior year of high school). Column (2) tells a similar story for Advanced Placement and International Baccalaureate courses. There might be a small discouragement effect for English, which reports a statistically significant decrease in AP and IB coursework, but the other three portions of the exam again suggest students are switching into more challenging courses. Considering all of these estimates reported in Columns (1) and (2), students seem to disproportionately take more advanced courses upon failing a portion of the TAKS, which can be

seen as students exerting extra effort.

Table 1.4 also reports the impact of the first administration of the TAKS exam on the number of failed courses in high school and the number of days absent. For the English, mathematics, and science portions of the TAKS exam in Column (3), students are actually less likely to fail a course. Failing a single course in high school carries a much lower penalty than failing a portion of the TAKS exam; while a student must still meet a overall GPA and credits requirement, a high school credential will not be withheld because of a failed math course unlike a failed math portion of the exam. This again can be taken as evidence that students below the minimum threshold exert more effort, which results in fewer failed courses. Column (4) reports the number of days a student was absent. There is an estimated 0.49 decrease in the number of days absent for students who just barely fail the English portion of the TAKS exam and a 0.44 day decrease associated with science; the other subjects report statistically significant increase in the number of absences of roughly equal magnitudes. These results can also be seen in Figure 1.6. The TEA records do not differentiate between excused and unexcused absences, but Column (4) suggests that while any change is small, there is not as clear a pattern of student response through attendance as there is through coursework or course difficulty.

As discussed above, another way a discouragement effect could surface is through changes in student behavior inside or outside the classroom that result in disciplinary infractions, harming the formation of student capital. The first panel of Table 1.5 reports estimates for the number of days of disciplinary action actually completed by the student. For every subject of the TAKS exam except social studies, failing is associated with an increase in the days of disciplinary action actually served. This is by and large consistent with students just below the threshold displaying discouragement effects and possibly acting out as a consequence. The second panel reports the

number of suspensions that were assigned for a student's disciplinary infractions. Of these, magnitudes are much smaller and do not display a consistent trend. Perhaps even though students who fail seem to experience more disciplinary infractions they are not offenses punishable by suspension, or perhaps this is not a dimension of student response that experiences a large change due to the TAKS exams. Taken as a whole, Table 1.5 suggests that just failing a section of the TAKS exam does result in a small increase in the number of days of disciplinary infractions, which could translate into more disruptive behavior in the classroom because students display discouragement effects.

Another way we could observe a behavioral response to the first administration of the TAKS test that would impact the formation of human capital is in the types of disciplinary infractions committed by students. Table 1.6 reports estimates for the six major types of disciplinary infractions described above in Section 1.5.3. Across the board the coefficients are small and do not tell a consistent story of students shifting disruptive behavior in response to the exams. Column (1) reports the impact of failing a given portion of the exit exam on being disciplined for having drugs or alcohol. Between the four subjects coefficients are both positive and negative. Moreover, a coefficient of one would suggest one or more incidences of drugs as a consequence of the exit exam, and both the mean as well as the estimates are all well below one. Column (2) reports the effect of just failing the TAKS on being disciplined for fighting. The coefficients are still very small and both positive and negative depending on the specific subject, which does not tell a consistent story of exit exams having an impact on incidence of fighting. Two more outcomes linked to violence, having a weapon (Column (3)) and committing assault (Column (5)) also report small point estimates on both sides of zero, suggesting no discernible effect of exit exams on violent behavior in a general sense. Column (6), violating local code of conduct, has some of the largest

estimates from the table but still does not tell a general story of exit exams having any impact on this general disciplinary infraction. As a whole, these tables suggest that failing the TAKS is only affecting small changes in student behavior among any of these observable dimensions before the end of high school for students in Texas.

Considering all the students in Texas within a five-point bandwidth of the threshold for passing a given section of the TAKS exam, these estimates back up the existing literature and add more evidence that students around the margin are experiencing heterogeneous effects with respect to their effort levels. While the effects vary some for the specific portion of the exam studied, I find students who fail are more likely to drop out of high school than their peers. As shown in Appendix Table A.1 I also support existing research showing the exit exams do have an impact on high school diploma receipt that increases with the number of retakes. After a student passes the exit-level TAKS test, there is no additional standardized test the next year so I am unable to compare scores year over year to discern effects on student ability on the exam. However, examining a different dimension of courses taken, I do find students just below the threshold to take more courses and more advanced courses than their peers just above the cutoff. It is also not simply that students who satisfy the exit exams have fewer remaining credits left to satisfy in order to obtain a high school diploma. Appendix Table A.2 shows the number of courses students take in non-TAKS subjects during their senior year. Students who fail a given portion of the exam take a very similar number of courses compared to students who passed the exam, so students seem to be switching their courseload rather than reducing it.

1.8 Heterogeneity and Robustness

1.8.1 Heterogeneity

All of the results of the preceding section suggest that the TAKS exam is having an impact on student behavior and human capital formation in ways beyond diploma receipt. But another question worth considering is whether there are heterogeneous effects among different sub-populations. There is compelling reason to believe that students in certain subgroups will have differential reactions to the same standard. If minorities or other subgroups are differentially affected by the exit exam, it is crucial to understand so in order to support these individuals in other compensating ways. All of the results reported in the previous tables pool across the entire state of Texas over a decade to determine individuals near the passing threshold. Because the TAKS test affects every student in public high school in Texas, there are a multitude of groups worth considering for evidence of systematic differences in individual response. I choose to focus on two specific subgroups likely to respond in additional ways to the TAKS exam: African American and Hispanic students.[19]

I examine the impact of the exams on these subgroups by estimating Equation 1.1 adding interaction terms for the model with an indicator for being a part of a given sub-population on the

[19]While students who receive free and reduced lunch, females, and Asians were also considered, the impact of belonging to any of these groups did not tell a consistent story beyond the behavioral responses for the whole sample already described.

same set of outcomes. Specifically, I estimate:

$$
\begin{aligned}
Y_i = \alpha_0 + \alpha_1 MET_{ti} + \alpha_2 Score_{ti} * MET_{ti} + \alpha_3 Score_{ti} * (1 - MET_{ti}) \\
+ \alpha_4 Group_i + \alpha_5 Group_i * MET_{ti} + \alpha_6 Group * Score_{ti} * MET_{ti} \\
+ \alpha_7 Group * Score_{ti} * (1 - MET_{ti}) + \epsilon_{ti}
\end{aligned} \tag{1.3}
$$

$$
\begin{aligned}
MET_{it} = \beta_0 + \beta_1 FAIL_{ti} + \beta_2 Score_{ti} * FAIL_{ti} + \beta_3 Score_{ti} * (1 - FAIL_{ti}) \\
+ \beta_4 Group_i * FAIL_{ti} + \beta_5 Group * Score_{ti} * FAIL_{ti} \\
+ \beta_6 Group * Score_{ti} * (1 - FAIL_{ti}) + \gamma_{ti}
\end{aligned} \tag{1.4}
$$

Where $Group_i$ is an indicator for belonging to a specific subgroup of interest. The first stage estimates of earning a score below the minimum threshold are reported in Appendix Table A.3 for both subgroups. Point estimates are similar to those for the whole sample from Table 1.2. Table 1.7 reports estimates for African American students. The first panel reports the impact of failing a given portion of the TAKS exam on dropping out of high school, the second panel reports estimates for the number of courses taken in a given subject the senior year of high school, the third panel displays the estimated impact of failing a portion of the exam on the number of advanced courses taken, and the fourth panel reports results for the number of courses failed. Looking at the coefficients reported on the interaction of being African American for the English, mathematics, and science portions of the exam have a statistically significant, negative effect on dropping out of high school; African Americans seem less sensitive on this dimension. The interaction between the subgroup and not meeting the standard for courses taken in panel 2 reports positive coefficients that are statistically significant for both mathematics and social studies, which can be taken as evidence that African Americans display a larger motivation effect for just failing an exam. This

pattern of a larger motivation effect for African American students is also present for the advanced courses panel, which again reports larger, statistically significant interaction terms for English, mathematics, and science. While less statistically significant, the final panel shows that African American students are less likely to fail a course after failing a portion of the exit exam relative to the whole sample, again suggesting a larger motivational effect. If African Americans are trying harder, the decrease in dropout due to failing a portion of the exam could be partly because students have higher grades and feel less pressured to forego enrollment in high school.

Another population of interest for a statewide standardized test in Texas is Hispanic students. Table 1.8 reports the effect of just failing a portion of the exam on a number of outcomes in the senior year of high school. Again, the first panel reports the impact of failing a given portion of the TAKS exam on dropping out of high school, the second panel reports estimates for the number of courses taken in a given subject the senior year of high school, the third panel displays the estimated impact of failing a portion of the exam on the number of advanced courses taken, and the fourth panel reports results for the number of courses failed. Unlike for African Americans, for failing the English, social studies, and science portions of the exam, Hispanic students are more likely to drop out of high school. For additional coursework taken the senior year of high school, however, for all four portions of the exam Hispanic students have negative, statistically significant interaction terms. Both of these trends are consistent with the motivation with heterogeneous effects framework where Hispanic students are more likely to choose to drop out rather than exert additional effort. The results on advanced courses taken in Panel 3 do not portray any clear trend. Taken together, Tables 1.7 and 1.8 suggest that exit exams do impact minorities differentially and furthermore that some subgroups are more likely to respond with additional effort while others are more likely to opt to exert less effort and drop out.

1.8.2 Robustness

In order for the estimates reported above to be credible estimates of the behavioral effects of the TAKS exam, it is important to demonstrate first that the empirical strategy is not picking up on some other underlying trend. Table 1.9 reports the estimates of Equation 1.1 on courses taken in each TAKS subject during the sophomore year of high school. I use the student's TAKS score and whether or not they meet the standard from the first administration during their junior year, but this information has not been revealed to the students during their sophomore year of high school because they have not taken the exit exam yet. Table 1.9 does not show the same trend as for students during their senior year, after the exams have been administered. Point estimates are negative and and less statistically significant. Furthermore, as shown in Table A.2, students above and below the threshold are not taking different numbers of courses their senior year, just changing the makeup. So taken together, Table 1.9 and Table A.2 do not show any evidence of students above the threshold being more motivated and simply being done with all of their high school coursework earlier. It appears that the changes in coursework are due to the consequences of the TAKS exam and not some earlier high school event.

Another possible concern with the results presented above is that the point estimates may be very sensitive to the empirical specification itself. Table 1.10 addresses this by displaying the number of courses taken in a given TAKS subject during a student's senior year at alternative specifications and bandwidths. Column (1) reports the preferred, linear specification with a five point bandwidth as a point of reference. This is analogous to Panel B of Table 1.3. Column (2) reports estimates using a quadratic term for the interaction with running variable and met standard. Column (3) reports point estimates for the preferred specification for the same five point bandwidth controlling for individual covariates. Column (4) reports estimates on the number of

subsequent courses taken in the TAKS subject senior year of high school for a wider bandwidth of seven points around the threshold score. Column (5) reports estimates for an even wider nine point bandwidth around the passing threshold. For all four subjects, estimates seem fairly robust to which specification is used, whether linear, quadratic, or with controls and, while a little more sensitive to bandwidth selection, still qualitatively the same. In only one instance, for social studies in the widest bandwidth, do the estimates seem to wash out completely. Additional Tables reporting stability for advanced courses and dropout are reported in Appendix Tables A.4 and A.5, respectively, showing similar robustness to empirical specification and bandwidth selection. These tables, paired with the earlier use of the Akaike Information Criterion to choose my preferred specification, suggest that linear specification is a valid choice and the results reported here are not too dependent on a specific functional form.

1.9 Policy Implications and Conclusion

Using a regression discontinuity framework, I examine the impact of the TAKS exit exam on several student outcomes before the end of high school. Looking at students right around the cutoff score that pass the first time the test is administered, I am able to estimate the impact of just barely failing a given subject of the TAKS exam on the number of courses taken for each subject on the exam, number of courses failed, whether these courses were advanced or AP or IB status, absences, number of disciplinary infractions and their resulting punishment, and type of disciplinary infraction. The variety in possible student outcomes captures new dimensions of student response to the initial administration of the exit exams both academically or behaviorally. I find students changing the courses they take their senior year and committing additional disciplinary infractions consistent with a motivational framework with heterogeneous effects.

Taken along with existing research, these results continue to expand our knowledge of how high stakes standardized testing impacts students along the threshold. Table A.1 does seem to suggest that the TAKS exit exams factor into graduation as they were designed to do. But beyond impacting the probability of obtaining a high school credential or dropping out, this research shows that students respond even before the end of high school. Students who score just below the threshold take additional classes in the TAKS subjects and are more likely to be observed behaving in ways consistent with students exerting discouragement effects and being more disruptive in the classroom compared to their peers just above the threshold.

From a policy perspective, these results continue to reinforce the importance of exit exams on students and the wider education environment. This policy was designed and implemented to impact high school instruction and graduation. Research has shown exit exams have had some success in this goal, but it clearly comes with some unanticipated behavioral effects. Lindo et al. [2010] describe the trade off between motivating some students to work harder, encouraging some students to drop out, and allowing others to exert lower effort towards performance standards in educational settings. This research provides evidence that this same trade off is present before the end of high school. Furthermore, if these exams are increasing the volume of disruptive events in high school they could impact the quality of the learning environment or require additional systems to manage disciplinary actions. Coupling this with results from Martorell [2004] and Clark and Martorell [2014] that find little impact of high stakes exit exams on post-secondary education and labor force outcomes, this research reinforces the idea that these impacts are largest in the intermediate run but could continue to spill into other areas of individual's lives.

This study sheds further light on some of the mechanisms that exit exams in Texas have on student outcomes for students who are at risk of failing the TAKS test. It confirms findings of

prior research that these exams do have a small impact on high school diploma receipt, and shows that students are responding along other margins than simply choosing to drop out or continue to enroll. If these choices impact human capital formation or possible job market opportunities in a traditional sense then it only increases the estimated magnitudes of the impact that exit exams are having on student behavior.

Table 1.1: Student Summary Statistics

	All	Below Cutoff	Above Cutoff	Mean Difference
Female	0.51	0.56	0.55	.0085***
	(0.50)	(0.50)	(0.50)	
FRL	0.38	0.49	0.47	.0148***
	(0.49)	(0.50)	(0.50)	
LEP	0.07	0.10	0.09	.0093***
	(0.25)	(0.30)	(0.29)	
Afr. Am.	0.13	0.19	0.18	.0103***
	(0.34)	(0.39)	(0.39)	
Hispanic	0.41	0.49	0.48	.0117***
	(0.49)	(0.50)	(0.50)	
White	0.41	0.29	0.31	-.0205***
	(0.49)	(0.46)	(0.46)	
HS Diploma	0.69	0.65	0.71	-.0600***
	(0.46)	(0.48)	(0.45)	
N	2900368	95777	166448	262225

Note: Table 1.1 reports summary statistics of student characteristics for the sample. Column (1) reports statistics for the whole sample, Column (2) for students below, and Column (3) above the cutoff score for passing the Mathematics exam. FRL indicates percent of the sample on free and reduced lunch. LEP indicates percent of the sample classified as limited English proficiency. N is the number of observations in the sample. * Indicates statistical significance at the 5% level, ** at the 1% percent level, and *** at the 0.1% level.

Table 1.2: First Stage Estimates of Passing Score on "Met Standard"

	English	Mathematics	Soc. Stud.	Science
Passing Score	0.798***	0.790***	0.764***	0.793***
	(0.007)	(0.007)	(0.006)	(0.006)
N	26282	34902	43131	39349
F-Test	13300.114	14731.997	15570.498	15842.897
R-Squared	0.7637	0.7112	0.6877	0.6822

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.2 reports estimates from Equation 1.2, the first-stage effect of failing a portion of the TAKS exam on the probability of not"met standard" in the TEA files. Column (1) reports the first stage estimate of scoring a failing score on the English portion of the exam on not meeting the standard, Column (2) for the mathematics portion, Column (3) reports the social studies portion of the exam, and Column (4), science.

Table 1.3: TAKS Exam Effects on Dropout and Courses Taken by Subject

	English	Mathematics	Soc. Stud.	Science
Dropout	-0.011***	0.036***	-0.007***	0.012***
	(0.003)	(0.001)	(0.002)	(0.000)
N	32978	42711	52749	49469
F-Test	13.767	962.318	19.054	54493.883
Mean	0.186	0.096	0.201	0.090
Courses Taken	0.141***	0.191***	0.015	0.134***
	(0.015)	(0.008)	(0.008)	(0.014)
N	26282	34902	43131	39349
F-Test	89.512	542.409	3.499	96.983
Mean	2.608	1.294	2.317	0.939

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.3 estimates Equation 1.1, a fuzzy regression discontinuity, on the probability of dropping out for each of the four subjects of the TAKS exam in Panel A. Column (1) reports the impact of failing the English portion of the exam, Column (2) reports failing mathematics, Column (3) the impact of failing Social Studies, and Column (4), Science. The second panel reports the effect of each TAKS section on the subsequent number of courses taken in that subject the students' senior year.

Table 1.4: Advanced Courses Taken, AP & IB Courses, Courses Failed, and Absences

	Advanced Courses	AP/IB Courses	Failed Courses	Days Absent
English	-0.008	-0.026***	-0.136***	-0.487***
	(0.010)	(0.003)	(0.040)	(0.015)
N	26282	26282	26282	32978
F-Test	0.695	84.637	11.565	1093.931
Mean	0.427	0.095	1.229	12.649
Mathematics	0.032*	0.008**	-0.036***	0.276*
	(0.013)	(0.003)	(0.007)	(0.137)
N	34902	34902	34902	42711
F-Test	6.397	8.287	29.780	4.044
Mean	0.562	0.141	0.974	11.846
Social Studies	-0.002	0.022***	0.107***	0.452**
	(0.003)	(0.000)	(0.006)	(0.140)
N	43131	43131	43131	52749
F-Test	0.295	20687.862	288.469	10.376
Mean	0.423	0.098	1.276	12.823
Science	0.028***	0.016**	-0.024***	-0.441***
	(0.006)	(0.005)	(0.003)	(0.091)
N	39349	39349	39349	49469
F-Test	23.845	9.204	63.591	23.503
Mean	0.583	0.149	0.999	11.688

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.4 estimates the effect of failing a portion of the TAKS exam on several additional outcomes before the end of high school. Column (1) reports the number of career advanced courses taken in the student's senior year by TAKS subject, Column (2) reports the number of AP & IB courses taken, Column (3) reports the number of failed courses in any subject in the senior year, and Column (4) reports the impacts on days absent during the senior year. The First panel estimates these impacts for the English portion, the second panel for mathematics, the third panel for social studies, and the final panel reports impacts of just failing the science portion of the TAKS exam.

Table 1.5: Days and Location of Disciplinary Actions

	English	Mathematics	Soc. Stud.	Science
Disciplinary Days	0.220***	0.536***	-0.353***	0.014***
	(0.012)	(0.084)	(0.033)	(0.004)
N	26839	38598	42164	45007
F-Test	346.866	40.839	115.103	13.773
Mean	1.706	1.594	1.765	1.707
Suspension	-0.006	0.016***	-0.003	0.005***
	(0.005)	(0.003)	(0.002)	(0.000)
N	26839	38598	42164	45007
F-Test	1.284	27.432	1.681	117.345
Mean	0.163	0.160	0.166	0.163

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.5 estimates Equation 1.1 on the effect of failing a portion of the TAKS exam on two types of disciplinary outcomes. Column (1) reports the impact of failing the English portion of the exam, Column (2) reports failing mathematics, Column (3) the impact of failing social studies, and Column (4), science. The first panel reports the impact on the number of disciplinary days actually served. The second panel reports estimates on the number of days served in suspension, whether in- or out-of-school.

Table 1.6: Types of Disciplinary Infractions

	Drugs	Fighting	Weapons	Truancy	Assault	Conduct
English	-0.005***	-0.002***	-0.000***	-0.023***	0.001***	-0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.004)
N	26839	26839	26839	26839	26839	26839
F-Test	51.729	65.082	17.730	684.385	25.389	0.019
Mean	0.010	0.011	0.001	0.039	0.002	0.209
Mathematics	0.001*	0.002***	-0.000	- 0.008***	0.002***	0.010***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
N	38598	38598	38598	38598	38598	38598
F-Test	5.391	17.664	0.111	181.179	486.034	65.925
Mean	0.009	0.010	0.001	0.034	0.001	0.201
Social Studies	0.005***	-0.005***	0.001***	-0.004***	-0.003***	-0.016***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
N	42164	42164	42164	42164	42164	42164
F-Test	54.495	58.031	819.084	78.706	282.836	505.869
Mean	0.009	0.011	0.001	0.036	0.002	0.210
Science	-0.001***	-0.003***	0.000	0.007***	0.000***	0.007***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)
N	45007	45007	45007	45007	45007	45007
F-Test	1624.884	317.667	1.774	10.842	742.411	44.524
Mean	0.011	0.009	0.001	0.032	0.001	0.205

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.6 estimates the effect of failing a portion of the TAKS exam on indicators for a number of different types of disciplinary infractions using Equation 1.1. Column (1) reports estimates for recorded disciplinary actions during senior year for drugs or alcohol, Column (2) reports estimates the impact of just failing a portion of the TAKS exam on an indicator for fighting, Column (3) reports results for whether any infraction was for having a gun, knife, or club on campus. Column (4) considers whether students were disciplined for truancy at the one, three or ten day interval. Column (5) reports whether students were disciplined for assault of a student, district employee, or non-district employee. Column (6) considers whether students were disciplined for violating local codes of conduct. For all results, Equation 1.1 was estimated using a five-point bandwidth.

Table 1.7: Heterogeneity: African Americans

	English	Mathematics	Social Studies	Science
Dropout	0.001	0.038***	-0.009***	0.016***
	(0.002)	(0.001)	(0.000)	(0.001)
Interaction	-0.060***	-0.009***	0.008	-0.022***
	(0.002)	(0.001)	(0.009)	(0.005)
N	32978	42711	52749	49469
Mean	0.186	0.096	0.201	0.090
Courses	0.139***	0.163***	-0.005	0.123***
	(0.019)	(0.009)	(0.008)	(0.020)
Interaction	0.016	0.148***	0.095***	0.058
	(0.019)	(0.004)	(0.000)	(0.032)
N	26282	34902	43131	39349
Mean	2.608	1.294	2.317	0.939
Advanced	-0.024***	0.021	0.008***	0.014***
	(0.004)	(0.016)	(0.001)	(0.002)
Interaction	0.080**	0.062***	-0.051*	0.070***
	(0.025)	(0.017)	(0.022)	(0.018)
N	26282	34902	43131	39349
Mean	0.427	0.562	0.423	0.583
Failed	-0.100*	-0.028*	0.079***	-0.009
	(0.042)	(0.011)	(0.001)	(0.007)
Interaction	-0.162***	-0.049	0.138***	-0.078
	(0.011)	(0.025)	(0.035)	(0.051)
N	26282	34902	43131	39349
Mean	1.229	0.974	1.276	0.999

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.7 reports Equation 1.3 for African American students. The first panel reports the impact of failing a given portion of the TAKS exam on dropping out of high school, the second panel reports estimates for the number of courses taken in a given subject the senior year of high school, the third panel displays the estimated impact of failing a portion of the exam on the number of advanced courses taken, and the fourth panel reports results for the number of courses failed. Column (1) reports the impact of failing the English portion of the exam, Column (2) reports failing mathematics, Column (3) the impact of failing social studies, and Column (4), science.

Table 1.8: Heterogeneity: Hispanic Students

	English	Mathematics	Social Studies	Science
Dropout	-0.018***	0.038***	-0.016*	0.004**
	(0.003)	(0.001)	(0.008)	(0.001)
Interaction	0.011***	-0.003**	0.012	0.013***
	(0.000)	(0.001)	(0.009)	(0.002)
N	32978	42711	52749	49469
Mean	0.186	0.096	0.201	0.090
Courses	0.173***	0.270***	0.032***	0.185***
	(0.001)	(0.001)	(0.009)	(0.008)
Interaction	-0.046*	-0.126***	-0.025***	-0.081***
	(0.023)	(0.015)	(0.001)	(0.010)
N	26282	34902	43131	39349
Mean	2.608	1.294	2.317	0.939
Advanced	0.024	0.019***	-0.016	0.008
	(0.022)	(0.001)	(0.011)	(0.014)
Interaction	-0.048*	0.019	0.021	0.034**
	(0.019)	(0.021)	(0.013)	(0.013)
N	26282	34902	43131	39349
Mean	0.427	0.562	0.423	0.583
Failed	-0.028	-0.016*	0.158***	-0.058***
	(0.042)	(0.007)	(0.013)	(0.010)
Interaction	-0.158***	-0.037	-0.079***	0.057***
	(0.003)	(0.021)	(0.012)	(0.013)
N	26282	34902	43131	39349
Mean	1.229	0.974	1.276	0.999

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.8 reports Equation 1.3 for Hispanic students. The first panel reports the impact of failing a given portion of the TAKS exam on dropping out of high school, the second panel reports estimates for the number of courses taken in a given subject the senior year of high school, the third panel displays the estimated impact of failing a portion of the exam on the number of advanced courses taken, and the fourth panel reports results for the number of courses failed. Column (1) reports the impact of failing the English portion of the exam, Column (2) reports failing mathematics, Column (3) the impact of failing social studies, and Column (4), science.

Table 1.9: Placebo: Courses Taken Sophomore Year

	English	Mathematics	Social Studies	Science
English	-0.003	-0.055***	-0.043***	0.016**
	(0.014)	(0.008)	(0.005)	(0.005)
N	24919	24919	24919	24919
F-Test	0.042	45.228	68.449	8.833
Mean	2.737	2.147	2.035	2.017
Mathematics	0.013*	-0.042***	-0.000	-0.005
	(0.006)	(0.006)	(0.002)	(0.005)
N	36598	36598	36598	36598
F-Test	5.417	44.270	0.043	1.339
Mean	2.711	2.129	2.033	2.006
Social Studies	-0.073***	-0.029***	0.011*	-0.038***
	(0.003)	(0.002)	(0.004)	(0.002)
N	38974	38974	38974	38974
F-Test	565.469	143.280	5.926	387.318
Mean	2.794	2.120	2.030	2.002
Science	0.026**	0.007	-0.011	0.004
	(0.008)	(0.005)	(0.007)	(0.010)
N	42349	42349	42349	42349
F-Test	10.259	2.046	2.606	0.150
Mean	2.676	2.124	2.036	2.015

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table 1.9 reports placebo estimates of Equation 1.1 on courses students took their sophomore year, a year before the first administration of the TAKS exit exam. Column (1) reports the impact of failing the English portion of the exam, Column (2) reports failing mathematics, Column (3) the impact of failing social studies, and Column (4), science. Each panel corresponds to courses taken in a given subject area during the sophomore year.

Table 1.10: Stability of Specification: Courses Taken

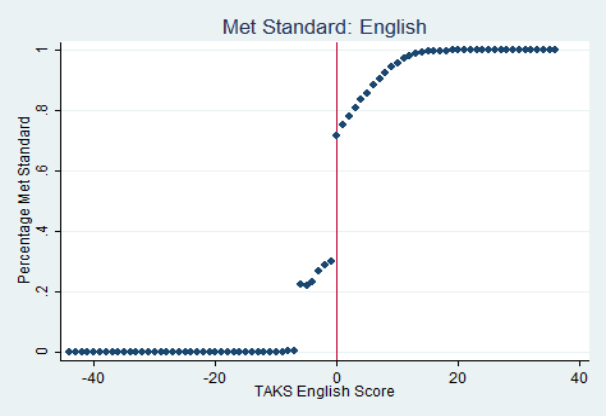
	Linear (BW5)	Quadratic	Covariates	BW 7	BW 9
English	0.157***	0.277***	0.113***	0.079**	0.104***
	(0.003)	(0.003)	(0.006)	(0.029)	(0.021)
N	35438	50222	32784	50222	65271
F-Test	2276.940	9809.722	307.176	7.185	23.718
Mean	2.659	2.658	2.653	2.658	2.654
Mathematics	0.191***	0.243***	0.223***	0.150***	0.148***
	(0.008)	(0.006)	(0.012)	(0.015)	(0.015)
N	34902	48472	31892	48472	61884
F-Test	542.409	1422.993	318.888	102.321	93.661
Mean	1.294	1.298	1.238	1.298	1.304
Social Studies	0.015	-0.017**	0.003	0.028**	-0.013
	(0.008)	(0.006)	(0.006)	(0.009)	(0.024)
N	43131	60182	40255	60182	77001
F-Test	3.499	8.539	0.337	9.024	0.290
Mean	2.317	2.311	2.306	2.311	2.313
Science	0.134***	0.138***	0.119***	0.131***	0.132***
	(0.014)	(0.005)	(0.012)	(0.018)	(0.021)
N	39349	54344	36079	54344	69411
F-Test	96.983	696.693	101.595	53.432	37.815
Mean	0.939	0.949	0.879	0.949	0.961

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

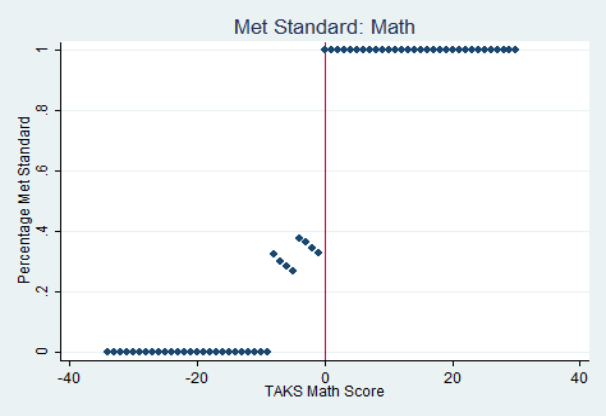
Note: Table 1.10 displays the estimated impact of failing a portion of the exit exam on courses taken senior year using different specifications and bandwidths as a check on the stability of the results. Panel 1 reports English courses taken after failing the English portion, Panel 2 reports the number of mathematics courses taken after failing the mathematics portion of the exam, Panel 3 is social studies courses taken after failing the social studies portion, and Panel 4 is science courses taken after failing the science portion. Column (1) reports the preferred linear specification at a bandwidth of five, Column (2) includes a quadratic term, Column (3) is a linear specification that includes individual controls, Column (4) estimates the impact using a linear specification but a wider bandwidth of seven points, and Column (5) uses a linear specification but a bandwidth of nine.

Figure 1.1: Graph of the Percentage of Students Meeting the Standard for Each TAKS Section.

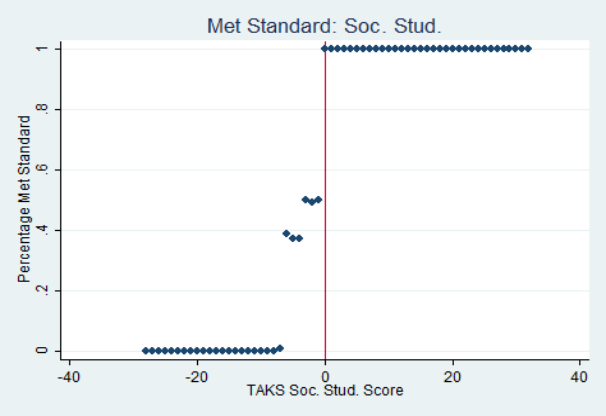
(a) English



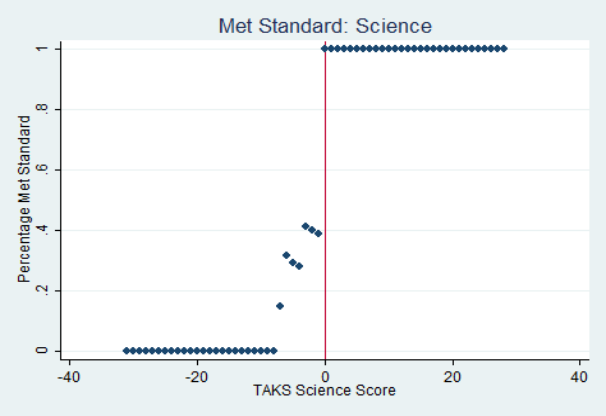
(b) Math



(c) Social Studies



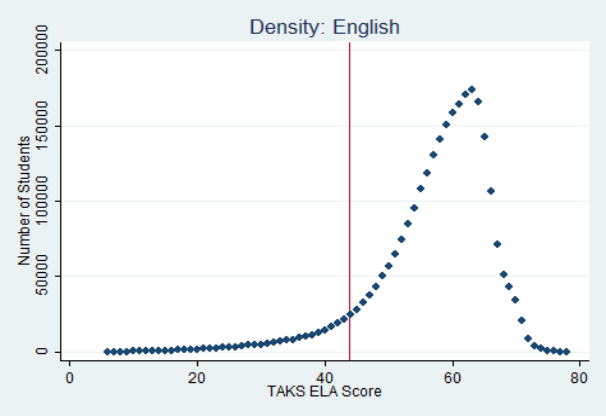
(d) Science



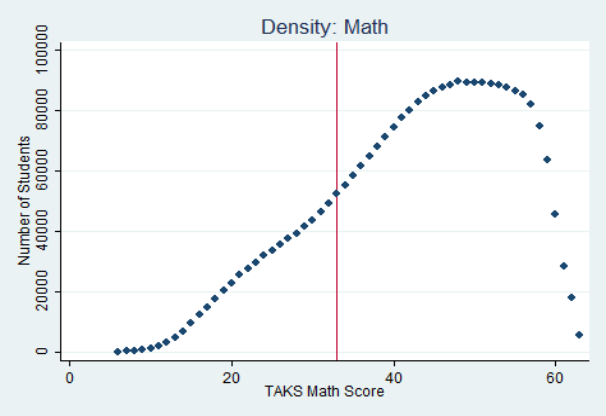
Note: Figure 1.1 displays the proportion of students reported by the TEA as "meeting the standard" for each score on the exit exam. Panel (a) displays the proportion of students for the English portion of the exam, Panel (b) displays meeting the standard for the mathematics portion of the exam, Panel (c) displays the social studies portion of the exam, and Panel (d) the science portion of the exam. Vertical line represents the pass threshold.

Figure 1.2: Full Support of the Densities of TAKS Scores by Subject.

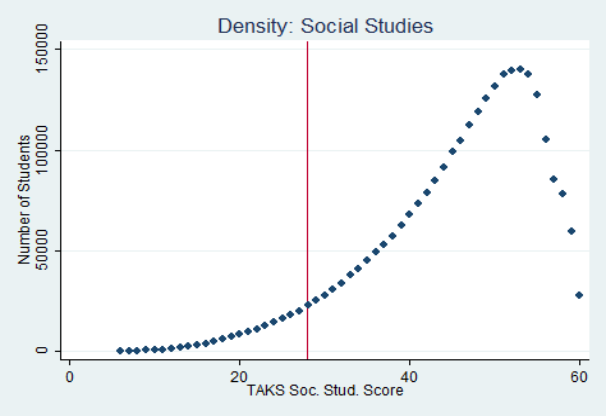
(a) English



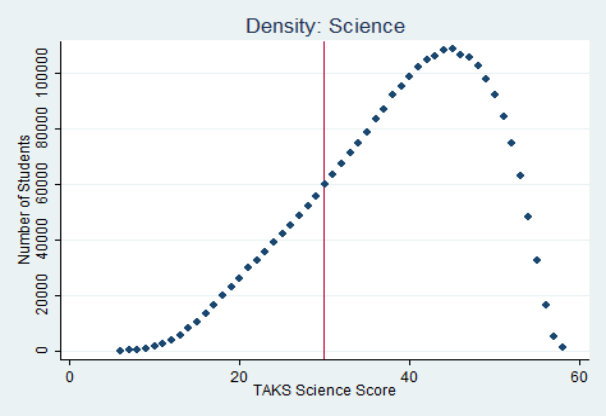
(b) Math



(c) Social Studies

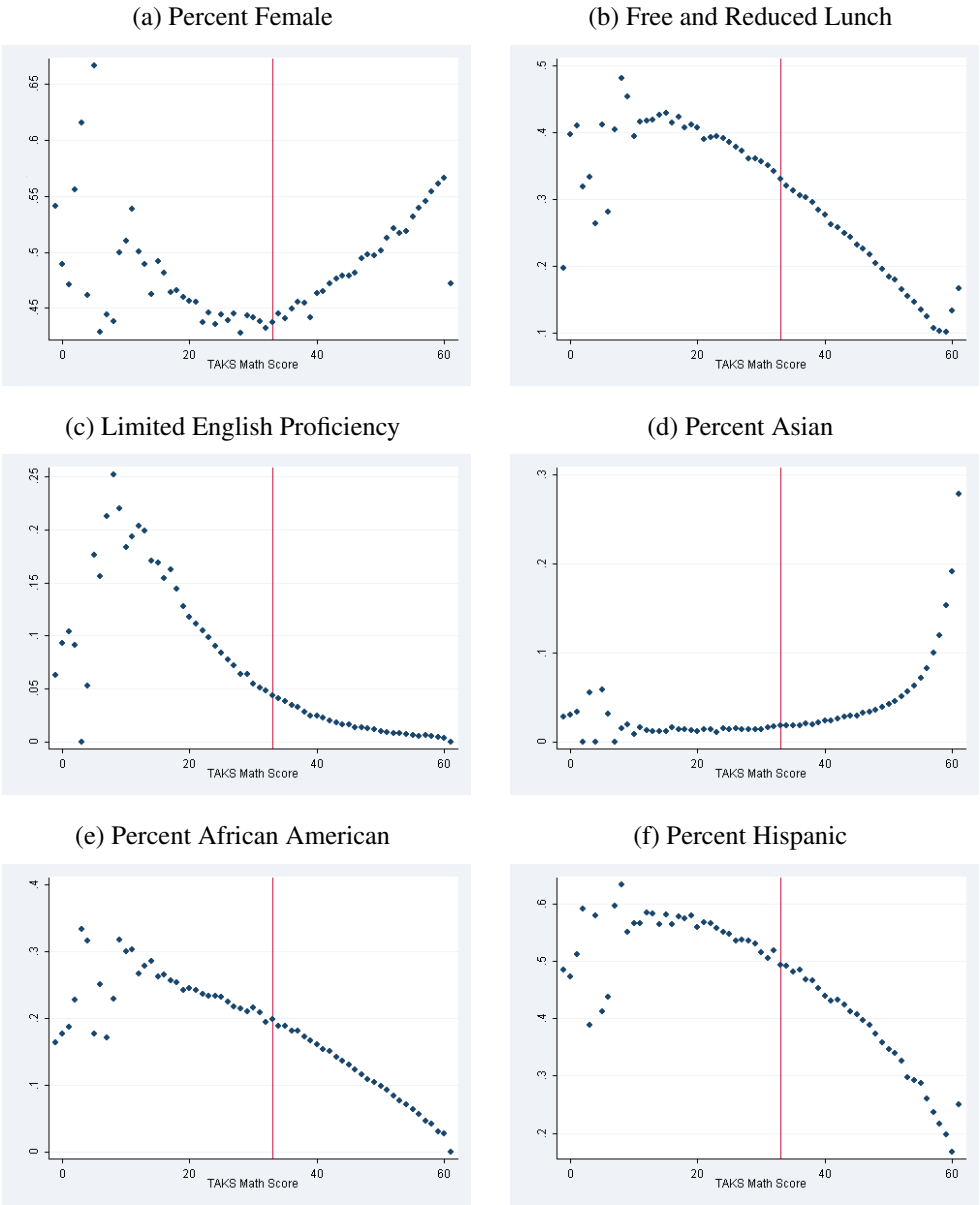


(d) Science



Note: Figure 1.2 displays the density of scores for each portion of the TAKS exit exam. Panel (a) displays number of students scoring each possible score for the English portion of the exam, Panel (b) displays number of students' scores for the mathematics portion of the exam, Panel (c) displays the social studies portion of the exam, and Panel (d) the science portion of the exam. Vertical line indicates the minimum score required to "pass."

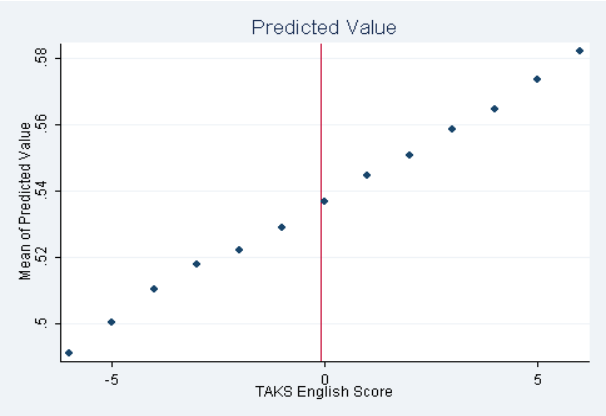
Figure 1.3: Graph of the Percentage Makeup of Covariates for the TAKS Math Exam.



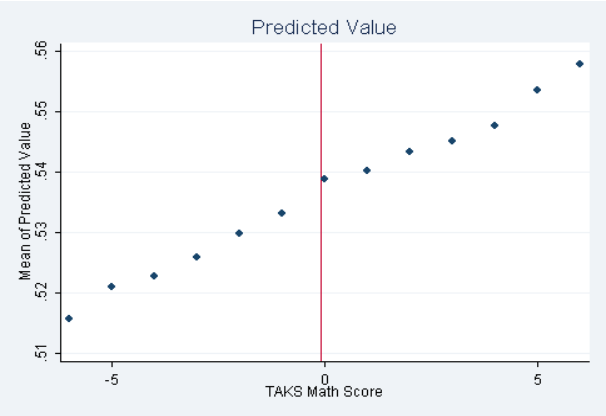
Note: Figure 1.3 displays the percentage makeup of the sample for several individual covariates based on their TAKS mathematics score. Panel (a) displays the percentage of the sample that is female, Panel (b) displays the percentage of the sample receiving free or reduced lunch, Panel (c) displays the percentage of the sample that is a limited English proficiency student, Panel (d) displays the percentage of the sample that is Asian, Panel (e) the percentage of the sample that is African American, and Panel (f) the percentage of the sample that is Hispanic. Vertical line represents the pass threshold.

Figure 1.4: Mean of the Predicted Value for Each Possible TAKS score by Subject.

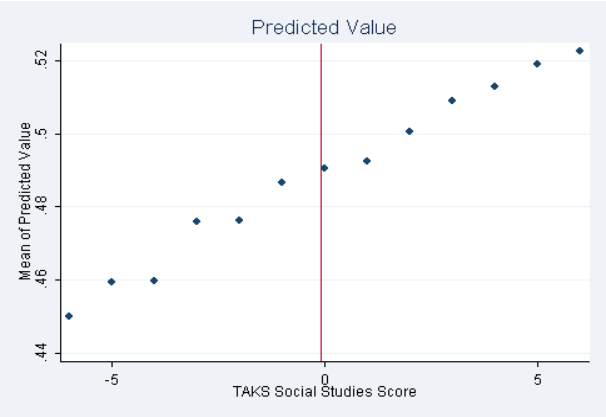
(a) English



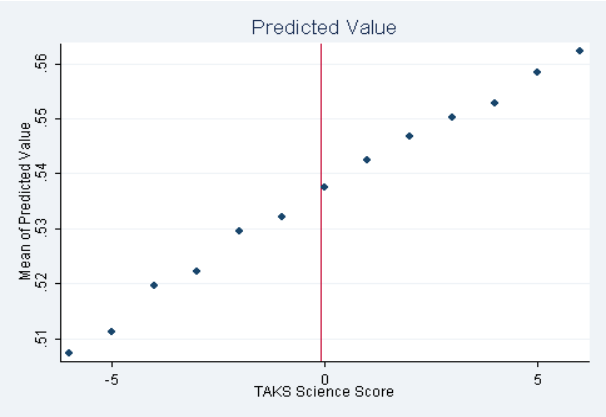
(b) Math



(c) Social Studies



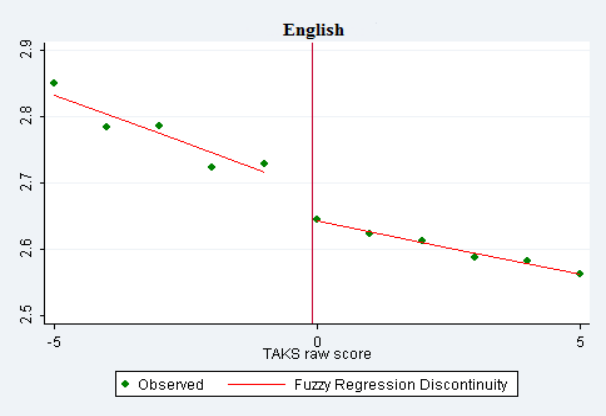
(d) Science



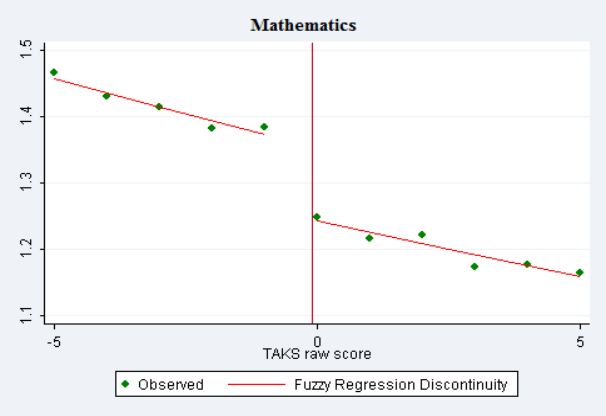
Note: Figure 1.4 displays the predicted values of a regression of high school diploma receipt on the individual covariates. Panel (a) displays the English portion of the exam, Panel (b) displays the mathematics portion of the exam, Panel (c) displays the social studies portion of the exam, and Panel (d) the science portion of the exam. Red line indicates the minimum score required to "pass."

Figure 1.5: Graph of Regression Discontinuity on Courses Taken for Each TAKS Section.

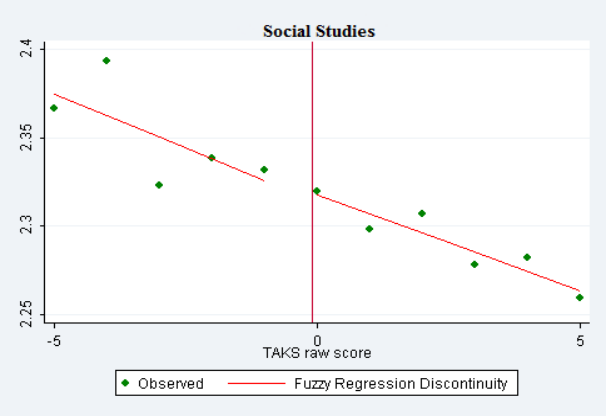
(a) English



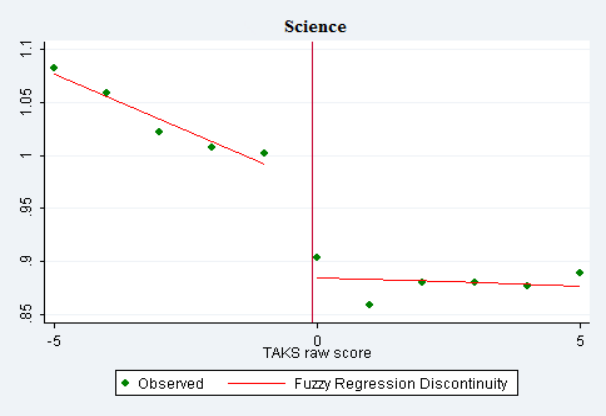
(b) Math



(c) Social Studies



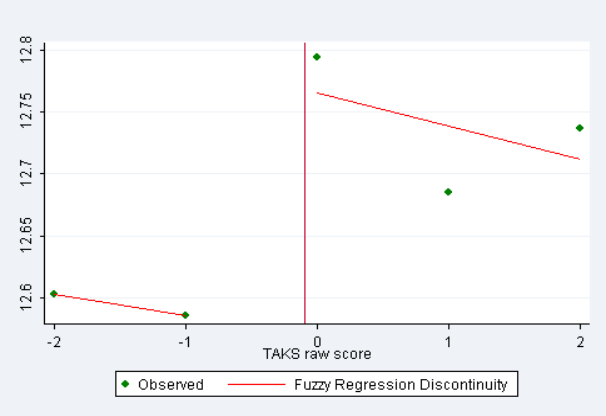
(d) Science



Note: Figure 1.5 displays the graphical results of Equation 1.1, the fuzzy regression discontinuity, on the number of courses taken the senior year of high school in a given subject by portion of the TAKS exit exam. Panel (a) displays number of courses taken for the social studies portion of the exam, Panel (b) displays number of courses for the mathematics portion of the exam, Panel (c) displays the English portion of the exam, and Panel (d) the science portion of the exam. Vertical line represents the pass threshold.

Figure 1.6: Graph of Regression Discontinuity on Days Absent for Each TAKS Section.

(a) English



(b) Math

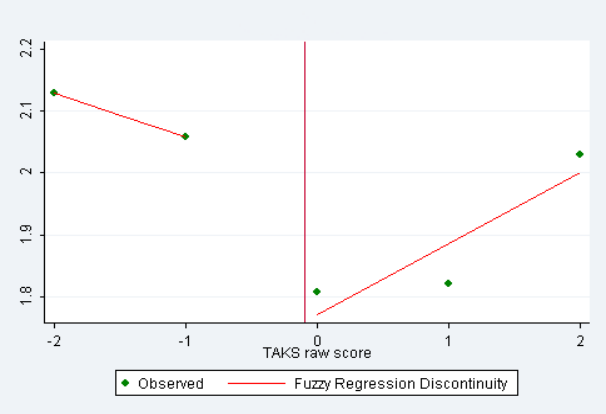
(c) Social Studies

(d) Science

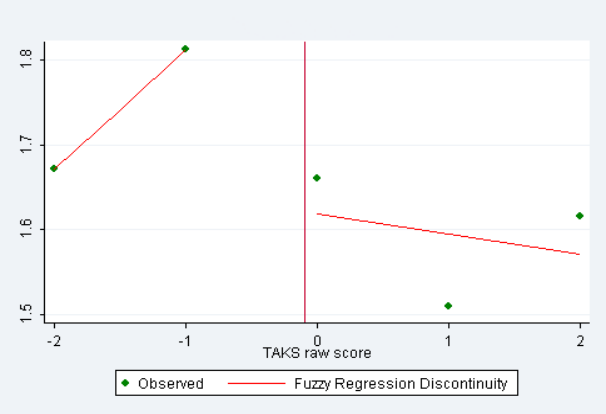
Note: Figure 1.6 displays the graphical results of Equation 1.1, the fuzzy regression discontinuity, on the number of days absent in the senior year of high school in a given subject by portion of the TAKS exit exam. Panel (a) displays number of days absent for the English portion of the exam, Panel (b) displays number of courses for the mathematics portion of the exam, Panel (c) displays the social studies portion of the exam, and Panel (d) the science portion of the exam. Vertical line represents the pass threshold.

Figure 1.7: Graph of Regression Discontinuity on Disciplinary Days for each TAKS section.

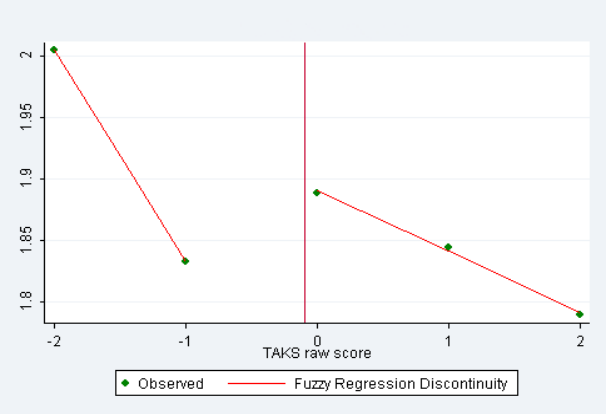
(a) English



(b) Math



(c) Social Studies



(d) Science

Note: Figure 1.7 displays the graphical results of Equation 1.1, the fuzzy regression discontinuity, on the number of disciplinary days assigned in the senior year of high school in a given subject by portion of the TAKS exit exam. Panel (a) displays number of disciplinary days for the English portion of the exam, Panel (b) displays number of disciplinary days for the mathematics portion of the exam, Panel (c) displays the social studies portion of the exam, and Panel (d) the science portion of the exam. Vertical line represents the pass threshold.

Chapter 2

Racing to a Diploma: Standardized Test Scores and High School Graduation Rates for Racial Minorities in Texas

2.1 Introduction

In the United States white students are more likely to graduate from high school than black or Hispanic students. For the cohort of public high school students born 1986-1990, the estimated high school graduation rates result in a black-white gap of 8.1 percentage points and a Hispanic-white gap of 8.5 percentage points (Murnane [2013]). These minority students also score lower on standardized tests than their white peers (Reardon and Galindo [2009], Clotfelter et al. [2009]).

It is hard to identify the reasons why these high school graduation gaps have persisted during the last three decades of the twentieth century. Along with graduation gaps, there are also well documented test score gaps that emerges early in school across races; black kindergartners score 0.64 of a standard deviation worse than whites on the tests contained within the Early Child Longitudinal Study (Fryer and Levitt [2004]). As education continues, the average gap in black mathematics scores stays relatively constant but reading scores fall relative to whites over time. Hispanic students, in contrast, enter kindergarten with much lower average math and reading skills compared to their white peers, but narrow this gap by roughly a third in the first two years of schooling before leveling off (Reardon and Galindo [2009]). These gaps change over time, with Hispanic mean scores rising relative to whites over the course of education (Clotfelter et al. [2009]).

Research also shows that, while important, simple differences in school quality cannot account for all of these measured gaps in both academic achievement during school and in graduation rates (Rivkin et al. [2005], Clotfelter et al. [2009], Fryer and Levitt [2004], Page et al. [2008], Hanushek and Rivkin [2006]). Many policies at the state and federal level exist to try to rectify these pervasive gaps, and there is an abundant literature in the economics of education examining various policies with the intent to increase graduation rates (Jacob [2001], Dobbie and Fryer Jr [2011], Neal [2006], Bound et al. [2010], Cornwell et al. [2013]).

But if we were able to set test scores equal between groups of minority students and their white counterparts, how much does the graduation gap shrink? Does the student's relative position in the distribution of ability or the type of school district they are enrolled in impact graduation in other meaningful ways? The goal is not to identify the causal effect of any single factor that contributes to test score differences between minority students and their white counterparts, but rather a summary measure of the total effect of the educational and home environment on graduation rates. If we could somehow guarantee that all students come out of high school with relatively equal ability, how far does that go in solving the problem of the graduation gap? Finding that it explains a large portion of the graduation gap would be important because it would show that these types of policy interventions are warranted. Conversely, finding that these differences in test scores, used as a proxy of cognitive ability, do not explain much of the gap would suggest that perhaps working on these types of policies can only solve so much of the problem in the long run.

This research aims to conduct this thought experiment. I try to quantify how much of the high school diploma gap can be accounted for by the differences in eighth grade test scores between races. I do this by employing linear probability models on high school diploma receipt by race for all students in Texas and then restricting the sample to specific portions of the population in order to

examine group differences. I find that, while controlling for demographic characteristics decreases the gap in high school graduation, it does not explain all of it. But, once I also control for cognitive ability measured on standardized tests at the end of middle school, this gap in diploma receipt is no longer large or statistically significant. Results vary for low socioeconomic status students, by whether the school district is urban or rural, and for different parts of the distribution of test scores. Controlling for campus-level characteristics for both the middle and high school where these students are enrolled suggests that school quality also matters, telling a school quality story. These findings confirm the previous results from the literature for a new time period and much larger Hispanic population, suggesting policies targeted at increasing school quality and student test scores continue to be viable methods for closing these graduation gaps.

I use administrative data on seven cohorts of public-middle school students in Texas who were enrolled in the eighth grade from 2003-2009. Having longitudinal administrative data on the universe of public school students for a state as large and diverse as Texas is an improvement over other frequently used survey data trying to measure high school graduation. Using the universe of public student data does not require the use of sample weights or lead to fears about misrepresenting the underlying population of public students in Texas. The time period is also of special interest because other research documents a shift in the trend in high school graduation rates around 2000, yet end before my observation period (Heckman and LaFontaine [2010]). Furthermore, by excluding General Educational Development (GED) credentials from my analysis, I am able to offer a consistent measure of diploma receipt in contrast to some debate in the literature about what that should entail. By leveraging some of the additional information the Texas Education Agency collects, I am also able measure the graduation gap among sub populations that have not received as much attention as simple racial categories.

A final benefit of this research is that Texas' large Hispanic population can be considered a bellwether in understanding how the changing makeup of racial minorities in the public education system nation-wide might impact achievement and graduation gaps in the future. As a fast-growing minority in the United States, research considers Hispanic students specifically. Gandara et al. [2003] find that in several measurable aspects the education quality of English language learners is inferior to native speakers' educational settings. Other research has focused on the intergenerational assimilation of Hispanics (Groger and Trejo [2002], Smith [2003]). Obtaining current estimates on Hispanic—white graduation gaps and some of the factors that seem to attribute to them can help with setting educational policy to support the success of the future US workforce.

The related literature makes clear the importance of rectifying this minority-white graduation gaps by highlighting the consequences these gaps in test scores and graduation rates have on minorities longer term, mostly in the labor market. While these articles address fundamentally different research questions, their conclusions are important for the high school diploma gap as well. Neal and Johnson [1996] conclude that observed wage gaps are largely skills gaps, which depend on differences in family demographics and school quality. The authors suggest that focusing on the obstacles minorities face in acquiring skills is what is needed from future research. Blau and Kahn [2005] assess how much the growing dispersion of cognitive skills impacts wage inequality. They find that greater dispersion of cognitive test scores in the United States plays a part in explaining U.S. wage inequality, but that greater residual inequality is considerably more important quantitatively than differences in the distribution of test scores in explaining higher U.S. wage inequality. These works and others show how large an impact on long-term outcomes these achievement and graduation gaps can have. This research is complementary to other results but does not inform our understanding of the relationship between high school diploma gaps and long term success

directly. Keeping in mind what is at stake in terms of the long term success of students lends more urgency to better understanding the early differences in student ability and how they impact high school diploma receipt.

A different branch of economic research considering the question of how student ability on standardized tests contributes to the high school diploma gap examines specific policies to measure their effectiveness in closing these documented gaps. Dobbie and Fryer Jr [2011] look at the Harlem Children's Zone and find that some large-scale interventions (like increasing school quality) have enough of an impact to alleviate the racial achievement gap in elementary school in both mathematics and English scores. But the authors are unable to address whether these specific interventions carry through beyond elementary school and subsequently impact educational attainment levels. Fryer and Levitt [2004] document the magnitude of achievement gaps between black and white students early in school and posit that a large share of this gap can be explained by the quality of the schools. Todd and Wolpin [2007] examine the contribution of lagged family and demographic controls on measured cognitive achievement on math and reading, finding these characteristics contribute a significant portion to the observed gaps. This research does not leverage the implementation of a specific event or policy change for identification, and thus does not contribute directly to our knowledge of how specific policies designed to close the high school diploma gap actually impact students. But the results here help better understand the relative importance of policies designed to close these gaps in cognitive achievement and high school graduation. As cognitive ability can account for most of the high school diploma gap, my results lend support to the idea that reforming the public education system can close these gaps nationwide.

This research also contributes directly to the literature documenting the gap between minority students and their white peers, both in academic achievement and high school graduation.

Murnane [2013] offers a survey of the trends in graduation rates in the United States from 1970-2010 overall and by several subgroups, as well as some of the models economists have posed to explain the persistent gaps. Data show a slow decline in graduation rates from 1970-2000, then a significant increase in high school graduation from 2000-2010. Murnane also covers some possible explanations for these trends and the paucity of causal evidence on what seems to impact high school diploma reciept. Heckman and LaFontaine [2010] bring consensus to the literature about what current graduation rates are by using several different datasets to tell a unified story of national trends in high school graduation. They report that minority graduation gaps are substantial and have not been converging in recent history. They also highlight the importance of using comparable definitions of graduation on comparable samples, a challenge given the information on high school graduates that exists, in order to reconcile some conflicting results in the literature. Their analysis, which shows that graduation rates are lower than many other widely used measures, covers students up to 2005.

I add to this existing literature documenting these gaps in several ways. Murnane [2013] does an excellent job synthesizing the literature, but most of the evidence comes from the pre-2000 era. Likewise, Heckman and LaFontaine [2010] are only able to observe students through 2005. In contrast, I study cohorts from 2003-2009, in the middle of a decade that Murnane [2013] points out stands in contrast to high school graduation rates for the preceding thirty years. Examining the most recent decade furthers our understanding of the current educational landscape, which is important for setting policy. My estimates of the high school diploma gap in Texas for minority students (9% for back students, 8% for Hispanic students) are in line with the current literature for this time period, suggesting that while we are unclear about the causes of these gaps we have a fairly consistent idea of what they are.

Furthermore, Heckman and LaFontaine [2010] point out that there has been renewed economic interest in high school graduation rates after the passage of the No Child Left Behind Act of 2001 (NCLB), and Murnane [2013] points out there is significant heterogeneity in graduation rates between states. My time period covers students under one consistent testing regime, the Texas Assessment of Knowledge and Skills, which was the state of Texas' standardized measure under NCLB. This again furthers our understanding of the current graduation gaps and not just recent trends. Focusing on Texas, a populous and diverse state, removes the confounding factors of differing education policies across states from the analysis, allowing for a more straightforward interpretation. Texas also has a large population of Hispanic students, a demographic trend which is surfacing in many other states and projected in more in the near future. Having a large population of Hispanics allows for closer examination of how certain factors impact them relative to other minorities or their white counterparts. Finally, I am able to leverage additional variables from the Texas Education Agency to focus on specific portions of the population, not simply limiting my analysis to racial categories. Examining graduation gaps for free or reduced lunch students and different types of schools continues to shed light on the trends Murnane [2013] documents among specific parts of the population.

The paper proceeds as follows: Section 2 describes the institutional detail surrounding the Texas public education system, Section 3 offers the research design, and Section 4 presents results. The final section offers a discussion and conclusion.

2.2 Institutional Detail

While no part of the public education system in Texas stands out as unique in a way that would be driving the results relative to other states or regions in the United States, it is important

to understand some of the main features leveraged in the subsequent analysis. A major component of the data for this analysis comes from student scores on the Texas Assessment of Knowledge and Skills (TAKS), the state-wide standardized test for scholastic achievement in Texas.

The TAKS test is administered state-wide to all students in public school in April of each year. The test consists of four sections: Mathematics, Reading, Science, and Social studies. One section is administered daily. The guidelines for administering the test are created in such a way to minimize distractions for test takers and to remove any unfair advantages for students.[1] The tests are untimed, and students are allowed as much time to respond to every question as is necessary.[2] Because the difficulty of the specific exam administered varies from year to year even trying to keep the standards constant, raw scores are then converted to a scaled score, which is directly comparable between years. While not binding, the minimum standard for passing is 2100 each year, and students who score 2400 or above achieve "commended performance."

The TAKS test was implemented in 2003 and the first cohort in my sample were eighth-graders the first year the test was in place. In my sample, I examine students who are eighth graders from 2003-2009 in order to observe their high school outcomes. Such students had been taking statewide standardized tests of a similar format yearly since the third grade, and so there is little chance results will be driven by differential effects of students of different racial groups being unfamiliar with the testing environment because it is the first administration.[3]

Another important piece of the Texas education system is whether and how a student ob-

[1]Guidelines require, among other concerns, no talking, no cell phones, and covering up any information around the classroom that could offer aid on a test.

[2]While not a requirement that administration of the exam continue beyond school hours, districts are allowed to offer students even that additional time.

[3] The TAKS test replaced the Texas Assessment of Academic Skills, which was in place right up until the implementation of the TAKS test.

tains a high school diploma five years after they take the TAKS test in the eighth grade.[4] A student in my sample must satisfy two major requirements in order to be awarded a high school diploma. First, a student must complete one of several approved courses of curriculum covering high school subjects. Throughout the time period of interest, the state of Texas offered several different types of high school diplomas: minimum high school program, recommended high school program, and distinguished achievement high school programs. While these distinctions are recorded on the student's final transcript, it is not clear that the actual high school credential is different in any meaningful way nor do employers screen on this additional information about the rigor of high school coursework. Therefore, in all my analysis, I do not distinguish between these different types of high school diplomas and pool all of them as "received diploma" for my outcome of interest.

The second major requirement in obtaining a high school diploma from a public school in Texas is passing the exit exam component of the TAKS test. The exit exam component is administered for the first time the spring of their junior year. Students who fail one or more portions are offered chances at several retakes before the end of their senior year. By the end of high school, students must score at or above a minimum threshold on each of the four sections (mathematics, science, English, and social studies) and meet a minimum number of credits and grade point average in order to receive a high school diploma. Other research shows that such high-stakes testing regimes constrain high school diploma receipt, and this will be reflected in the sample. However, as the state consistently mandated such a requirement for all students in

[4]The "on-time" graduation date is used in order to get a standard measure of high school completion. While students who repeat grade could very well finish a regular high school diploma track one (or more) additional years after the rest of their 8th grade cohort, the proportion of students who end up doing this is quite small. Alternative, more inclusive definitions of "on-time" receipt do not substantially change any findings.

Texas during my time period, obtaining a high school diploma adequately reflects satisfying all the necessary requirements and my outcome of interest will correctly reflect this. Discussion of whether or not these standards to obtain a high school are sufficient or too strenuous are beyond the scope of this paper.

Murnane [2013] makes the excellent point that high school graduation rates are a challenge to study in general because of the data limitations. Even with administrative data as good as that from the Texas Education Agency (TEA), this dataset suffers from one source of bias that can not be corrected for with complete certainty: students leaving one school to enroll in another in the state outside the purvey of the TEA, or moving between states. For such students, all observations after switching are not recorded, and thus these students can go on to obtain an on-time diploma and will not show up as in the data with one because the TEA was never notified. Even the records of students leaving the public education system are incomplete. I code all students who happen to fall into this category as "not graduated." There is some reason to think students who leave the sample due to this reason may vary systematically by race. Table 2.1 reports the number of students in my sample who are no longer enrolled in a public high school in Texas at the start of the 10th grade. This number of students can be viewed as a partial proxy for students who are leaving the sample but not dropping out because the start of 10th grade is still before the compulsory schooling age for most students. Of students who leave the sample, Hispanic students are 55 % of the total. Because of this, it is possible all estimates of ultimate graduation are biased downward.

2.3 Research Design

2.3.1 Data

Data for this analysis comes from administrative records from the TEA supplied by the Texas Education Research Center. The TEA administers K-12 public education in Texas. The TEA records several key variables for students that are included as demographic controls in the empirical specifications: sex, age, and race of the student, whether the student receives free or reduced lunch or is a special education student, campus and district of enrollment, TAKS scores for each student for all subjects, and graduation information.

The sample of interest is constructed of all 8th graders enrolled in Texas public schools from 2003-2009. The TAKS test was first implemented and administered in 2003 and continued through 2012, so all cohorts fall under a directly comparable standardized testing regime. I collect demographic information about the student from the student enrollment files from the fall of their eighth grade year and pair it with their TAKS mathematics and reading scores from the spring of 8th grade. I then record the high school campus the student enrolls in for 9th grade to add in additional high school campus-level controls both from the Academic Excellence Indicator System (AEIS) from the TEA and calculated directly from the enrollment files (Texas Education Agency [2012]).[5] To this dataset of individual-level student demographic information, test scores, and middle- and high school-level campus characteristics I then merge in the graduation information at the student level. As mentioned previously, while Texas offers several levels of high school diploma, this research treats all of them as the same credential.[6]

[5]The AEIS are TEA-constructed aggregates and publicly available. These variables from the AEIS include demographic makeup of the campus and average SAT score of the high school campus, among others.

[6]While there are several levels of high school diploma offered by the TEA as the same credential, in keeping with the more recent literature this research does not include a GED certificate as an equivalent measure (Heckman and

Because of the relative importance of individual scores on both the mathematics and reading portions of the TAKS and the race of the student, only individuals who have all of this information recorded are included in the sample. Additionally, students who take an alternate form of the TAKS test are omitted from the sample in order to capture students who should be otherwise comparable between years.[7] The percentage of students receiving waivers is reported in Table 2.1. While the percentage of Asian students receiving waivers is much lower at only 3% of the sample, for black, white, and Hispanic students this percentage ranges between 11 and 16%. This leaves 1,882,356 eighth graders with demographic information and TAKS reading and mathematics scores in my sample.

Table 2.1 reports summary statistics for all public 8th graders with a recorded TAKS score from 2003-2009. Column (1) reports summary statistics for the entire sample. Column (2) reports summary statistics for only white students. Column (3) reports these statistics for black students; Column (4) reports them for Hispanic students. Columns (5) and (6) report them for Asian Students and all minorities pooled together, respectively. The first row reports the percentage of students obtaining an on-time high school diploma. The difference in means implies that, in Texas during my sample period, there was 9% graduation gap between black and white students, and a 8% graduation gap between Hispanic and white students. Asian students have a 6% higher graduation rate than their white counterparts. Figure 2.1 graphs the respective graduation rates for each racial category for every cohort in my sample. Table 2.1 also reports the average TAKS reading and math scores for each racial category, both in raw terms (out of 2700) and standardized to have a mean of

LaFontaine [2010]). Because of this, a student who dropped out of high school but earned the GED equivalent on-time would still be counted as failing to earn a high school credential.

[7] The TAKS offers several alternative formats, both Alternative and Modified for students with documented learning disabilities in special education, as well as a form of the test in Spanish, in Braille, and on a computer. Because the results of these exams are not directly comparable, these students are omitted from subsequent analysis.

zero and a standard deviation of 1. Again, in line with other estimates, black and Hispanic students score worse then their white peers by roughly a third of a standard deviation.

A handful of interesting facts stand out from the demographic characteristics reported by the TEA, also included in the Demographics panel of Table 2.1. Across each racial category and for the whole sample, students are roughly 13 years at the start of the eighth grade and are roughly the same percentage female. Hispanic students, however, are much more likely to be classified as limited English proficiency students (15% of Hispanics versus 1% of black students and no white students). Also, both black and Hispanic students are much more likely to be on free or reduced lunch, the indicator for low socioeconomic status in the sample (75% of Hispanic students and 61% of black students, compared to 18% of white students).

Because much of the subsequent empirical work relies on the assumption that minority students differ in both observable and possibly unobservable ways from their white counterparts, the distribution of test scores are displayed visually. Figure 2.2 graphs the distribution of TAKS Mathematics scores for the whole sample and each racial group. Figure 2.3 graphs the distribution of TAKS Reading scores, again for the whole sample and each racial group individually. For each of the two figures, all seven cohorts of 8th graders are pooled. Both figures display similar trends; White and Asian students perform better than black and Hispanic students.

It is worth noting that empirical estimates could suffer from additional bias if the TAKS test is racially biased. Other research using measures of cognitive ability as a student level control often use Armed Forces Qualification Test (AFQT) scores (Neal and Johnson [1996], Cameron and Heckman [2001]). The AFQT is shown to be free of racial bias, but in other datasets it is not administered at a consistent age or at the same level of schooling for every student. One benefit of using TAKS scores for eighth graders is that the exact same exam is administered at exactly

the same time within each cohort, and scores are directly comparable between cohorts. However, possible racial bias in the TAKS test could impact students in several ways. If minority students systematically score lower than their white peers, then this will overstate the effect test scores have on high school graduation. Conversely, if the TAKS test systematically under-predicts minority student ability then any perceived differences in test scores could understate the impact ability has on graduation. Measuring the validity of these possible test-induced biases is beyond the scope of the paper.

2.3.2 Empirical Design

As mentioned previously, this paper addresses the differences in graduation rates in the spirit of Neal and Johnson [1996] and Johnson et al. [1998] and accounts for test scores by including them as explanatory variables of interest in linear regressions that estimate receiving a high school diploma. To do this, I estimate the following linear probability model:

$$Grad_i = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \Theta_t Year_i + \epsilon_i \tag{2.1}$$

Where $Grad_i$ is an indicator for whether individual i received an on-time high school diploma, $Black_i$ is an indicator for whether a student is black, $Hispanic_i$ is an indicator for whether an individual is Hispanic, $Asian_i$ is an indicator for being Asian, and Θ is a vector of cohort fixed effects (2003-2009).[8] The coefficients of interest are the βs, as they represent the differential effect of belonging to a given minority relative to white students, the omitted racial group. For this and all following specifications, I run the regressions independently for males and females

[8] While individuals often identify with more than one race or ethnicity, for the period in question the Texas Education Agency only recorded one race. For this dataset, the racial categories are mutually exclusive.

because of the voluminous other research showing the differences between sex in educational settings.

We are ultimately interested in the impact of the distribution of standardized test scores on high school diploma receipt for different minorities. However, the literature documents the extent to which minority students vary in observable characteristics from their white peers and how these differences factor in to academic achievement and graduation rates. Therefore, I next estimate:

$$\begin{aligned} Grad_i = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \Theta_t Year_i + \\ \gamma_1 Age_i + \gamma_2 FRL_i + \gamma_3 LEP_i + \epsilon_i \end{aligned} \tag{2.2}$$

in order to gain an understanding of how much these characteristics contribute to diploma gaps in this sample. Here, the new variable Age_i is the reported age of the student at the start of the 8th grade (in years), FRL_i is an indicator for whether the student receives either free or reduced lunch, and LEP_i is an indicator for whether the individual is a limited English proficiency student. In order to shed some light on how school quality in general, and not cognitive ability of the individual, I also report estimates of Equation 2.2 with school fixed effects in Column (4) of most tables. Because I am interested in isolating the contribution of eighth grade test scores specifically, I then estimate:

$$\begin{aligned} Grad_i = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \Theta_t Year_i + \\ \gamma_1 Age_i + \gamma_2 FRL_i + \gamma_3 LEP_i + \\ \alpha_1 Math_i + \alpha_2 Reading_i + \epsilon_i \end{aligned} \tag{2.3}$$

where $Math_i$ is the student's scaled mathematics score on the TAKS eighth grade exam and $Reading_i$ is the student's scaled reading score. Finally, to make try and account for possible

school-level differences that could be contributing to both achievement and graduation outcomes, I also estimate Equation 2.3 with middle school and high school campus fixed effects.

After reporting these five specifications, I then focus on specific sub-populations in Texas to see if the graduation gaps in these groups are systematically different from the entire sample. Specifically, I examine only those students who are low socioeconomic status (as measured by free or reduced lunch receipt) and by type of school district (as defined by the TEA). Examining each of these groups in turn sheds further light on what type of students seem to be driving the results for the whole sample.

A final specification of interest allows for nonlinear returns to test scores across the range of ability. The TEA reports two additional thresholds along with the student score; whether the student failed to meet the minimum performance standard (scoring below 2100) or above the commended performance standard (above 2400). I therefore estimate a spline regression with knots at these two thresholds of the following form:

$$\begin{aligned} Grad_i = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 Asian_i + \Theta_t Year_i + \\ \gamma_1 Age_i + \gamma_2 FRL_i + \gamma_3 LEP_i + \\ \alpha_1 Math_i + \alpha_2 d_1(Math_i - 2100) + \alpha 3 d_2(Math_i - 2400) + \\ \alpha_3 Reading_i + \alpha_4 d_3(Reading_i - 2100) + \alpha_5 d_4(Reading - 2400) + \epsilon_i \end{aligned} \tag{2.4}$$

where d_1 is an indicator variable for whether an individual's math score is above 2100, d_2 is an indicator for whether an individual's math score is above 2400, d_3 is an indicator for whether an individual's reading score is above 2100, and d_4 is an indicator for whether an individual's score is above 2400. The results section covers the exact order of the specifications run for each table and

what the estimates suggest for the whole sample and specific sections of males and females below. All specifications report robust standard errors.

2.4 Results

Table 2.2 examines the relationship of several different sets of demographic and cognitive characteristics on the probability of obtaining a high school diploma for males. Column (1) estimates Equation 2.1 on the minority gap in obtaining a high school diploma controlling for only a vector of races and cohort fixed effects. It is worth noting that the point estimates for blacks (-0.070) and Hispanics (-0.064) imply that blacks are 7 percentage points less likely to obtain a high school diploma than white students and Hispanic students are 6.4 percentage points less likely than whites. These are both statistically significant and in line with general summary statistics on high school graduation rates reported both in Table 2.1 and other published work (Murnane [2013], Heckman and LaFontaine [2010], etc.). Table 2.3 reports the same estimates for females: controlling for just race and cohort we observe a 10.7 percentage point gap in graduation relative to whites. Hispanic females also have a higher gap then their male peers, at -0.09.

But while the goal is to characterize the racial differences in high school graduation, there is clearly much more that determines an individual's educational outcome than just the color of their skin. Column (2) therefore also includes all the individual-level characteristics reported by the TEA as additional controls. The point estimate on being black drops to -0.025, a little less than half of the originally gap. This implies that together these characteristics are important, but still do not explain all of the graduation gap for black students. For Hispanic students, including these same controls changes the point estimate considerably; it changes to a positive 1.9 percentage point advantage over their white counterparts, effectively closing the diploma gap. For black females,

reported in Table 2.3, the diploma gap drops to 5.9 percentage points, roughly the same decrease in absolute value (0.045 for males and 0.048 for females). For Hispanic females, the difference in diploma receipt drops to only 0.003 after controlling for demographic information.

Column (3) then adds in additional controls for student ability by introducing standardized TAKS reading and math scores. For males, adding in reading and math scores again increases the coefficient on black students, to 0.026, and on Hispanic students to 0.049. Both of these estimates can be interpreted as black and Hispanic students actually performing better than their white peers once accounting for demographic characteristics and test scores. Additionally, Hispanic students seem to do much better than their black, or even white, peers, suggesting that addressing differences in individual characteristics are a bigger factor in increasing graduation rates for Hispanic students. Females (reported in Table 2.3) again show the same trend, with black females graduating at a 0.5 percentage points rate greater than their white peers once accounting for demographic and cognitive ability controls. For Hispanic females, once accounting for both demographic and cognitive controls their coefficient increases to 0.030.

While it is interesting that adding in test scores to the other demographic controls has that equalizing effect on diploma gaps, there is also the distinct chance that it is a school quality story driving the differences. Column (4) of Table 2.2 tries to consider this by reporting estimates of the same Equation 2.2 including the school fixed effects but not test scores. In both Tables 2.2 and 2.3, comparing the estimates between Columns (2) and (4) shows that accounting for school fixed effects in addition to demographic characteristics continues to close the high school graduation gap for black males, but does not close it completely like accounting for test scores does in Column (3). Similarly for black females, including school fixed effects lessens the percentage point diploma gap by roughly half, but does not close it completely. For Hispanic males, who seem to have more of

the gap explained by demographic characteristics, the estimate in Column (4) of 0.014 is roughly in line with the estimate without school fixed effects: 0.019. For Hispanic females, including school fixed effects increases the point estimate by 0.004.

Column (5) of Tables 2.2 and 2.3 estimate Equation 2.3 with school fixed effects. For black males, controlling for both school fixed effects and test scores increases their relative graduation advantage over their white peers by a further one percentage point. Hispanic males, on the other hand, have estimates roughly equal between the two specifications, 0.049 without school fixed effects and 0.042 with. Black females also show substantial increases in the probability of obtaining a high school diploma once accounting for school fixed effects; the point estimate increases by 0.025. Hispanic females, on the other hand, report the same coefficient between the two specifications. Taken together, Tables 2.2 and 2.3 paints a picture of differences in student characteristics accounting for a large difference in the graduation rates between black and Hispanic students and their white classmates, with additional evidence of a school quality story being present in Texas as well, especially for black students.

After documenting these overall patterns for Texas in Tables 2.2 and 2.3, I next consider the same specifications for only those students on free or reduced lunch. Free or reduced lunch receipt is the only available proxy for low socioeconomic status in the data. Low socioeconomic status students might be at higher risk of dropping out of high school due to other factors outside the classroom, so focusing on just those students rather than the whole sample paints a clearer picture of the graduation gaps between poor minority students and their poor white peers. The summary statistics point out that the group of students on free or reduced lunch is disproportionately black and Hispanic students. Table 2.4 reports the same estimates as Table 2.2 for only those students on free or reduced lunch. Several important features stand out. First, in Column (1), which estimates

Equation 2.1 with only a vector of racial indicators and cohort fixed effects, the large, statistically significant difference between black students and their white peers is no longer present. Likewise, the graduation gap for Hispanics is no longer present; relative to white students on free or reduced lunch Hispanic students are actually more likely to graduate for both males and females. Again, adding demographic controls seems to matter more for Hispanic than black students: the coefficient for black males does not change at all and only changes by 0.004 for females between Columns (1) and (2) but increases by a statistically significant 0.025 for both Hispanic males and females. Including school-level controls in Columns (4) and (5) again suggest that a school quality story is also at work, especially for black students. Overall, the same trends are present in the two tables, but since minority students start relatively more equal in terms of graduation rates, accounting for measured differences in individual and school level characteristics results in estimates that suggest that black and Hispanic students on free or reduced lunch are relatively more successful than their comparable white peers.

Because of the relative importance in closing the graduation gap that test scores seem to display for both males and females, it is worth examining whether there seem to be non-linear effects of test scores for racial minorities. Many of the results presented above, while not causal, suggest that distinct from other demographic or school-level characteristics, differences in test scores can account for substantial portions of the diploma gap. It is therefore worth considering whether there are nonlinearities with regard to test scores and graduation rates. While there are no tangible benefits to the classification, the TEA classifies any student who scores above a 2400 on a portion of the TAKS exam as achieving commended performance on that subject. These groups are still sizable portions of the overall sample; 17% of students meet the commended performance standard for mathematics, 39% of students meet the standard for commended performance for

reading. In addition to a commended performance threshold, the TEA also sets a minimum performance standard (at 2100). While eighth graders in my sample period were not penalized for scoring below this threshold on any portion of the exam, eventually the standard becomes binding at the exit level exam in the spring of their junior year. Because there are no consequences for students scoring below this threshold, the group of students with scores below this point can be considered the low ability group without fear of confounding the effects with other institutional policies that might impact their eventual graduation rate. Again, the proportions of students failing to meet the minimum performance standard is non-trivial: roughly 32% of students fail to meet the mathematics performance standards and 14% of students fail to meet the minimum standard for reading.

Estimates of Equation 2.4 and related variants are reported in Table 2.6 and 2.7 for males and females, respectively. In order to compare how allowing for these nonlinearities at the minimum performance and commended performance knots with a linear spline changes estimates of graduation rates, Column (1) reports estimates of Equation 2.3 (analogous to Column (3) of Tables 2.2 and 2.3). Then Column (2) controls for math scores linearly while allowing a spline for reading scores, and Column (3) controls for reading scores linearly while allowing a spline for math scores. Next, Column (4) estimates Equation 2.4 with the full spline for both portions of the test. Finally, Column (5) estimates the same equation as Column (4) but includes school fixed effects. For black males and Hispanic males and females, the estimated graduation rates are very similar across these specifications. Black females seem to have more of a school quality story as portrayed by the increase in coefficients from Column (4) to Column (5). However, one interesting trend emerges: across both males and females and for all specifications, similar trends for the different parts of the ability curve are present. For both math and reading, students on the lower parts of

the spline gain additional percentage point returns to graduation. For math, the returns are highest at the lowest portion; for reading, it is the intermediate range. For both math and reading, the commended performance portion is associated with a negative relative return, suggesting that once students are relatively high enough up the distribution of skills the returns lessen.

One final way to continue to add to our understanding of how the graduation gaps for minorities vary between different segments of the population is to examine the type of school district a student attends. It is reasonable to believe school quality, something that clearly has an impact on graduation rates, could vary considerably by whether the school district is rural, urban, suburban, or something else. Tables 2.8 and 2.9 does this for five separate types of school districts as classified by the TEA for males and females: major urban school districts, major suburban districts, independent towns, rural school districts, and charter school districts. It should come as no surprise that due to the geography of Texas and the location of its major cities these different types of school districts make up relatively different percentages of the whole sample of students. Because of where the population of Texas residents is located, the last row of Tables 2.8 and 2.9 also reports what percentage of the whole sample students in that type of school district make up; these range from 34% of students being enrolled in a major suburban district to just 1% of students being enrolled in a charter school district.

Each column in Tables 2.8 and 2.9 reports the estimates of Equation 2.3 with school fixed effects for a given type of school district. Each of these columns could be compared to Column (5) of Tables 2.2 and 2.3, which reports the same regression for the whole sample. Column (1) reports this for major urban districts. Column (2) reports this for major suburban districts, Column (3) reports this for independent towns, Column (4) reports estimates for rural school districts, and

Column (5) reports this for charter school districts.[9] While for all five types of school districts, after accounting for individual characteristics, test scores, and school fixed effects black and Hispanic students graduate at relatively higher rates than their white counterparts, there are still some interesting differences. In major urban school districts, both black and Hispanic students graduate at relatively lower rates than for the whole sample. This is also true for students in major suburban school districts, where Hispanics do relatively much worse (though still have a positive graduation differential) than the statewide estimates. Conversely, black and Hispanic students who are enrolled in independent towns do much better than both their white peers in those towns and than black and Hispanic students in Texas as a whole. These results suggest that school quality most likely effects students in nonlinear ways, but that it also has measurable effects. Policies that try to raise school quality could then be effective if properly targeted.

2.5 Discussion and Conclusion

Some research brings up the question of whether minority students under-invest in skills because the returns in the labor market are lower (Neal and Johnson [1996]). If this were the case then the graduation gaps could be seen as a rational response to labor market conditions as students accurately assess that it is not worth it to continue schooling as a minority student compared to their white peers. At the same time, if this were the case any policy aiming to rectify differences in demographic characteristics through government transfers or academic achievement through school programs to raise test scores could never fully close the gap, because it is not the only factor driving the existence of the gap in the first place.

[9]The TEA reports several other types of school district that are omitted for brevity: other central city (15.2% of the sample), other central city suburban (13.0%), non-metropolitan fast growing (1.2%), non-metropolitan stable (7.8%).

However, the results presented above do suggest that policies that address inequality in both socioeconomic levels or academic achievement also address the problem of graduation gaps. These estimates suggest these programs, if properly targeted, could in fact close the graduation gaps for minorities in Texas. The results from the decomposition lend further support to this argument, and additionally suggest that there are not other, unaccounted-for factors in the public education system that would depress graduation rates for minorities and make broad policies less effective.

This paper attempts to further our understanding on the high school diploma gap between white and minority students in Texas and shed light on what observable characteristics contribute to these gaps. It does so first by estimating the graduation rates from public high schools in Texas across racial categories. It then examines specific portions of the student population individually: students who receive free or reduced lunch, students who are classified as high or low ability for either reading or mathematics using standards set by the TEA, and by type of school district attended. Each of these factors matters to the measured graduation gap, but overall the same trends hold. Hispanic students seem to recover more of the diploma gap by simply accounting for individual characteristics. For both black and Hispanic students, controlling for both individual characteristics and test scores eliminates the diploma gap, and in many instances suggests that relative to their white equivalent these minority students actually do better. Finally, including school-level variables further raises the graduation rates for minorities, suggesting that there is also a school quality story at work, but not one that can account for the entirety of graduation gaps on its own.

Having a better knowledge of how some of these major components contribute to the high school diploma gap helps inform future policy in Texas and nation-wide. Obtaining estimates for a large Hispanic population, who seem to have a larger response than black students to simply

accounting for individual demographics, help inform education policy for the future as changing demographic trends continue to increase their proportion of the population in many states. This research supports many of the current trends for graduation rates documented in the literature and also adds to our understanding of exactly how successful specific parts of the population are in obtaining a high school diploma relative to their peers of other races.

Table 2.1: Student Summary Statistics

	All	White	Black	Hispanic	Asian	All Minority
HS Degree	0.78	0.82	0.73	0.74	0.88	0.75
	(0.42)	(0.93)	(0.44)	(0.44)	(0.32)	(0.43)
TAKS Reading	2293.62	2375.04	2251.13	2226.26	2366	2240.44
	(254.74)	(210.77)	(214.38)	(277.60)	(274.46)	(266.65)
TAKS Math	2184.64	2250.89	2112.21	2134.2	2331.02	2141.36
	(218.50)	(206.16)	(182.97)	(214.39)	(243.65)	(215.46)
Math (STD)	0.00	0.30	-0.33	-0.23	0.67	-0.20
	1.00	(0.94)	(0.84)	(0.98)	(1.11)	(0.99)
Reading (STD)	0.00	0.32	-0.17	-0.26	0.29	-0.21
	1.00	(0.83)	(0.84)	(1.09)	(1.08)	(1.05)
Demographics						
Age	13.15	13.12	13.16	13.19	13.03	13.17
	(0.45)	(0.38)	(0.49)	(0.48)	(0.43)	(0.48)
Female	0.49	0.49	0.47	0.49	0.51	0.49
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
LEP	0.07	0.00	0.01	0.15	0.08	0.11
	(0.25)	(0.05)	(0.07)	(0.36)	(0.27)	(0.32)
FRL	0.49	0.18	0.61	0.75	0.3	0.69
	(0.50)	(0.39)	(0.49)	(0.43)	(0.46)	(0.46)
% TAKS Waiver	0.12	0.114	0.167	0.115	0.037	0.124
	(0.363)	(0.318)	(0.373)	(0.320)	(0.189)	(0.329)
N	1882356	743729	255932	812674	70021	1138627
% of Total Sample	1	0.4	0.14	0.43	0.04	0.6
Attrition	185249	49720	29955	102136	2770	135529

Notes: Table 2.1 reports summary statistics for the full sample in Column (1) and for white students only in Column (2), black students in Column (3), Hispanic students in Column (4), Asian students in Column (5), and all minorities together in Column (6). Data come from the 2003-2009 eighth grade cohorts in the Texas Education Agency's administrative data.

Table 2.2: Linear Regressions - Full Sample Males

	(1)	(2)	(3)	(4)	(5)
Black	-0.070***	-0.025***	0.026***	-0.004*	0.038***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Hispanic	-.064***	0.019***	0.049***	0.013***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.074***	0.076***	0.037***	0.053***	0.028***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Other	-0.085***	-0.062***	-0.042***	-0.052***	-0.036***
	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
Math			0.078***		0.071***
			(0.001)		(.001)
Reading			0.043***		0.040***
			(0.001)		(0.001)
Controls					
Age		-0.213***	-0.169***	-0.190***	-0.154***
		(0.001)	(0.001)	(0.001)	(0.001)
FRL		-0.085***	-0.051***	-0.063***	-0.039***
		(0.001)	(0.001)	(0.001)	(0.001)
LEP		-0.152***	-0.031***	-0.154***	-0.045***
		(0.002)	(0.002)	(0.002)	(0.002)
Spec. Ed		-0.120***	-0.039***	-0.082***	-0.016***
		(0.010)	(0.009)	(0.010)	(0.010)
Cohort FE	X	X	X	X	X
School FE				X	X
R-Squared	0.016	0.085	0.135	0.146	0.185
N	962336	962336	962336	962336	962336

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.2 reports estimates of linear probability models of various characteristics on high school diploma receipt for males. Column 1 reports Equation 2.1, Column (2) reports Equation 2.2, Column (3) reports Equation 2.3, Column (4) reports Equation 2.2 with school fixed effecs, and Column (5) reports Equation 2.3 with campus fixed effects.

Table 2.3: Linear Regressions - Full Sample Females

	(1)	(2)	(3)	(4)	(5)
Black	-0.107***	-0.059***	0.005***	-0.019***	0.034***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Hispanic	-0.090***	-0.003***	0.030***	0.002	0.029***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.075***	0.075***	0.038***	0.059***	0.038***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Other	-0.092***	-0.071***	-0.048***	-0.058***	-0.040***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Reading			0.083***		0.076***
			(0.001)		(.001)
Math			0.043***		0.039***
			(0.001)		(0.001)
Controls					
Age		-0.207***	-0.165***	-0.184***	-0.150***
		(0.001)	(0.001)	(0.001)	(0.001)
FRL		-0.093***	-0.059***	-0.068***	-0.046***
		(0.001)	(0.001)	(0.001)	(0.001)
LEP		-0.151***	-0.027***	-0.152***	-0.040***
		(0.002)	(0.002)	(0.002)	(0.002)
Spec. Ed		-0.115***	-0.027***	-0.082***	-0.024***
		(0.007)	(0.007)	(0.008)	(0.007)
Cohort FE	X	X	X	X	X
School FE				X	X
R-Squared	0.021	0.101	0.155	0.161	0.202
N	920020	920020	920020	920020	920020

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.3 reports estimates of linear probability models of various characteristics on high school diploma receipt for for females. Column 1 reports Equation 2.1, Column (2) reports Equation 2.2, Column (3) reports Equation 2.3, Column (4) reports Equation 2.2 with school fixed effects, and Column (5) reports Equation 2.3 with campus fixed effects.

Table 2.4: Linear Regressions: Free and Reduced Lunch Recipients - Males

	(1)	(2)	(3)	(4)	(5)
Black	-0.020***	-0.020***	0.066***	0.033***	.078***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Hispanic	0.050***	0.075***	0.093***	0.067***	0.086***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asian	0.185***	0.176***	0.121***	0.164***	0.118***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Other	-0.064***	-0.056***	-0.043***	-0.054***	-0.040***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.012)
Reading			0.046***		0.044***
			(0.001)		(0.001)
Math			0.103***		0.101***
			(0.001)		(.001)
Controls					
Age		-0.239***	-0.186***	-0.231***	-0.181***
		(0.001)	(0.001)	(0.001)	(0.001)
LEP		-0.153***	-0.009***	-0.155***	-0.016***
		(0.002)	(0.002)	(0.002)	(0.002)
Spec. Ed		-0.138***	-0.044***	-0.125***	-0.034***
		(0.012)	(0.011)	(0.012)	(0.012)
Cohort FE	X	X	X	X	X
School FE				X	X
R-Squared	0.017	0.109	0.174	0.163	0.279
N	475904	475904	475904	475904	475904

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.4 reports estimates of linear probability models of various characteristics on high school diploma receipt for only those males receiving free or reduced lunch. Column 1 reports Equation 2.1, Column (2) reports Equation 2.2, Column (3) reports Equation 2.3, Column (4) reports Equation 2.2 withschool fixed effects, and Column (5) reports Equation 2.3 with campus fixed effects.

Table 2.5: Linear Regressions: Free and Reduced Lunch Recipients - Females

	(1)	(2)	(3)	(4)	(5)
Black	-0.018***	-0.014***	0.043***	-0.024***	.079***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Hispanic	0.027***	0.052***	0.071***	0.060***	0.078***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asian	0.188***	0.176***	0.125***	0.183***	0.143***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Other	-0.060***	-0.058***	-0.047***	-0.053***	-0.041***
	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)
Reading			0.041***		0.039***
			(0.001)		(0.001)
Math			0.107***		0.104***
			(0.001)		(.001)
Controls					
Age		-0.238***	-0.188***	-0.230***	-0.183***
		(0.001)	(0.001)	(0.001)	(0.001)
LEP		-0.148***	-0.009***	-0.149***	-0.015***
		(0.002)	(0.002)	(0.002)	(0.002)
Special Ed		-0.129***	-0.036***	-0.115***	-0.029***
		(0.009)	(0.009)	(0.009)	(0.009)
Cohort FE	X	X	X	X	X
School FE				X	X
R-Squared	0.020	0.097	0.161	0.216	0.265
N	446767	446767	446767	446767	446767

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.5 reports estimates of linear probability models of various characteristics on high school diploma receipt for females on free or reduced lunch. Column 1 reports Equation 2.1, Column (2) reports Equation 2.2, Column (3) reports Equation 2.3, Column (4) reports Equation 2.2 school fixed effects, and Column (5) reports Equation 2.3 with campus fixed effects.

Table 2.6: Linear Regressions: Test Score Spline - Males

	(1)	(2)	(3)	(4)	(5)
Black	0.026***	0.030***	0.026***	0.031***	0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Hispanic	0.049***	0.051***	0.047***	0.051***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.037***	0.041***	0.053***	0.055***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Other	-0.042***	-0.043***	-0.044***	-0.044***	-0.037***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Reading	0.043***		0.034***		
	(0.001)		(0.001)		
Math	0.078***	0.075***			
	(0.001)	(0.001)			
Below Read		0.033***		0.004***	0.005***
		(0.001)		(0.001)	(0.001)
Middle Read		0.127***		0.117***	0.109***
		(0.001)		(0.001)	(0.001)
Commended Read		-0.051***		-0.027***	-0.026***
		(0.001)		(0.001)	(0.001)
Below Math			0.135***	0.136***	0.131***
			(0.001)	(0.001)	(0.001)
Middle Math			0.089***	0.071***	0.061***
			(0.001)	(0.001)	(0.001)
Commended Math			-0.035***	-0.018***	-0.017***
			(0.001)	(0.001)	(0.001)
Controls					
Demographics	X	X	X	X	X
Cohort FE	X	X	X	X	X
School FE					X
R-Squared	0.135	0.142	0.144	0.149	0.197
N	962336	962336	962336	962336	962336

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.6 reports estimates of linear probability models of various characteristics on high school diploma receipt for males with a linear spline for math and reading scores. Column 1 reports Equation 2.3, Column (2) reports Equation 2.2 with math score and linear spline for reading, Column (3) reports Equation 2.2 with reading scores and linear spline for math, Column (4) reports Equation 2.4, and Column (5) reports Equation 2.4 with school fixed effects.

Table 2.7: Linear Regressions: Test Score Spline - Females

	(1)	(2)	(3)	(4)	(5)
Black	0.005***	0.009***	0.008***	0.012***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Hispanic	0.030***	0.032***	0.030***	0.033***	0.030***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Asian	0.038***	0.041***	0.051***	0.053***	0.050***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Other	-0.048***	-0.049***	-0.049***	-0.049***	-0.041***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Reading	0.043***		0.034***		
	(0.001)		(0.001)		
Math	0.083***	0.080***			
	(0.001)	(0.001)			
Below Read		0.032***		0.012***	0.011***
		(0.001)		(0.001)	(0.001)
Middle Read		0.114***		0.103***	0.093***
		(0.001)		(0.001)	(0.001)
Commended Read		-0.046***		-0.025***	-0.024***
		(0.001)		(0.001)	(0.001)
Below Math			0.120***	0.123***	0.117***
			(0.001)	(0.001)	(0.001)
Middle Math			0.116***	0.101***	0.092***
			(0.001)	(0.001)	(0.001)
Commended Math			-0.035***	-0.022***	-0.021***
			(0.001)	(0.001)	(0.001)
Controls					
Demographics	X	X	X	X	X
Cohort FE	X	X	X	X	X
School FE					X
R-Squared	0.155	0.160	0.162	0.165	0.211
N	920020	920020	920020	920020	920020

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.6 reports estimates of linear probability models of various characteristics on high school diploma receipt for females with a linear spline for math and reading scores. Column 1 reports Equation 2.3, Column (2) reports Equation 2.2 with math score and linear spline for reading, Column (3) reports Equation 2.2 with reading scores and linear spline for math, Column (4) reports Equation 2.4, and Column (5) reports Equation 2.4 with school fixed effects.

Table 2.8: Linear Regressions: By School Type - Males

	Major Urban	Major Suburban	Independent	Rural	Charter
Black	0.023***	0.025***	0.088***	0.067***	0.060**
	(0.004)	(0.002)	(0.006)	(0.012)	(0.020)
Hispanic	0.037***	0.025***	0.066***	0.043***	0.059***
	(0.003)	(0.002)	(0.005)	(0.007)	(0.019)
Asian	0.043***	0.026***	0.043**	0.069	0.017
	(0.006)	(0.002)	(0.017)	(0.048)	(0.029)
Other	-0.029	-0.029*	-.034	-0.042	-0.045
	(0.021)	(0.012)	(0.031)	(0.039)	(0.123)
Reading	0.042***	0.038***	0.048***	0.025***	0.039***
	(0.001)	(0.001)	(0.003)	(0.004)	(0.007)
Math	0.085***	0.064***	0.087***	0.044***	0.049***
	(0.001)	(0.001)	(0.002)	(0.004)	(0.007)
Controls					
Age	-0.160***	-0.142***	-0.167***	-0.143***	-0.133***
	(0.003)	(0.002)	(0.005)	(0.008)	(0.010)
FRL	-0.029***	-0.036***	-0.057***	-0.025***	0.014
	(0.003)	(0.002)	(0.005)	(0.006)	(0.012)
LEP	-0.038***	-0.043***	-0.044**	-0.035	-0.009
	(0.004)	(0.004)	(0.012)	(0.019)	(0.021)
Spec. Ed	0.006	-0.030*	-0.015	-0.043	-0.156
	(0.022)	(0.014)	(0.055)	(0.084)	(0.119)
Cohort FE	X	X	X	X	X
School FE	X	X	X	X	X
R-Squared	0.263	0.217	0.217	0.244	0.464
N	178756	324481	58332	32154	10190
% of Sample	18.6	33.7	6.1	3.3	1.1

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.8 reports estimates of linear probability models of high school diploma receipt broken up into school type for males. All estimates are obtained using Equation 2.3 with high school fixed effects. Column (1) reports only for students attending a "Major Urban" school district, Column (2) reports estimates for students enrolled in a "Major Suburban" school district, Column (3) reports estimates for students in "Independent Towns," Column (4) reports estimates for students in "Rural" school districts, and Column (5) reports estimates for only students in charter schools.

Table 2.9: Linear Regressions: By School Type - Females

	Major Urban	Major Suburban	Independent	Rural	Charter
Black	0.024***	0.019***	0.069***	0.054***	0.054*
	(0.004)	(0.003)	(0.007)	(0.013)	(0.023)
Hispanic	0.027***	0.012***	0.048***	0.046***	0.034
	(0.003)	(0.002)	(0.005)	(0.007)	(0.020)
Asian	0.063***	0.030***	0.028	-0.051	0.006
	(0.006)	(0.003)	(0.017)	(0.060)	(0.029)
Other	-0.032	-0.046***	-.078*	-0.023	-0.130
	(0.023)	(0.013)	(0.036)	(0.040)	(0.138)
Ztaks_r	0.039***	0.035***	0.051***	0.031***	0.042***
	(0.002)	(0.001)	(0.003)	(0.004)	(0.008)
Ztaks_m	0.090***	0.072***	0.083***	0.040***	0.048***
	(0.001)	(0.001)	(0.003)	(0.004)	(0.008)
Controls					
Age	-0.159***	-0.144***	-0.160***	-0.118***	-0.120***
	(0.002)	(0.002)	(0.004)	(0.006)	(0.011)
FRL	-0.033***	-0.045***	-0.073***	-0.020***	-0.002
	(0.003)	(0.002)	(0.004)	(0.006)	(0.014)
lep	-0.036***	-0.038***	-0.040**	-0.037*	0.001
	(0.004)	(0.004)	(0.011)	(0.018)	(0.023)
Special Ed	-0.004	-0.023*	-0.017	-0.036	-0.105
	(0.017)	(0.011)	(0.039)	(0.067)	(0.087)
Cohort FE	X	X	X	X	X
School FE	X	X	X	X	X
R-Squared	0.282	0.233	0.234	0.248	0.490
N	167458	314470	65221	31068	9083
% of Sample	18.2	34.2	7.1	3.4	1.0

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table 2.9 reports estimates of linear probability models of high school diploma receipt broken up into school type for females. All estimates are obtained using Equation 2.3 with high school fixed effects. Column (1) reports only for students attending a "Major Urban" school district, Column (2) reports estimates for students enrolled in a "Major Suburban" school district, Column (3) reports estimates for students in "Independent Towns," Column (4) reports estimates for students in "Rural" school districts, and Column (5) reports estimates for only students in charter schools.

Figure 2.1: Graduation Rates by Race, 2003-2009

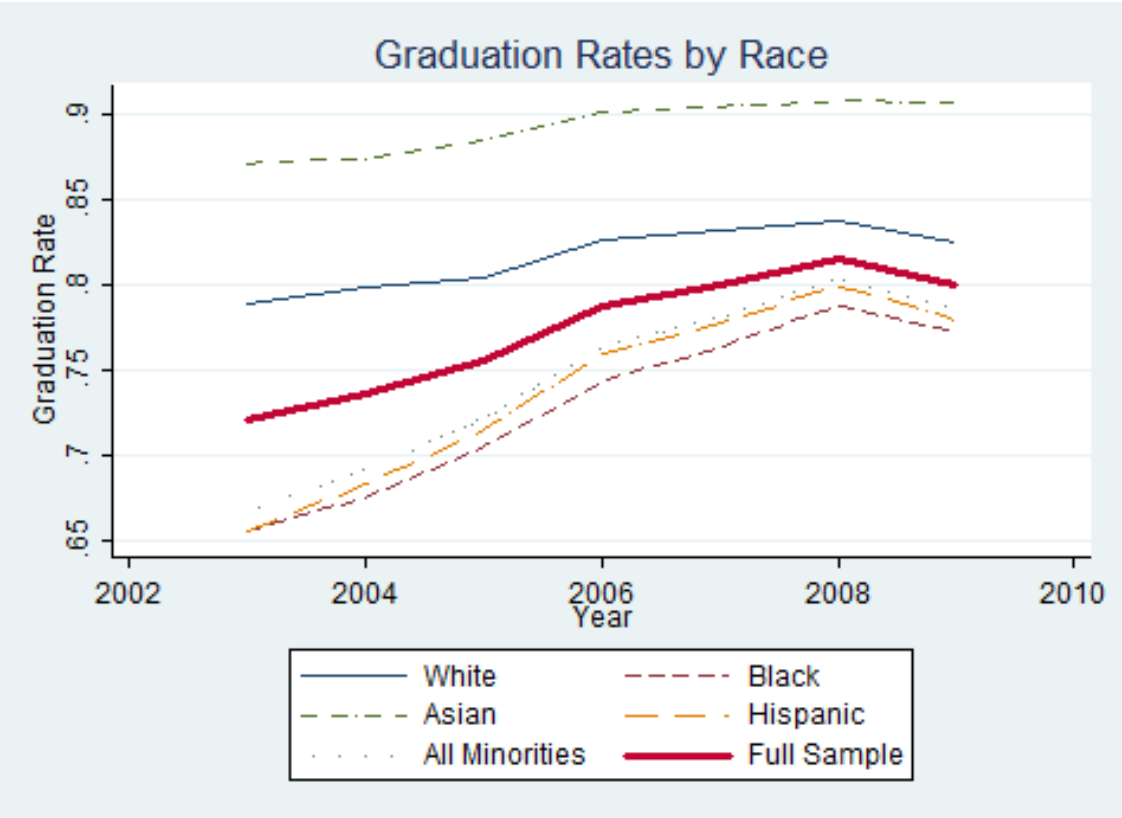


Figure 2.1 displays the overall graduation rate of students in Texas from the 2003-2009 8th grade cohorts and graduation rates for different racial groups over the same period. While overall the graduation rates increase for most of the period, black and Hispanic students are below their white and Asian peers.

Figure 2.2: Distribution of 8th Grade TAKS Math Scores by Race

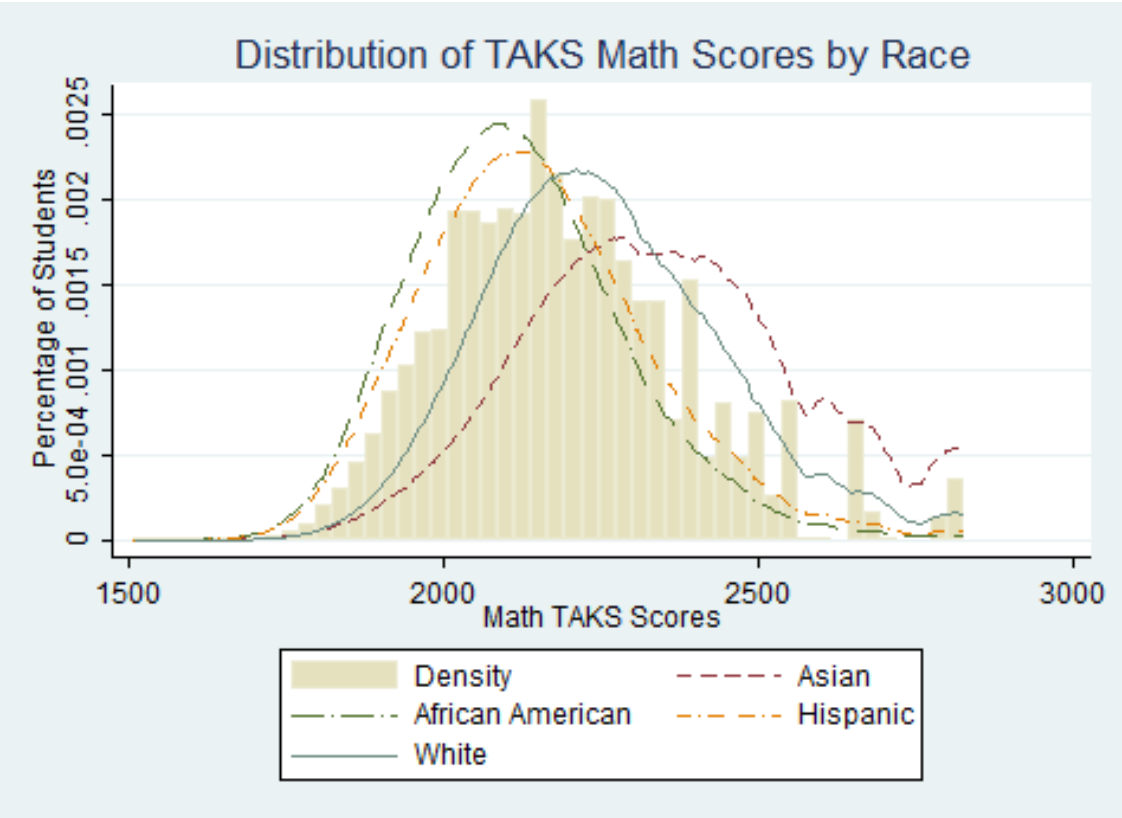


Figure 2.2 displays the distribution of Math TAKS scores for all public eighth-graders in Texas overall and by racial subgroups 2003-2009. The solid red line represents the overall graduation rate of public high school students in Texas. The dashed red line represents black graduation rates, the yellow dashed line represents Hispanic graduation rates, the thin dotted line graphs the graduation rate for all minorities together, the solid blue line graphs graduation rates for white students, and the dashed green line represents graduation rates for Asians.

Figure 2.3: Distribution of 8th Grade TAKS Reading Scores by Race

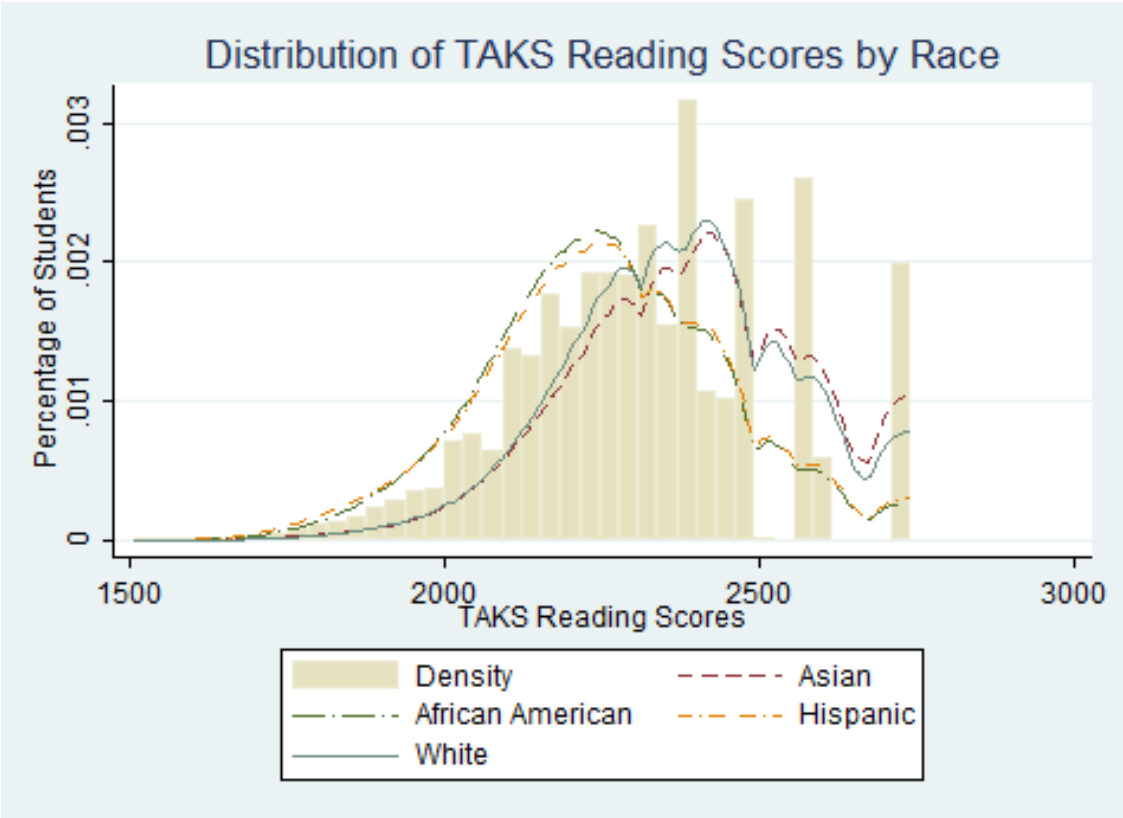


Figure 2.3 displays the distribution of Reading TAKS scores for all public eighth-graders in Texas overall and by racial subgroups. The solid red line represents the overall graduation rate of public high school students in Texas. The dashed red line represents black graduation rates, the yellow dashed line represents Hispanic graduation rates, the thin dotted line graphs the graduation rate for all minorities together, the solid blue line graphs graduation rates for white students, and the dashed green line represents graduation rates for Asians.

Chapter 3

Spreading HOPE? Evaluating Merit-Based Scholarship Programs in Tennessee

3.1 Introduction

The goal of educational mobility has been a part of American society for decades. More recently, merit-based scholarships have begun subsidizing the cost of college for those who do not qualify for needs-based aid. Since the Georgia legislature created the Georgia HOPE (Helping Outstanding Pupils Educationally) Scholarship in 1993, it has become arguably the best known of any state-funded merit scholarship and the inspiration for the federal program under the same name. Several states have implemented a number of programs similar in scale based on the Georgia HOPE scholarship, which offers the total cost of tuition and fees to attend a public university in the state of Georgia for any student who has met certain requirements.

But what effect have these merit-based scholarships had on educational enrollment? Dynarski [2000] estimates that HOPE has increased 18-19 year-old college attendance by 7.0-7.9%. The estimated effect of the HOPE scholarship, while impressive, also implies that only roughly 20% of the total scholarship dollars awarded are spent inducing students who would not otherwise have gone to college to enroll. A full 80% of the money awarded was to high school graduates who would have most likely enrolled in college anyway. Furthermore, the effects seem concentrated in white, upper-middle income segments of the population. Impacts of a program of this scope that

has been repeatedly copied deserve to be well studied to make sure the program is reaching the goals the legislation initially intended.

For all the praise the Georgia HOPE scholarship has received in the decades since its implementation, much less attention has been paid to states who adopted similar programs in the wake of the perceived success of Georgia's merit-based scholarship program. Dynarski [2000] cautions readers that because Georgia was on the low end of educational attainment, the results may not extend to other states. Dynarski also cautions that Georgia's success may be due to the simplicity, transparency, scale and publicity of the program. But have similar merit-based scholarship programs in neighboring states affected educational outcomes the same way? Have similar scholarship programs increased overall college attendance? What distributional effects have these programs had on educational achievement outside the state of Georgia?

Tennessee's Education Lottery Scholarship (TELS) is structured in much the same way as Georgia's HOPE scholarship but is more lenient in student aid eligibility requirements and was implemented without the initial limits on family income eligibility for the merit scholarship. Because a much larger portion of the student population is immediately eligible for scholarships, examining the Tennessee program may shed additional light on the true effectiveness of state-wide merit scholarships. Examining a different state a decade after the initial Georgia scholarship was put in place should offer more evidence of whether the policy is readily extendable for other states.

In keeping with the current literature, this paper examines the effect of introducing the Tennessee Education Lottery Scholarship on college attendance for students. It employs a difference-in-difference framework to evaluate the current college enrollment for students in Tennessee before and after the introduction of the TELS, relative to other students in the United States. It then goes on to examine the effects by family income level to test implications of a model of human capital

accumulation with credit constraints. Finally it offers several robustness checks on the specification of choice and employs the synthetic control method suggested by Abadie et al. [2010] in an effort to get a fuller picture of the true impact of merit scholarships in Tennessee.

The results suggest that the Tennessee Education Lottery Scholarships have not created the anticipated increase in college attendance in the state. While the specific point estimates vary some by specification, age group, relative control group and specific sub-populations evaluated, the overall story is one of no discernible effect, or perhaps a small decrease in college enrollment. While these results conflict with the earliest estimates, they fall in line with the more recent literature that suggests no effects of these high-dollar programs in any of the states they have been implemented in.

The paper is organized as follows. The next section offers a short review of the relevant literature on merit-based scholarships. The third section discusses the Tennessee lottery scholarship program in detail. The fourth section outlines the theoretical framework; the fifth section presents the data used. The subsequent section offers the empirical framework and results. Section seven considers some policy implications; the final section concludes.

3.2 Review of Current Literature

There is a growing body of work on the effect of financial aid on college attendance in general, characterized by an ongoing debate of whether or not merit-based scholarships have an impact on student outcomes. Dynarski [2000, 2008] makes a significant contribution to our understanding of merit-based scholarships by studying whether the Georgia HOPE program increases college enrollment or simply transfers funds to students who are already college-bound. Using the Current Population Survey (CPS) and the Integrated Postsecondary Educational Data System (IPEDS) by

a difference-in-difference analysis of HOPE Scholarship eligibility on the college attendance rate in Georgia, she finds an economically positive and statistically significant effect of the program, as discussed previously.

Cornwell et al. [2009] use IPEDS and National Center for Education Statistics (NCES) data to examine: overall effect on enrollment, enrollment by institution type, enrollment by ethnicity, and the effect of Georgia HOPE on interstate migration. The authors focus on the effect at an institutional level rather than individual effects; the authors examine recent high school graduates who enroll as college freshmen and find small, statistically insignificant results. Their sample focuses only on freshmen who enter college the fall immediately following graduating from high school, and thus they infer that much of the effect Dynarski finds is from students who delayed entry to college initially. They also find a large part of HOPE's effect is keeping students from leaving the state for college and estimate larger, statistically significant gains for blacks in several types of institutions. Several other authors have examined interesting individual-level effects of the HOPE program within specific institutions[1]. Many of these studies show results that comment on unanticipated incentive effects of the Georgia HOPE program, but are too institution-specific to give a clear evaluation of the distributional effects of the merit scholarship.

Sjoquist and Winters [2012] use the same methodology as Dynarski [2008] but use the larger 5% census sample and get point estimates that are much smaller. Sjoquist and Winters also argue clustering standard errors by state and year as Dynarski does greatly understates the confidence interval as shown in Conley and Taber [2011]; using more appropriate econometric

[1] Henry et al. [2004] examine credit hours taken, grade point average, and overall college persistence of HOPE recipients in the University System of Georgia. Cornwell et al. [2005] study course loads, withdrawal rates, and summer enrollment using University of Georgia data. Dee and Jackson [1999] study individual characteristics on Georgia Tech students.

techniques they still find positive effects but they become statistically insignificant.

Tennessee has not had the same level of inquiry into the first-order effects of its own state-level merit scholarship. Pallais [2009] examines some of the student responses in Tennessee to the creation of the TELS using microdata from student ACT scores to examine whether objective scholastic achievement of high school students changed due to the creation of the scholarship. She finds a marginal increase in ACT scores close to the cutoff of 19 to receive funding, that females in Tennessee are more responsive to funding incentive and that African Americans responded little to the scholarship. However, she also finds that the scholarship did not reach a stated goal of "inducing more students to prefer to stay in Tennessee." Sjoquist and Winters [2014] examine the graduation rate of the 25 states that have adopted merit programs, finding no meaningful impact of these programs, which includes examination of Tennessee individually as one part of the bigger picture they try to show. They focus on an older cohort of individuals, aged 24-30, and are more focused on completion rates than attendance. They also estimate their preferred specification pooling all states with merit scholarships in order to estimate the total effect of these types of policies. Concurrent with this research, Bruce and Carruthers [2014] use a Regression-Discontinuity framework for administrative data from Tennessee, finding substitution out of 2-year into 4-year institutions, but no effect on the overall probability of attending college or migration. This research is an important compliment to the empirical strategy employed here, but the regression discontinuity design reports local average treatment effects for students right around the ACT score cutoff. In contrast, my estimates are for the whole sample of college-aged students in Tennessee which should give a clearer sense of the total effect of the program.

This paper proposes to fill in the hole in the literature by evaluating the first order effects of the introduction of the Tennessee Education Lottery Scholarship on current college enrollment in

Tennessee. Focusing on one state rather than imposing additional difficulties in estimation by aggregating a bunch of treated states should generate internally valid effects in Tennessee. Supplying estimates on current college enrollment for different groups using several different control groups should answer questions on overall and distributional effects of the scholarship on college attendance in an attempt to add to the ongoing debate on the effectiveness of merit-based scholarships.

3.3 The Tennessee Education Lottery Scholarship

After repealing a constitutional ban on lotteries, the Tennessee Legislature created the Tennessee Education Lottery Scholarship Program in 2003, a decade after the Georgia HOPE scholarship. Tennessee became the thirteenth state to begin offering state-funded merit scholarships on a grand scale. The first scholarships were awarded to students entering college in the fall of 2004. The TELS requirements are designed to be available to a large set of potential college entrants and have a straightforward application process. In order to be awarded a scholarship, students must be a resident of the State of Tennessee and complete the Free Application for Federal Student Aid (FAFSA).[2] Currently, entering freshmen must also have a minimum score of 21 on the ACT or 980 on the SAT, two standardized tests common for admission at American universities, *or* have a 3.0 high school GPA. At the time of inception the cutoff score was 19 ACT or 890 SAT, which was quickly raised slightly, but since this change is minor and the GPA requirement has not changed it seems reasonable to treat the increase in standards as one that would not have a major effect in the number of awards per year.[3] In order to maintain the scholarship, students must re-file a FAFSA

[2] "State resident" is defined by the TELS as being a resident for one year by September 1 of the year of application date.

[3] All empirical work later will not account for this increase. There is also debate in the legislature to change the requirements to "and" rather than "or," but this change does not affect the time period examined.

each year and meet similar GPA requirements.[4] While enrolled in college, if the TELS scholarship is lost one year, it can be regained in the future if the student raises their GPA sufficiently. Funding can be renewed yearly until the student is awarded their baccalaureate degree (four-year institutions), received the award for five years, or taken more than 120 hours. Unlike in Georgia, there has never been an eligibility cap based on family income for the TELS. Because no segments of the population were initially limited in applying for the scholarship, the eligible population is larger than in Georgia.

While structured in a very similar way to the Georgia HOPE scholarship, there are several key differences between the TELS and HOPE that could alter the impact of the program. In many ways, the Georgia scholarship seems targeted at a much more specific group of recipients. Tennessee has a more uniform application process: every student seeking aid simply fills out a FAFSA. In contrast, Georgia has different requirements by income level: students with family income under $50,000 must fill out a FAFSA; students with a family income above $50,000 complete a one page supplement but no FAFSA. The Georgia HOPE scholarship also has more stringent academic requirements for eligibility. A student in Georgia must be a state resident, have graduated from an eligible Georgia high school, and have at least a B (3.0) grade point average. To qualify for the scholarship in Tennessee, the student must be a Tennessee resident and score either a 21 on the ACT *or* have a 3.0 GPA in high school. Tennessee does not require the high school degree to be from an eligible *in-state* high school, only that they are a state resident.

Additionally, not all Georgia students were initially eligible for the scholarship: in the first

[4]Students must have a GPA of 2.75 or greater at the 24- and 48-semester hours intervals, the one and two year marks, in order to retain the scholarship. Continuing students must have a GPA of 3.0 or above at the 72-semester hour level, the three year mark, and any 24-hour interval above this. Students who have a GPA of 2.75-2.99 may still retain their funding on a semester-by-semester basis.

year (1993) only students with a family income below $66,000 were eligible to receive funds. In 1994 the income cap was raised to $100,000, and in 1995 the income cap was removed completely; however, there has never been an income cap in Tennessee. Initially in Georgia, national Pell Grant awards were subtracted from HOPE scholarship awards; a student receiving the full Pell amount could only receive a $400 book stipend. It was not until 2000 that a student in Georgia could receive both a Pell Grant and a HOPE scholarship. In contrast, Tennessee Pell Grant recipients are eligible for a non-renewable extra stipend in addition to TELS funds. However, while there may be a smaller pool of eligible recipients in Georgia, those that do get scholarships receive more substantial sums than in Tennessee. In Georgia, the award has always been full tuition and required fees at any public university or college in Georgia. In Tennessee the award has always been a dollar amount determined by the legislature that has increased from $3,000 to $4,000 over time. This amount has covered between 63% and 78% of average tuition and fees in Tennessee as reported by the Tennessee Higher Education Commission but has never covered full tuition and fees at public schools.

There is not a clear prediction for how the differences in the eligible base between Georgia and Tennessee will effect response to treatment by a scholarship. It is possible focusing a larger financial incentive on a smaller population targeted at those students with academic credentials more likely to succeed, as in Georgia, will generate more response than in Tennessee. On the other hand, if offering some amount of merit aid to a larger eligible population unambiguously reduces credit constraints and increases enrollment, these differences could create a larger positive effect in Tennessee than has been previously recorded in Georgia.

Finally, it is worth noting that the quality of both in-state and out-of-state students who study in Tennessee may be different than in Georgia. The US News and World Report ranks the

best university in the state of Tennessee as Vanderbilt University, which is ranked 17th nationally. The flagship university in the Tennessee system is the University of Tennessee, which is ranked 101st. In Georgia, the best university in the state is arguably Emory, which is ranked 20th in the nation. Both the Georgia Institute of Technology and the University of Georgia are fine institutions in their own right, ranked 36th and 62nd, respectively. While this type of in-state institutional disparity could signal differing student quality in the states of Tennessee and Georgia, one can also make the argument that in-state attendance at highly competitive private institutions like Emory and Vanderbilt are not much affected by this level of merit aid; the scholarship is not a substantial portion of tuition and paying for education is typically not the only obstacle to entry. If the average quality of higher education for students close to the margin of enrollment does not vary much, any differences in educational quality for the top end of students will not be a limitation of this study. Human capital theory is ambiguous about whether students right at the cutoff for eligibility might be better able to succeed at institutions like the University of Tennessee over more rigorous institutions like Georgia Tech.

3.4 Conceptual Framework

The underlying economic model being tested here is an extension of Becker's Human Capital model with credit constraints as put forth in Dynarski [2000]. Classical Human Capital theory dictates that children from low-income families could be subject to a liquidity constraint when considering higher education that, if binding, will reduce the amount low-income youth invest in higher education to a suboptimal level given idiosyncratic human capital endowments. Working off of the key assumption that the level of debt assumed by a student in a given year of schooling is a decreasing function of parent's income, Dynarski shows that financial aid unambiguously in-

creases the optimal level of schooling, but that students from low-income families should have a more elastic response to any given subsidy. This model is less clear about what effect an increase in aid should have on middle- and upper- income families. This leads to two empirical areas of interest: the magnitude of increases in aid on college attendance in lower-income families and whether the magnitudes and statistical significance change by income level.

Intuitively, creating a substantial scholarship could lower liquidity constraints on those lower- and middle-income students that qualify for and receive funding. Assuming that education is a normal good, this should increase the optimal amount of schooling, and we would expect to see an increase in current college enrollment. Cornwell et al. [2009] point out that in the case of the Georgia HOPE scholarship, the aid decreased the cost of in-state college attendance both overall and relative to the cost of attending college in other states for eligible students. Holding the cost of attending college constant, the Tennessee Education Lottery Scholarship should decrease the cost of attendance in Tennessee relative to other states, causing an increase in both enrollment overall and relative percentage of students enrolled in Tennessee. However, since the Tennessee scholarship is a set dollar amount rather than a waiver for full tuition and fees, it is possible macroeconomic factors or expectations of this aid on the colleges' part could increase the cost of attending college, creating no or a negative net effect on college enrollment.

It is also possible there might be an effect of the scholarship on students who decide to delay entering college for a year or two or who had originally intended never to enroll in college but were enticed by the scholarship to enter at a later age. Alternatively, the scholarship could help students who were enrolled but on the margin of dropping out complete additional years of higher education. Examining an age range for students who would be older first time freshmen, here defined as the 20-22 year-old range, should allow for the ability to examine these outcomes

as well.

3.5 Data

In order to measure the effect of the TELS on Tennessee college enrollment on eligible students, I use the yearly American Community Survey (ACS) data from 2001 to 2010, paired with the 2000 decennial US census data collected from the Public Use Microdata Sample (PUMS).[5] The earlier work discussed in the review of current literature evaluating the Georgia HOPE scholarship controls for regional and state-wide trends by combining two major geographic control groups: all the states in the South Atlantic and East South Central Census regions, and all the states in the Southern Regional Education Board division. This combined control group consists of: Alabama, Arkansas, Delaware, the District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia. In contrast, by the time Tennessee enacted their merit scholarship many more states had programs in existence. Because a majority of the states that enacted merit scholarships are in the South, it is less clear this same group of states is the appropriate control group to examine the TELS with. Instead, the preferred control group for most empirical specifications is all other states in the United States, because of the large sample size it offers and its relevance to the question of whether there should be continued expansion of these merit programs. Several alternate geographic control groups constructed considering merit scholarship programs in place are offered as robustness checks on this selection.[6]

[5]This data was extracted from Ruggles et al. [2010] from the IPUMS website.

[6]Additional geographic control groups considered but not reported include States that only had merit scholarship programs in place before Tennessee, states that physically border Tennessee, states that border Tennessee but did not have a merit scholarship program in place during the period of interest, and the same southern geographic control

I consider my full sample of "Youth" to be the population aged 16-26. This large window is chosen to try and incorporate any students who graduate high school early and students who may delay entry into college but do not so in order to change careers later in life. Summary statistics for the full sample and two subgroups of greater interest, 18-19 year-olds and 20-22 year-olds, are reported in Table 1 along with statistics for the entire United States. While the yearly sample size grows over time, even the smallest-sampled year in Tennessee (2002) has 2,424 students, of which 438 18-19 year-olds were currently enrolled in higher education. The "TN Makeup" panel reports the percentage of white and African American students in Tennessee by year. In every year sampled the population is predominately white and this stays relatively constant over time. African Americans are by far the largest minority in Tennessee and, when combined with white residents, make up almost all of the sample; because of this, some specifications considering minorities restrict attention to African Americans only. The final panel of 3.1 reports the percentage of students currently enrolled in education beyond high school. While there is an upward trend in enrollment for both 18-19 year-olds and 20-22 year-olds both in Tennessee and in the United States as a whole over time, the percentages do not give a clear picture that enrollment had a significant increase around the time the scholarship was implemented.

Most of the subsequent analysis restricts its attention to persons 18 or 19 years of age, as they are the segment of the population that is most likely to be entering higher education for the first time and therefore most likely to be affected by the addition of merit aid.[7] Previous research has highlighted the importance of accounting for family characteristics that may influence college

group as previous work on HOPE without regard to merit scholarship programs in the state.

[7]The ACS survey is administered in such a way that anyone who declines to report educational attainment and current enrollment in school, the two variables of biggest interest, are considered to have "no education" and are "not currently enrolled." This allows for the possibility of attenuation bias in my sample but the low percentage of responses in these categories suggests this is not a major issue.

attendance. The ACS provides data to construct two useful proxies for these family characteristics: family income and family education. I lose about 6% of the whole sample due to missing family income observation, this varies a bit more for younger individuals with 13.5% missing for 18-19 year-olds and 8.5% for 20-22 year-olds. "Family education" is constructed by taking the highest level of education of both the mother and father, to allow for one-family homes. Results are qualitatively similar for defining "family education" to be mother's education. Specifications, tables, and figures are all calculated using the sample weights provided by the Census.

An additional consideration of ACS data is that, unlike "point in time" surveys, the ACS is constructed to reflect yearly averages. Because "currently attending school" is an average from all of 2004 (as opposed to a specific point in time, like the March CPS) and the first scholarships were not given until the fall, it is possible that this could attenuate the effect of the program in the first year.

3.6 Empirical Framework and Results

3.6.1 Empirical Specification

The ideal experiment would estimate the treatment effect of offering a given student a merit scholarship on college attendance. Even under the assumption of perfect information for individuals, which is unreasonable, the researcher can never know the true counterfactual outcome of an individual's enrollment decision conditional on receiving a scholarship or not. However, lack of perfect information confounds this problem substantially. Perhaps the local average treatment effect of the scholarship on the population actually eligible for the award is of the most interest; a dataset that links individuals with high school and post-secondary education records would allow for identification of students who met all the requirements for eligibility and their educational

outcomes would allow for this. Then possible margins to study the impact of a merit scholarship would include students just above and below the cutoff in a given year or students with the same credentials before and after the implementation of the scholarship. Unfortunately, the ACS does not offer specific high school information like GPA or ACT scores for individuals. Because we cannot know on an individual level in the ACS whether a student meets scholarship eligibility, a difference-in-difference strategy is employed to examine average effects on larger groups of people most likely to be affected by the legislation. While this estimate will include some students who were the right age during the period of study but did not meet the scholarship requirements, it is the finest local average treatment effect the ACS allows for estimation.

The implementation of the TELS at a statewide level relative to other southern states lends itself well to difference-in-difference estimation. I begin by examining the level of current college undergraduate attendance by year in Tennessee. Under the difference-in-difference framework, I examine current college attendance in Tennessee and the rest of the United States, before and after the TELS was created. This is done by estimating current college enrollment using a "before" and "after" variable:

$$attend_i = \pi_i + \alpha_n State_i + \beta_t Year_i + \gamma Male_i + \delta Black_i + \theta(TN_i * After_i) + \psi_1 \textit{Family}_i + \epsilon_i \quad (3.1)$$

Where *attend* is an indicator for being currently enrolled as a college undergraduate, *State* is a vector of indicators for the state the student resides in, *Year* is the year of a given observation, *Male* is an indicator for the student being a male, *Black* is an indicator for the student being African American, ψ is a vector for family controls, and *TN*After* is the interaction of Tennessee and an indicator for the scholarship being in place (after 2003). Standard errors are clustered at the state

level. Our parameter of interest is θ, the coefficient on *TN*After*[8]. This model is valid under the assumption that before the introduction of the scholarship college enrollment trends in Tennessee and the rest of the United States are similar.

These trends are graphed for the full sample of youth in Figure 1. The change in current college enrollment over time for 18-19 year-olds can be more clearly seen in Figure 2 and Figure 3 for 20-22 year-olds. Table 3.2 reports the estimates for Equation 3.1 on all 16-26 year-olds, 18-19 year-olds, and 20-22 year-olds, respectively. Robust standard errors are reported in parentheses below each estimate. Column (1) reports estimates for the whole sample of youth using all other states in the United States as the control group, clustered at the state by year level. Surprisingly, the point estimate suggests a 0.4% decrease in college attendance in Tennessee since the TELS has gone into effect, which is statistically significant at the one-percent level. Column (2) reports estimates considering only 18 and 19 year-olds, as they are arguably the group most likely to be enrolled in college and therefore most affected by the scholarship. The point estimate is slightly larger in magnitude, at -0.0111, band is still statistically significant. This estimate of the TELS having a negative effect on college attendance is stronger when considering the subgroup of 20-22 year-olds. Column (3) reports coefficients for Equation 1 considering only this age group, which suggests a 0.6% decrease in college attendance in Tennessee after the scholarship. The confidence intervals do not include zero for any of the three age groups. These initial specifications suggest that instead of finding a more pronounced effect because more students were eligible from the beginning of the Tennessee program, the Tennessee Education Lottery Scholarship has not had a pronounced effect on college attendance.

[8]Equation 1 was also run as a Probit model, but the results did not differ qualitatively from OLS estimates.

Tennessee was one of the last states to introduce a statewide merit scholarship program; only North Carolina has introduced one since 2003. However, it is possible that there are substantial lagged effects or some other regional or time trend affecting estimates that the simplicity of a model such as Equation 3.1 is unable to control for. A more flexible model would allow for both interactions by year and the inclusion of additional family characteristics. This additional flexibility is estimated by:

$$attend_i = \pi_i + \alpha_n State_i + \beta_t Year_i + \gamma Male_i + \delta Black_i + \theta_{n*t}(TN_i * Year_i) + \psi_1 \textit{Family}_i + \epsilon_i \quad (3.2)$$

Where *attend* is an indicator for being currently enrolled as a college undergraduate, *State* is a vector of indicators for the state the student resides in, *Year* is a vector of year dummies, *Male* is an indicator for the student being a male, *Black* is an indicator for the student being African American, *TN*Year* is the vector of interactions between Tennessee and each year during our period of observation and *Family* is a set of additional family covariates. Our parameters of interest are θ for the interaction of Tennessee and all years after 2003, when the scholarship was put in place. This reduced form effect implicitly assumes that all relative effect of attendance changes in Tennessee after 2003 is due to the creation of the TELS. These estimates are reported in Table 3.3, which reports Equation 3.2 for both the 18-19 year-old and 20-22 year-old age groups of interest.

3.6.2 Results

The first column of Table 3.3 reports Equation 3.2 using all other states as the control group. Standard errors of the estimated effects, clustered at the state by year level, are displayed in parentheses in the table. Estimates implying a strong effect of current college enrollment would have small or negative coefficients on the *Year*TN* interactions before 2003, and larger positive effects after. The reported coefficients do not give any indication that implementing the TELS impacted

college attendance in Tennessee over time. Estimates for the *Year*TN* interactions are mostly small, negative, and statistically significant; several confidence intervals include zero. Column (4) reports the same specification for 20-22 year-olds. While the coefficients themselves change a little, the overall pattern of no statistical effect of the scholarship on college enrollment prevails.

While it is interesting to note the Tennessee Education Lottery Scholarship does not seem to have had as large an effect on college attendance in Tennessee as HOPE did in Georgia, it is possible that comparing Tennessee to all other states is not appropriate. At the time HOPE was implemented in Georgia, it was the only state in the region with such a scholarship program in place, thus a southern geographic control group clearly isolated the effect of the program. Columns (2) and (5) estimate Equation 2 for the set of southern control states from the South Atlantic and East South Central Census regions, and all the states in the Southern Regional Education Board division.[9] The estimates are very similar to those for the entire United States in magnitude for the years before implementation of the scholarship. They are slightly more negative after the introduction of the TELS. Using only southern states as controls does not seem to tell a substantially different story.

As an additional consideration, many of the states included in the "Southern" control group have implemented merit scholarship programs of their own, which may alter their own college enrollment trends during the period of interest. Sjoquist and Winters [2014] classifies all 25 states with merit scholarship programs into two categories: weak merit programs, which do not offer much aid, and strong merit programs, which offer substantial aid to a large segment of the population. The states that offered strong merit programs are: Florida, Georgia, Kentucky, Louisiana,

[9] Alabama, Arkansas, Delaware, the District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

Nevada, New Mexico. South Carolina, Tennessee, and West Virginia [10]. Columns (3) and (6) report estimates considering only those states that have no merit scholarship program in place, weak or strong. These additional states that have some sort of merit scholarship and so are excluded are: Alaska, Arkansas, California, Delaware, Idaho, Illinois, Massachusetts, Maryland, Michigan, Mississippi, Missouri, Montana, New Jersey, New York, North Dakota, Oklahoma, South Dakota, Utah, Washington, and Wyoming. Again coefficients for both age groups give the indication that the trend in college enrollment before and after the implementation of the scholarship was not differentially impacted. This table suggests that college enrollment in Tennessee starts and stays below the national trend.

Students who are not eligible to enter the scholarship should not be included in the count of students who are exposed, but don't respond to, treatment. Conditioning on only those students with at least a high school education resolves this issue, but brings up an endogeneity issue of whether students responded to the implementation of the scholarship by accelerating or delaying graduation by a year. Table 4 reports the percentage of individuals with at least a high school education for each year of the sample. While only suggestive, it gives the impression that these students are not, in fact, responding to the scholarship in this way. Figure 3.4 graphs current college enrollment for 18-19 year-olds who have at least a high school education. Taken together, these trends suggest that students are not delaying or accelerating high school graduation in response to the implementation of the scholarship.

The coefficients in Column (1) of Table 3.5 estimate Equation 3.2 for the whole United States and report a stronger, positive statistically significant effect of the *Year*TN* interactions for

[10]Tennessee was not omitted from the sample because it is the state of interest.

the 18-19 year-old age group; however, these estimates are also positive before implementation in 2003, implying that Tennessee high school graduates are more likely to go to college in general and not in response to the scholarship. 20-22 year-olds, reported in Column (4), exhibit a pattern of no discernible effect with estimates on both sides of zero. Equation 3.2 is reported in Column (2) for 18-19 year-olds and Column (5) for 20-22 year-olds considering only the states in the Southern control group, conditional on high school completion. Columns (3) and (6) consider only those states that never implemented a merit scholarship program of any type. Using either of these alternate control groups does not change the estimates qualitatively. In all specifications, "Controls" includes both family education and family income. While these results should shed some light on the effect of the TELS, they do not take into account enrollment by type of institution. Overall, Table 3.5 gives the impression that the Tennessee Education Lottery Scholarship does not seem to be affecting college attendance in Tennessee in the way policy-makers intended.

While there does not seem to be much of an effect on college enrollment for either age for any control group, distributional effects are also an important aspect of the program; it is worth examining interesting subsamples to see if certain groups were more affected. Table 3.6 reports estimates for Equation 3.2 for the subpopulation of minorities and for African Americans specifically. Columns (1) and (4) report coefficients considering only African Americans using the rest of the United States as a control group. The magnitude of the coefficients for 18-19 year-olds is more negative and statistically significant; while it still offers no clear prediction it seems relative to African Americans in the rest of the United States, they go to college less frequently after the introduction of the scholarship. Columns (2) and (5) consider all minorities and are similar to estimates for only African Americans. Columns (3) and (6) consider all minorities only in the states that never implemented a merit scholarship. Estimates for 20-22 year-olds are mildly

positive and statistically significant in many years, telling a story of perhaps some older students entering college but even here several confidence intervals include zero and, especially with the estimated effect on younger minority students, there seems to again be a story of no real effect caused by the introduction of the TELS.

The previous tables suggest that the Tennessee Education Lottery Scholarship has had little to no positive effect on current college enrollment in Tennessee. However, the theory of human capital being tested focuses on the impact of the scholarship on different ranges of the income distribution. Table 3.7 examines the impact of the Tennessee Education Lottery Scholarship by income thresholds for both the 18-19 year-old and 20-22 year-old age group. This is done by estimating:

$$\begin{aligned} attend_i = \pi_i + \alpha_n State_i + \beta After_i + \gamma Male_i + \delta Black_i + \theta_{n*t}(TN_i * After_i) \\ + \rho(TN * IncomeBin_i) + \omega(IncomeBin_i * After_i) \\ + \mu(TN * After * IncomeBin_i) + \phi IncomeBin_i + \eta FamEduc_i + \epsilon_i \end{aligned} \tag{3.3}$$

Where *attend* is a binary indicator of current undergraduate enrollment, *TN*Year* is the interaction of Tennessee and year dummy variables, *Male* is an indicator for the student being a male, *Black* is an indicator for the student being African American, *IncomeBin* is an indicator for total family income as reported to the ACS segmented into $25,000 bins, and *FamEduc* is the maximum reported level of education between both parents. In this saturated model, the parameter of interest is the interaction of Tennessee and After with each income category. All coefficients in Table 3.7 were estimated using Equation 3.2, relative to the year 2000 and the $125,000-$150,000 income range. This range was used because families with income at this level may have the financial means to allow children to enroll in higher education without any credit constraints and should have a fairly inelastic demand for higher education as a group. Columns (2) and (4) consider the same specifi-

cation for only those states that had no merit scholarship program in place throughout the period of interest. The point estimates would suggest an increase in attendance for younger students in the middle- and upper-income levels. This result is in line with distributional arguments in favor of the idea that merit scholarships seem to subsidize those who would have already enrolled in a university, rather than induce more students to attend in the first place. This may be due to lessened credit constraints or additional parental support. Taken together, these different specifications give the idea that there has not been a significant effect on college attendance in Tennessee after the introduction of the TELS.

3.6.3 Synthetic Control Method

The previous discussion has made it clear that due to the geographic differences and change in merit scholarship policy at the state level over time, there is not a clear "best control group". While the preferred control groups are sensible, it might be worthwhile to implement the synthetic control method as put forth by Abadie et al. [2010]. The goal is to construct a "synthetic Tennessee" out of weighted characteristics of all the United States. Doing so will make explicit the contribution of each individual observation in the control group to the construction of the synthetic Tennessee. This process will also make explicit the similarities and differences of the control and treatment groups. Since the econometrician is trying to construct a control group that does not violate the parallel trends assumption in the pre-treatment period, it is important to have a clear picture of whether this is a reasonable assumption or not.

One drawback of this approach is that in order to code it properly, there can only be one unit (or region) that is treated, so the individual-level observations from the ACS must be collapsed down into the state by year level. Because a lot of the idiosyncratic variation is thrown out, these

estimates can be seen more of a robustness check than a fundamentally different way to answer the questions above. Because of this, several additional macro-level controls are added to try to pick up more variation on the state level. State unemployment rates, from 2000-2010, are gathered from the Bureau of Labor Statistics in an attempt to control for economic conditions. Current College Attendance, lagged by one and two years, is included to pick up underlying population differences in college attendance by state. Finally, the average log family income, average education level (measured in years) and percentage of the population that is male and percentage that is black are included.

Table 3.8 reports the weights estimated for 18-19 year-olds. The constructed Tennessee draws from eight states, several of which are in "strong merit group"; others are not geographically close to Tennessee. This list suggests that it is not easy to draw an appropriate control group simply by intuition. The second panel reports the values of each control for the synthetic Tennessee constructed using the weights in Table 3.8. Overall, the synthetic Tennessee seems to do a fairly good job matching in the pre-treatment period, matching more closely on the non-unemployment controls. Ideally, we would like to see a real and synthetic Tennessee moving in concert in the pre-treatment period. Then, after the introduction of the scholarship in 2004, we would see the real Tennessee have a noticeable increase in current college attendance over the synthetic Tennessee control group. However, Figure 5 tells a different story. The real and synthetic Tennessee do not seem to diverge in the post-reform period. Table 3.9 reports the weights estimated for 20-22 year-olds. The group of eight states comprising the synthetic Tennessee for the 20-22 year-olds changes slightly, but again includes some states with large merit programs in place and some states that do not seem to share much in common with Tennessee on a map (like Rhode Island). Panel B reports the values of each control variable for the second synthetic Tennessee using the weights from Table

3.8. Figure 6 tells a similar story to that for the 18-19 year-olds. In both cases Tennessee does not seem to deviate from the control group after the introduction of the scholarship, implying that the program is not creating a noticeable effect.

3.7 Policy Implications

By and large, the estimates in the preceding section lend further evidence to suggest that merit scholarships in Tennessee are not having any noticeable effect on college attendance. What is driving these results? Perhaps Tennessee-specific program features are preventing the scholarship from having a noticeable effect on college attendance. While the Tennessee Higher Education Commission does not offer program evaluation at nearly as rigorous a level of analysis as this paper provides, they do offer summary statistics which, paired with these estimates, shed more light on who benefits from the program and who does not. Perhaps a difference in the quality of the Tennessee Student compared to students in other states in the control group is driving the results. Of qualifying recipients in 2010, only 56% of students in Tennessee met both the GPA and standardized test score requirements. This implies that almost half of the Tennessee recipients who received money would not have even qualified for funds in other states. It is possible that these marginal students enroll and pay fees but do not attend, or drop out of courses so quickly they are not counted as an "attending student" by the ACS. This may also be suggested in the fact that the renewal rate for the Fall 2009 scholarship cohort (receiving funds in the Fall of both 2009 and 2010) was only 54%. While not all of these dropped out of college entirely and may have only fallen below the 3.0 GPA minimum, this again suggests a high rate of attrition even after receiving funds.

Perhaps the negative effects on minorities are driven by some program-specific feature. The

program may be simply subsidizing the cost of college for students from demographics that would have already attended anyway, rather than inducing new portions of the population to enroll. A full 61% of scholarship recipients are from families whose income is $48,000 a year or more. This lines up with the suggestive (but statistically insignificant) results from Table 3.7 that the program only has an effect on the middle- and upper-income brackets.

Perhaps the credit constraints for poorer students are still binding even at these subsidy levels. The monetary stipend for a TELS scholarship covered 63% of a public four-year institution's 2010 tuition and fees. Historically, this percentage cost of attendance covered by the scholarship award has hovered between 60-80 percent for the TELS. This does not include consideration of what percentage of the total cost of attending college, considering textbooks room, and board, that the scholarship covers.

3.8 Conclusion

This research evaluates the impact of merit scholarships in Tennessee on current college enrollment. Using yearly data from the ACS, I use a difference-in-difference framework to examine whether there are higher college attendance rates among the population eligible for the scholarship (Tennessee residents after 2003). Then, it considers a synthetic control method to try to get around the issue of selecting a reasonable control group in a period when many other states implemented similar merit scholarships. The estimates suggest that among most demographic groups the scholarships have no discernible effect. This finding holds up regardless of the specific age range considered or geographic control group.

While perhaps at odds with the initial findings from Dynarski [2000], these results, or rather a lack of an estimated effect, are in line with more recent literature regarding merit scholarships

both as a whole and related work on Tennessee specifically. Using a regression discontinuity strategy and focusing on students very close to the threshold of eligibility on the ACT test, Bruce and Carruthers [2014] also find that merit scholarships do not significantly change the likelihood of attending any college. They point out that it is possible students farther away from the ACT cutoff could be responding to the merit scholarship but not picked up because they are outside the bandwidth of the regression discontinuity. Coupled with the estimates reported here, which implicitly use students across the range of ACT scores —unavailable in this data set —add evidence that students in other ranges aren't responding in differential ways.

Sjoquist and Winters [2014] also support the lack of an estimated impact of merit scholarship programs in Tennessee found here with research combining different programs across different states. There are important differences between their work and the research presented here, but overall both findings are consistent. Sjoquist and Winters [2014] are more focused on older students and eventual completion rates. In contrast, this focuses instead on younger students and their immediate enrollment decisions. I also compliment their work by using the same control group of "no merit" scholarship states in some specifications, and even extend it by employing a synthetic control method to remove any possible biases in selection of a control group. While examining Tennessee specifically is a less researched piece in the growing merit scholarship literature, this paper can be taken as mounting evidence that the scholarship programs have not been as effective as they were intended to be.

Table 3.1: Summary Statistics on College Enrollment, 2000-2010

Variables	All US		Tennessee		
	Observations	Attendance (%)	Observations	Attendance (%)	% Black
2000	2,032,547		40,906		17.66
18-19	400,595	34.97	7,918	32.21	
20-22	542,532	37.79	11,052	34.02	
2001	152,907		2,717		14.13
18-19	27,294	32.74	468	29.49	
20-22	38,927	40.56	704	37.78	
2002	136,995		2,424		14.03
18-19	24,318	34.11	438	32.65	
20-22	35,134	42.63	617	37.76	
2003	153,816		2,758		13.42
18-19	27,159	34.4	450	30.44	
20-22	39,438	43.47	699	36.48	
2004	153,363		2,794		14.03
18-19	27,175	35.89	487	27.31	
20-22	39,324	45.2	720	38.61	
2005	367,717		7,299		15.52
18-19	65,178	36.12	1,245	30.9	
20-22	92,218	43.52	1,861	36.97	
2006	404,422		8,085		17.13
18-19	80,177	44.52	1,576	42.01	
20-22	102,769	45.41	2,030	40.05	
2007	409,687		8,123		15.57
18-19	82,004	45.94	1,611	43.08	
20-22	105,401	46.45	2,139	42.45	
2008	405,301		8,083		16.19
18-19	80,313	45.2	1,590	40.63	
20-22	102,817	45.88	1,986	40.58	
2009	411,096		8,216		17.25
18-19	82,817	46.33	1,596	40.91	
20-22	105,751	46.86	2,084	39.73	
2010	414,888		8,130		18.09
18-19	82,488	47.14	1,627	41.79	
20-22	108,961	47.72	2,145	41.77	

Notes: Table 3.1 reports summary statistics on college enrollment for the whole United States and Tennessee from the American Community Survey. College Attendance and Racial Makeup reported in unweighted percentages. Observations reports the raw number of persons in the sample.

Table 3.2: Simple Difference Estimates on Effect of Tennessee Merit Scholarship by Age Group

	16-26 year-olds	18-19 year-olds	20-22 year-olds
	(1)	(2)	(3)
Male	-0.050***	-0.099***	-0.116***
	(0.0008)	(0.0018)	(0.0021)
Black	-0.032***	-0.067***	-0.091***
	(0.0016)	(0.0031)	(0.0036)
Fam Inc	Y	Y	Y
Fam Educ	Y	Y	Y
TN $*$ After	-0.005	-0.011	0.006
	(0.0061)	(0.0104)	(0.0098)
Observations	2,524,563	620,222	569,005
R-squared	0.029	0.069	0.100

Robust standard errors in parentheses clustered State by Year
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 3.2 reports ordinary least squares regression of current college attendance on State, Year, Male, Black, family controls, and an interaction term for TN*After, including a constant using data from the American Community Survey. Column (1) reports this for the whole sample of youth, 16-26 year-olds. Column (2) reports this for the subsample 18-19 year-olds. Column (3) reports estimates for 20-22 year-olds.

Table 3.3: Estimates of Equation 3.2 Including Covariates by Age Group

	18-19 Year-Olds			20-22 Year-Olds		
	All States (1)	South (2)	No Merit (3)	All States (4)	South (5)	No Merit (6)
TN (2001)	-0.009**	-0.005	-0.016***	-0.011*	-0.014*	-0.012
	(0.0038)	(0.0045)	(0.0053)	(0.0059)	(0.0080)	(0.0073)
TN (2002)	0.027***	0.028***	0.027***	0.005	-0.001	0.009
	(0.0042)	(0.0059)	(0.0054)	(0.0064)	(0.0079)	(0.0071)
TN (2003)	-0.026***	-0.024***	-0.025***	-0.020***	-0.025***	-0.025***
	(0.0040)	(0.0047)	(0.0064)	(0.0064)	(0.0080)	(0.0075)
TN (2004)	-0.020***	-0.020***	-0.025***	0.037***	0.032***	0.025***
	(0.0042)	(0.0051)	(0.0069)	(0.0063)	(0.0075)	(0.0087)
TN (2005)	-0.010***	-0.014***	-0.020***	-0.028***	-0.028***	-0.039***
	(0.0031)	(0.0039)	(0.0046)	(0.0054)	(0.0071)	(0.0061)
TN (2006)	0.002	-0.002	-0.005	0.027***	0.027***	0.017***
	(0.0039)	(0.0051)	(0.0058)	(0.0054)	(0.0073)	(0.0064)
TN (2007)	0.006	-0.001	0.008	-0.005	-0.005	-0.019***
	(0.0047)	(0.0065)	(0.0052)	(0.0054)	(0.0071)	(0.0063)
TN (2008)	-0.028***	-0.032***	-0.035***	0.012**	0.007	0.006
	(0.0038)	(0.0052)	(0.0047)	(0.0051)	(0.0069)	(0.0069)
TN (2009)	-0.022***	-0.025***	-0.023***	-0.023***	-0.023***	-0.027***
	(0.0031)	(0.0043)	(0.0042)	(0.0051)	(0.0066)	(0.0065)
TN (2010)	-0.012***	-0.017***	-0.010*	-0.013**	-0.014*	-0.018***
	(0.0044)	(0.0058)	(0.0054)	(0.0055)	(0.0073)	(0.0061)
Controls	Y	Y	Y	Y	Y	Y
Observations	620,222	416,932	269,718	569,005	386,342	238,324
R-squared	0.069	0.068	0.060	0.100	0.090	0.111

Robust standard errors in parentheses clustered State by Year
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 3.3 reports coefficients for Equation 3.2 including several family controls for both age groups. Columns 1 and 4 include Tennessee by year indicators including family income, Columns 2 and 5 use a group of southern control states Columns 3 and 6 use states without a merit scholarship.

Table 3.4: Summary Statistics for HS graduation by Year

	Census year										
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
						16-26					
All US											
	1,250,219	97,944	87,676	100,879	101,492	238,962	268,230	274,870	272,306	280,573	282,741
	61.51	64.05	64.00	65.58	66.18	64.99	66.32	67.09	67.19	68.25	68.15
TN											
	25,803	1,729	1,555	1,786	1,852	4,777	5,326	5,518	5,503	5,662	5,649
	63.96	63.76	65.27	66.34	67.25	67.76	67.38	69.04	69.17	70.16	70.88

Notes: Table 3.4 reports total and percentage of individuals with at least a high school education for each year in the sample for the whole United States and Tennessee specifically. A dip right before the enactment of the policy in 2003 in Tennessee, or a spike in graduation from high school, could indicate students were aware of the policy and delayed graduation because of this. However, trends look fairly smooth both in Tennessee and in the United States over time.

Table 3.5: Estimates Conditional on High School Graduation

	18-19 Year-Olds			20-22 Year-Olds		
	All States	South	No Merit	All States	South	No Merit
	(1)	(2)	(3)	(4)	(5)	(6)
TN (2001)	-0.012**	-0.000	-0.018**	-0.007	-0.011	-0.006
	(0.0057)	(0.0068)	(0.0084)	(0.0074)	(0.0100)	(0.0089)
TN (2002)	0.021***	0.023***	0.035***	-0.011	-0.015	-0.005
	(0.0066)	(0.0078)	(0.0067)	(0.0081)	(0.0104)	(0.0082)
TN (2003)	0.014**	0.020***	0.015*	-0.029***	-0.033***	-0.038***
	(0.0060)	(0.0072)	(0.0088)	(0.0073)	(0.0098)	(0.0080)
TN (2004)	-0.023***	-0.015**	-0.030***	0.045***	0.040***	0.031***
	(0.0062)	(0.0068)	(0.0088)	(0.0076)	(0.0099)	(0.0093)
TN (2005)	0.015***	0.018***	0.007	-0.034***	-0.035***	-0.046***
	(0.0048)	(0.0058)	(0.0056)	(0.0069)	(0.0094)	(0.0076)
TN (2006)	0.024***	0.025***	0.018**	0.031***	0.030***	0.019**
	(0.0056)	(0.0071)	(0.0079)	(0.0073)	(0.0101)	(0.0078)
TN (2007)	0.044***	0.047***	0.047***	0.002	0.002	-0.015*
	(0.0061)	(0.0080)	(0.0058)	(0.0073)	(0.0098)	(0.0076)
TN (2008)	-0.019***	-0.018***	-0.023***	0.012*	0.006	0.005
	(0.0050)	(0.0062)	(0.0051)	(0.0066)	(0.0092)	(0.0083)
TN (2009)	-0.004	0.001	-0.002	-0.032***	-0.032***	-0.038***
	(0.0051)	(0.0070)	(0.0053)	(0.0066)	(0.0087)	(0.0081)
TN (2010)	-0.005	-0.001	-0.005	-0.019***	-0.020**	-0.025***
	(0.0055)	(0.0071)	(0.0065)	(0.0069)	(0.0094)	(0.0075)
Controls	Y	Y	Y	Y	Y	Y
Observations	376,720	256,017	159,470	495,113	340,240	209,075
R-squared	0.060	0.060	0.055	0.070	0.062	0.081

Robust standard errors in parentheses clustered State by Year
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 3.5 reports estimates for Equation 3.2 for those students with at least a high school education. Columns 2 and 5 consider only those students in the southern control states. Columns 3 and 6 consider states with no merit scholarship program in place.

Table 3.6: Estimates for Minority Groups

	18-19 Year-Olds			20-22 Year-Olds		
	Afr. Amer	All Minority	No Merit	Afr. Amer	All Minority	No Merit
	(1)	(2)	(3)	(4)	(5)	(6)
TN (2001)	-0.016*	-0.035***	-0.022	-0.035***	0.027***	-0.014*
	(0.0087)	(0.0060)	(0.0158)	(0.0113)	(0.0079)	(0.0075)
TN (2002)	0.022**	-0.028***	0.006	-0.054***	0.011	0.008
	(0.0097)	(0.0075)	(0.0158)	(0.0102)	(0.0078)	(0.0071)
TN (2003)	0.000	-0.049***	-0.016	0.013	-0.002	-0.025***
	(0.0072)	(0.0067)	(0.0115)	(0.0173)	(0.0101)	(0.0073)
TN (2004)	-0.135***	-0.101***	-0.141***	0.119***	0.117***	0.025***
	(0.0097)	(0.0068)	(0.0187)	(0.0107)	(0.0104)	(0.0085)
TN (2005)	0.057***	0.026***	0.054***	-0.045***	-0.004	-0.040***
	(0.0091)	(0.0057)	(0.0131)	(0.0098)	(0.0071)	(0.0062)
TN (2006)	-0.053***	-0.054***	-0.049***	0.078***	0.103***	0.017**
	(0.0081)	(0.0057)	(0.0150)	(0.0082)	(0.0066)	(0.0065)
TN (2007)	-0.041***	-0.008	-0.049***	0.014	0.052***	-0.019***
	(0.0089)	(0.0061)	(0.0127)	(0.0092)	(0.0068)	(0.0063)
TN (2008)	-0.036***	-0.042***	-0.052***	0.127***	0.105***	0.006
	(0.0074)	(0.0049)	(0.0122)	(0.0097)	(0.0076)	(0.0069)
TN (2009)	0.004	-0.016***	-0.012	0.038***	0.067***	-0.028***
	(0.0081)	(0.0054)	(0.0139)	(0.0010)	(0.0079)	(0.0066)
TN (2010)	-0.031***	-0.022***	-0.029*	-0.030***	-0.015**	-0.019***
	(0.0084)	(0.0053)	(0.0153)	(0.0087)	(0.0063)	(0.0062)
Controls	Y	Y	Y	Y	Y	Y
Observations	74,851	165,990	27,989	72,174	163,946	238,324
R-squared	0.090	0.084	0.088	0.094	0.088	0.107

Robust standard errors in parentheses clustered State by Year
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 3.6 reports estimates for Equation 3.2 for the subpopulation of minority students. Columns 1 and 4 consider African Americans. Columns 2 and 5 consider all minorities. Columns 3 and 6 consider all minorities for states with no merit scholarship program in place.

Table 3.7: Estimates for Merit Scholarship by Income Bin

	18-19 Year-Olds		20-22 Year-Olds	
	All US	No Merit	All US	No Merit
	(1)	(2)	(3)	(4)
TN	0.017*	-0.036**	0.013	0.010
	(0.0093)	(0.016)	(0.0076)	(0.0131)
After	0.047***	-0.007	0.012	0.026*
	(0.0106)	(0.0189)	(0.0091)	(0.0134)
TN*After	-0.080***	-0.049**	-0.007	-0.023*
	(0.0105)	(0.0188)	(0.0095)	(0.0131)
<$25,000	-0.152***	-0.154***	-0.119***	-0.129***
	(0.0089)	(0.0167)	(0.0120)	(0.0152)
$25,000-$50,000	-0.108***	-0.113***	-0.089***	-0.086***
	(0.0096)	(0.0179)	(0.00837)	(0.0129)
$50,000-$75,000	-0.063***	-0.079***	-0.056***	-0.052***
	(0.0092)	(0.0220)	(0.0075)	(0.0110)
$75,000-$100,000	-0.026***	-0.050***	-0.035***	-0.037***
	(0.0077)	(0.0177)	(0.0076)	(0.0117)
$100,000-$125,000	-0.002	-0.028	-0.023***	-0.019
	(0.0086)	(0.0197)	(0.0077)	(0.0133)
TN*After <$25,000	0.027***	0.018	-0.016*	-0.012
	(0.0096)	(0.0194)	(0.0092)	(0.0136)
TN*After $25,000-$50,000	0.083***	0.050***	0.061***	0.074***
	(0.0114)	(0.0176)	(0.0091)	(0.0114)
TN*After $50,000-$75,000	0.091***	0.051**	-0.005	0.005
	(0.0112)	(0.0238)	(0.0087)	(0.0139)
TN*After $75,000-$100,000	0.064***	0.016	-0.015*	-0.016
	(0.0096)	(0.0171)	(0.0090)	(0.0147)
TN*After $100,000-$125,000	0.059***	0.016	0.013	0.027
	(0.0107)	(0.0234)	(0.0105)	(0.0182)
Controls	Y	Y	Y	Y
Observations	609,328	154,472	552,126	231,601
R-squared	0.078	0.062	0.106	0.116

Robust standard errors in parentheses clustered by State
*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 3.7 reports estimates for Equation 3.3 for both age groups. Columns 1 and 3 use the full "All States" sample; Columns 2 and 4 consider only those states in the "No Merit" group.

Table 3.8: Synthetic Control Weights and Variables for 18-19 Year-Olds

Panel A: Synthetic Control Weights

State	Unit Weight
AK	.001
GA	.309
KY	.15
MI	.093
NV	.063
ND	.135
OR	.082
SC	.168

Panel B: Real and Synthetic Variables

Variable	Tennessee	Synthetic TN
One-Year Lag	.299	.298
Two-Year Lag	.293	.296
Black	.205	.205
Log Inc.	10.423	10.448
Male	.508	.517
Family Educ.	76.300	75.900
Unemp. 2000	4.0	3.8
Unemp. 2001	4.7	4.6
Unemp. 2002	5.3	5.4
Unemp. 2003	5.7	5.7
Unemp. 2004	5.4	5.5
Unemp. 2005	5.6	5.5
Unemp. 2006	5.2	5.2
Unemp. 2007	4.8	5.0
Unemp. 2008	6.6	6.3
Unemp. 2009	10.5	10.0
Unemp. 2010	9.8	10.0

Notes: Panel A of Table 3.8 reports the relative weights using the synthetic weighting process for 18-19 year-olds and all fifty states. Any state not listed has a weight of zero in the synthetic Tennessee. Panel B displays the precision of variable match between the synthetic Tennessee and the Tennessee variables in the data as a check for how closely the synthetic Tennessee matches the real state.

Table 3.9: Synthetic Control Weights and Variables for 20-22 Year-Olds

Panel A: Synthetic Control Weights

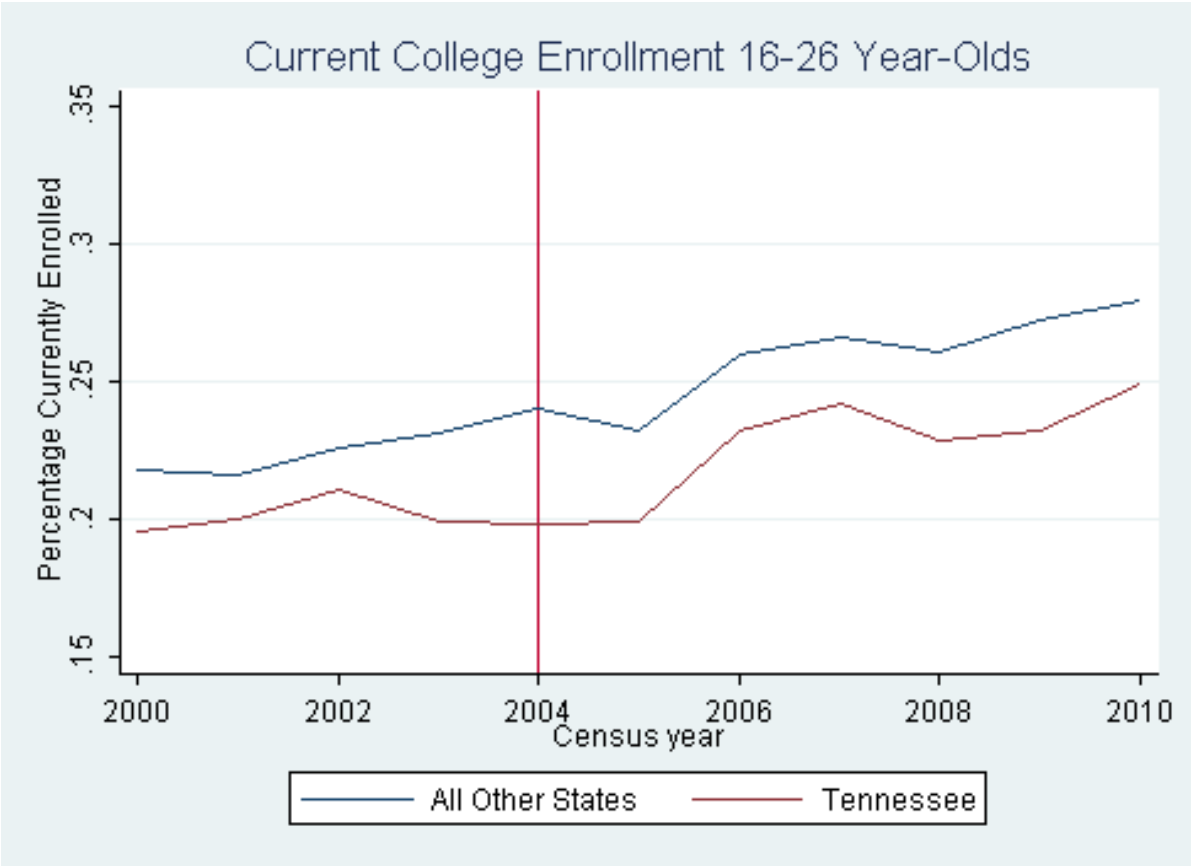
State	Unit Weight
AL	.092
GA	.189
KY	.254
ME	.148
MI	.066
NV	.052
RI	.042
SC	.157

Panel B: Real and Synthetic Variables

Variable	Treated	Synthetic
One-Year Lag	.345	.346
Two-Year Lag	.346	.345
Black	.189	.180
Log Inc.	10.112	10.155
Male	.495	.497
Family Educ.	73.351	73.654
Unemp. 2000	4.0	3.8
Unemp. 2001	4.7	4.7
Unemp. 2002	5.3	5.4
Unemp. 2003	5.7	5.8
Unemp. 2004	5.4	5.4
Unemp. 2005	5.6	5.5
Unemp. 2006	5.2	5.3
Unemp. 2007	4.8	5.1
Unemp. 2008	6.6	6.4
Unemp. 2009	10.5	10.3
Unemp. 2010	9.8	10.4

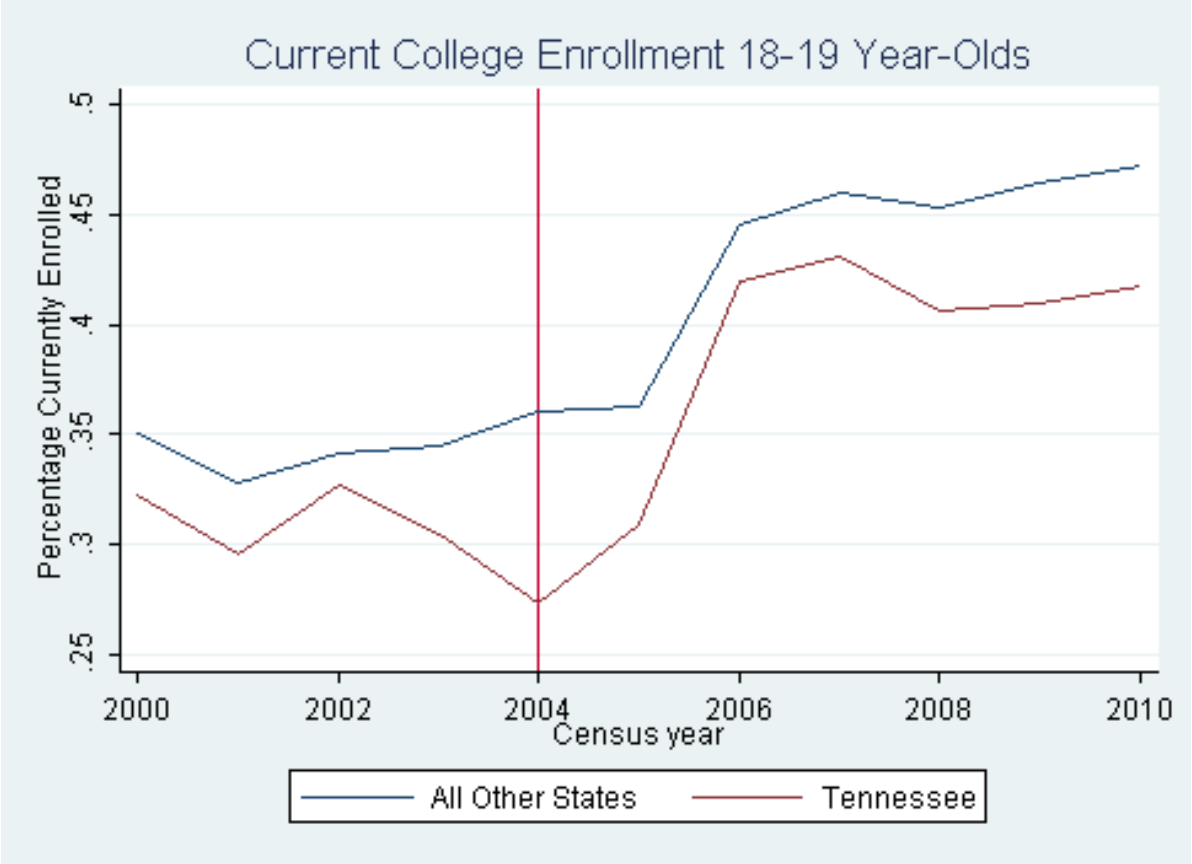
Notes: Panel A of Table 3.9 reports the relative weights using the synthetic weighting process for 20-22 year-olds and all fifty states. Any state not listed has a weight of zero in the synthetic Tennessee. Panel B displays the precision of variable match between the synthetic Tennessee and the Tennessee variables in the data as a check for how closely the synthetic Tennessee matches the real state.

Figure 3.1: Average college enrollment in Tennessee by year for 16-26 year-olds



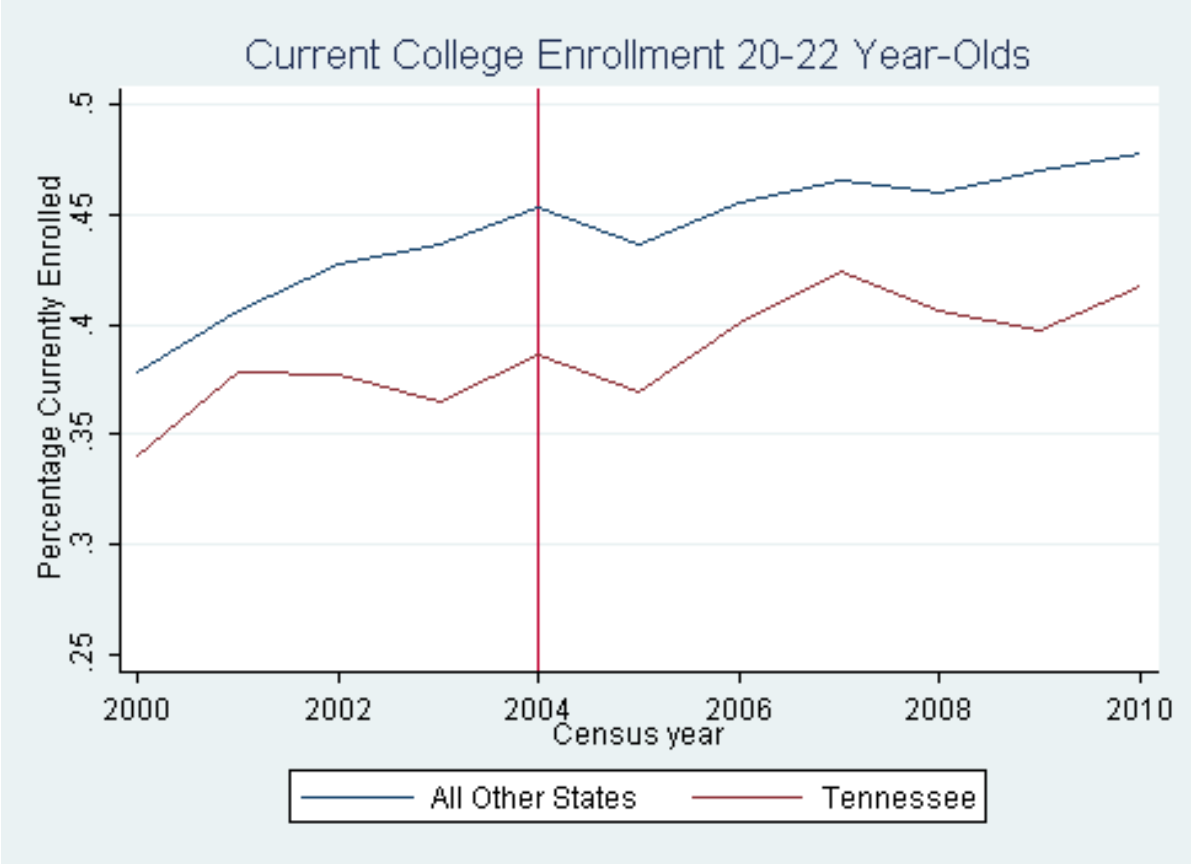
Notes: Figure 3.1 graphs the average college enrollment in Tennessee versus the rest of the United States for all youth aged 16-26. "All States" uses all observations available from the rest of the United States.

Figure 3.2: Average college enrollment in Tennessee by year for 18-19 year-olds



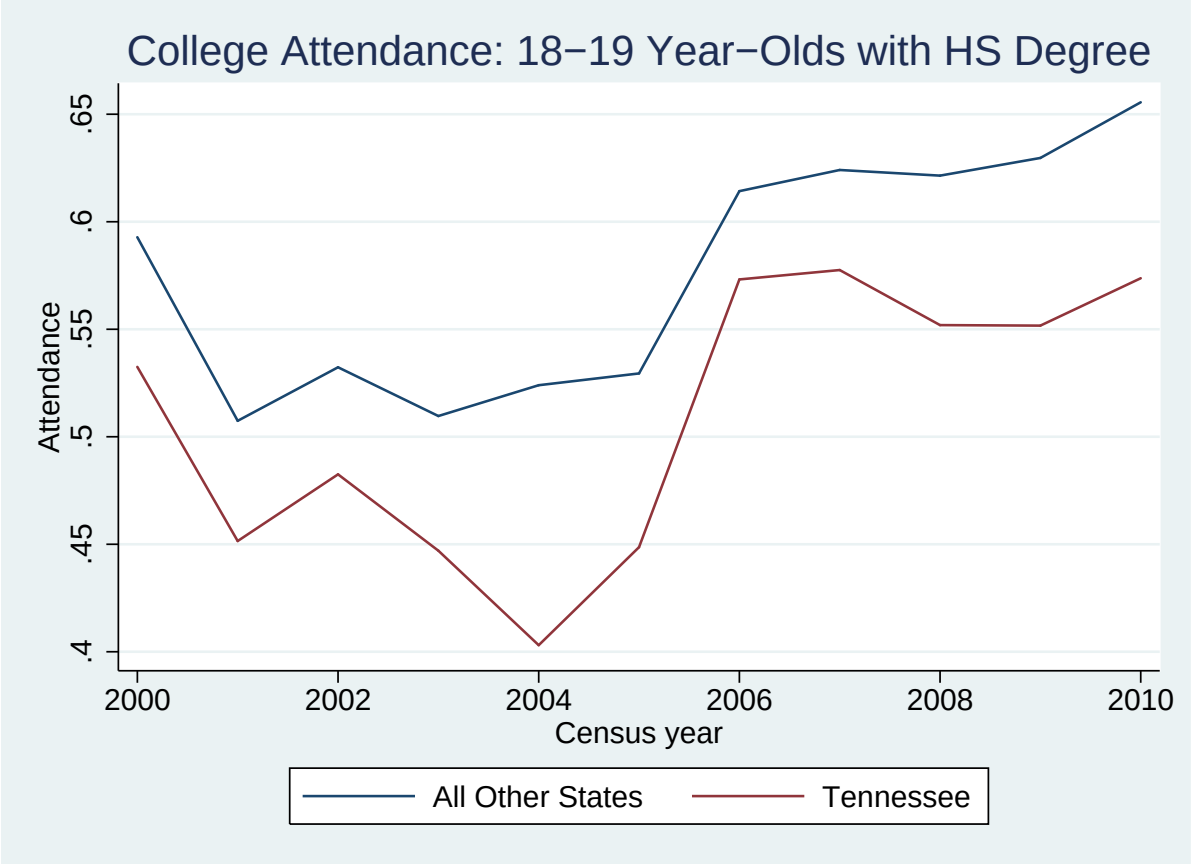
Notes: "All States" uses all observations available from the rest of the United States.

Figure 3.3: Average College Enrollment in Tennessee by Year for 20-22 Year-Olds



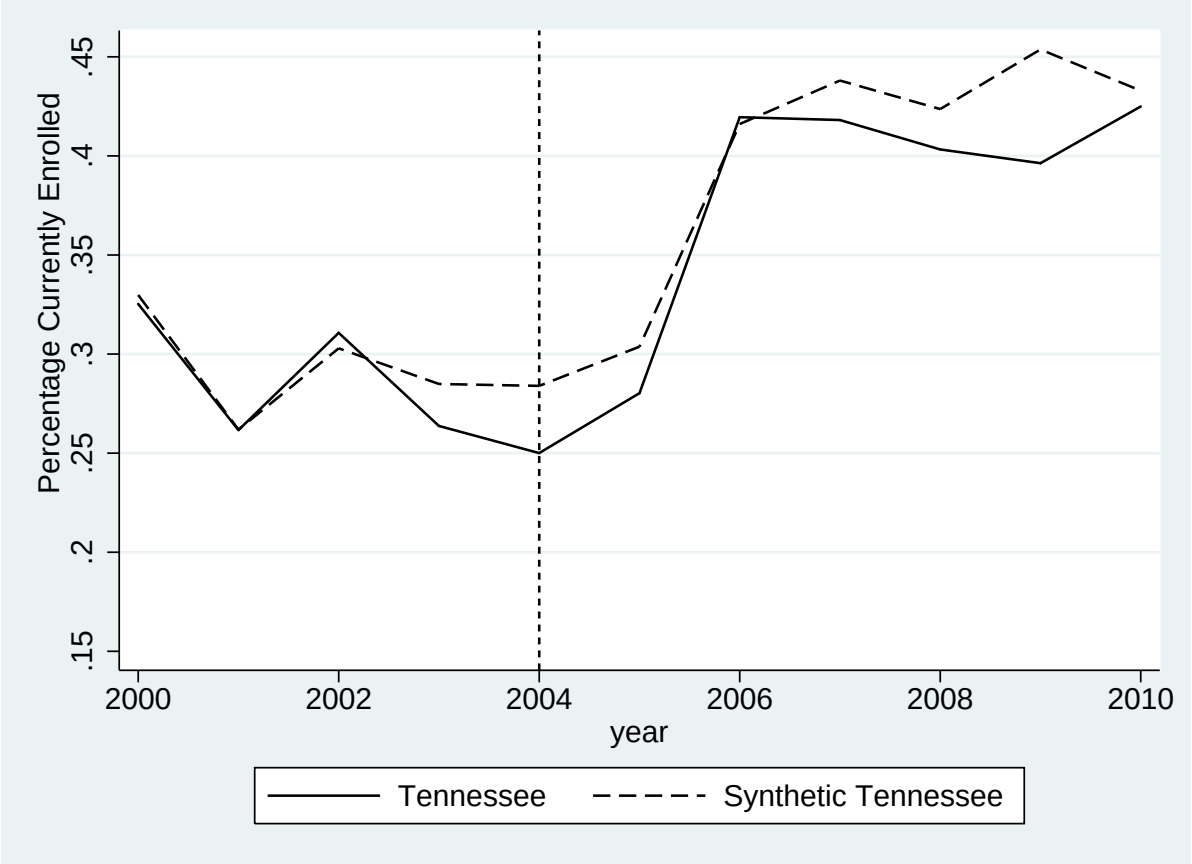
Notes: Figure 3.3 graphs college enrollment for 20-22 year-olds for the whole United States and Tennessee from 2000 to 2010.

Figure 3.4: Average college enrollment in Tennessee by year for 18-19 year-olds with a High School degree.



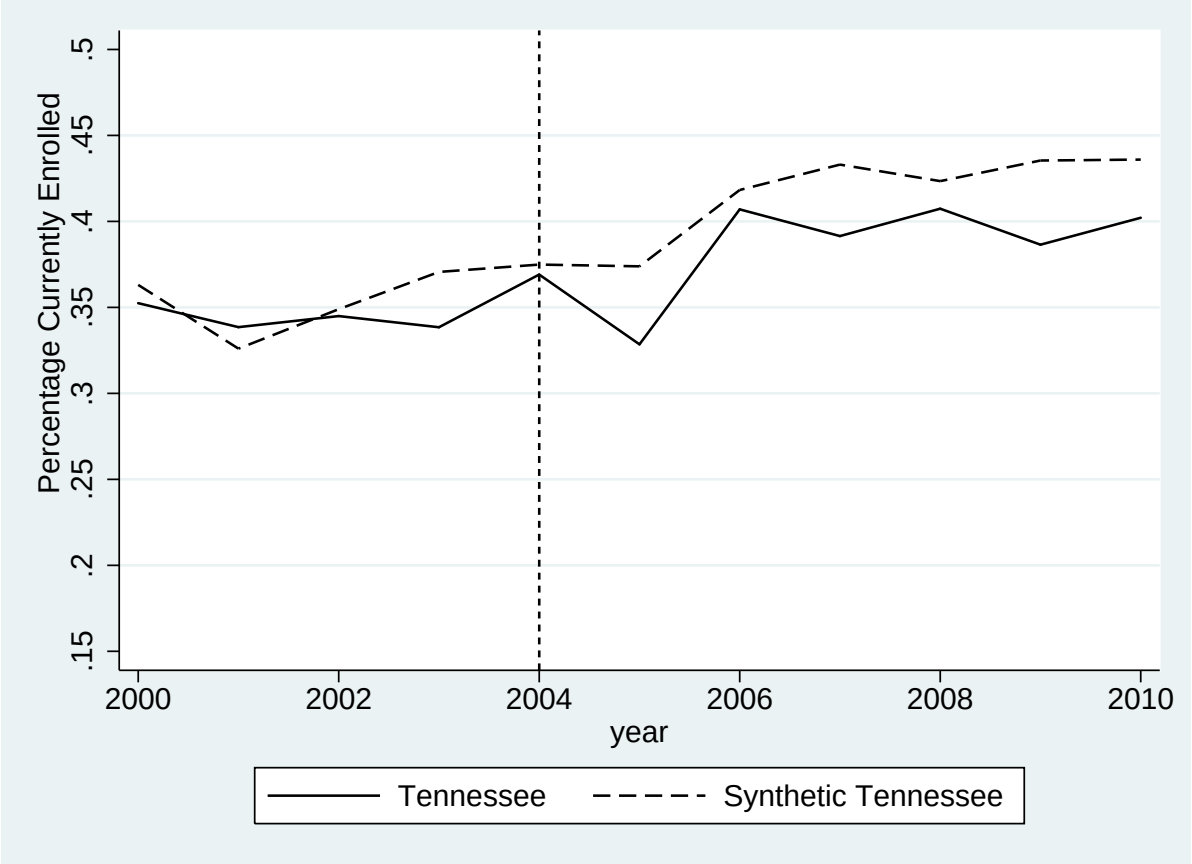
Notes: Figure 3.4 graphs average college enrollment in Tennessee by year for 18-19 year-olds with a High School degree.

Figure 3.5: With Lag and Unemployment Controls for 18-19 Year-Olds



Percentage of current enrollment for Tennessee and the Synthetic Tennessee constructed.

Figure 3.6: With Lag and Unemployment Controls for 20-22 Year-Olds



Percentage of current enrollment for Tennessee and the Synthetic Tennessee constructed.

Appendices

Appendix A

TAKS-ing Students? Texas Exit Exam Effects on Human Capital Formation - Appendix

This Appendix includes several additional tables and figures to accompany Chapter 1. Table A.1 reports reduced-form estimates of passing a given portion of the exam on the probability of obtaining a high school diploma for each of the five administrations. Moving across columns displays subsequent administration for a cohort. For three of the four segments (English, Mathematics, and Science) as the number of retakes increases failing an exam has a larger, more negative impact on the probability of receiving a high school diploma. These estimates support previous findings in the related literature.

Table A.2 reports estimates of the impact of failing a given portion of the TAKS exam on the number of courses taken in several additional subjects, mostly elective credits for the student. Column (1) reports the impact of failing a given portion of the exam on the number of fine art courses taken in the senior year, Column (2) reports the change in number of physical education credits taken senior year. Column (3) reports the number of credits foreign language, Column (4) reports the total number of credits taken across all TAKS subjects, and Column (5) reports the total number of high school credits taken during senior year.[1] While the point estimates are slightly larger than in many other specifications in Column (5), they can be taken as evidence that students

[1] Foreign Language credits include any language at any level, which varies by school and diploma track.

who fail a portion of the exam do not consistently take more high school credits than students who pass. This is important because it suggests that the extra courses in failed subjects are not a mechanical response.

Table A.3 reports estimates from Equation 1. the first-stage coefficients for African American and Hispanic students (analogous to Table 1.2). Looking at the coefficients reported on the interactions by race, African American students to not differ substantially in the probability of being classified as meeting the standard based on their score. Hispanic students, in contrast, do have statistically significant interaction terms.

Table A.4 reports estimates of failing a given portion of the exam on additional courses taken in that subject using several alternative specifications as a check on the stability of the results. Column (1) reports the preferred specification, for easy comparison. Column (2) reports a quadratic specification, Column (3) reports the preferred specification with covariates included as well, Column (4) reports a wider bandwidth at seven questions, and Column (5) reports the widest estimated bandwidth at nine questions. Across all portions of the exam and alternate specifications, results are fairly stable but lessen as the bandwidth is widened.

Table A.5 reports the impact of failing a portion of the TAKS exam on the probability of dropping out of high school before the end of senior year as an additional check on stability. The same alternate specifications are reported in the same order as Table A.4. Again, results appear fairly stable between alternate specifications but wash out as the bandwidth continues to widen.

Figure A.1 displays an example of the report a student receives about their performance on each of the four sections on the TAKS exam. Each portion of the student's performance on the TAKS exit exam is reported in a different panel. Several important features of this score report

stand out. As noted by the report date in the top left corner, students get their results roughly one month after the administration of the spring exam, giving them plenty of time to adjust their behavior and course load for their senior year. Also, each portion of the exam clearly displays the student's (scaled) score on that portion of the exam, the minimum score for the exam, and how the student did relative to that score in a bar graph.

Finally, Figure A.1 displays the percentage of students obtaining a high school diploma by score on the " last chance" exam administered the spring of the senior year for students. Panel (a) displays percentage of students obtaining a high school diploma for the English portion of the exam, Panel (b) displays this for Mathematics, Panel (c) for Social Studies, and Panel (d) for the Science portion of the exam. In support of previous literature, students who fail a portion of the exit exam in the "last chance" administration are less likely to graduate form high school, but additionally I show that these effects are much more pronounced for mathematics and science.

Table A.1: High School Diploma Receipt by TAKS Administration

	First Admin	Second Admin	Third Admin	Fourth Admin	Fifth Admin
English	0.005	0.003	-0.014	-0.063**	-0.048*
	(0.007)	(0.014)	(0.015)	(0.020)	(0.024)
N	102106	25359	19077	9842	6458
F-Test	0.549	0.037	0.796	9.604	4.031
R-Squared	0.006	0.006	0.011	0.018	0.023
Math	-0.025***	-0.027*	-0.067***	-0.149***	-0.209***
	(0.007)	(0.011)	(0.067)	(0.013)	(0.016)
N	92944	33615	36521	23660	14743
F-Test	13.695	5.877	39.239	125.420	165.424
R-Squared	0.004	0.007	0.015	0.045	0.072
Social Studies	-0.005	0.017	0.012	-0.018	-0.049
	(0.007)	(0.014)	(0.016)	(0.021)	(0.029)
N	116727	20863	15111	7628	3675
F-Test	0.565	1.453	0.524	0.697	2.925
R-Squared	0.008	0.005	0.009	0.017	0.023
Science	-0.007	-0.012	-0.059***	-0.091***	-0.159***
	(0.006)	(0.009)	(0.010)	(0.011)	(0.014)
N	105579	45748	45094	32583	20742
F-Test	1.400	1.582	38.014	65.854	130.623
R-Squared	0.000	0.004	0.013	0.037	0.053

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table A.1 reports estimates of the impact of passing a given section of the TAKS on high school diploma receipt over the number of administrations. Each column reports a sequentially later administration for a given cohort.

Table A.2: Elective Courses in High School

	Art	P.E.	Language	TAKS Courses	All Courses
English	-0.035***	0.000	0.011	0.010	0.057
	(0.003)	(0.006)	(0.009)	(0.050)	(0.039)
N	26282	26282	26282	21201	21201
F-Test	125.629	0.001	1.464	0.039	2.173
Mean	0.695	0.461	0.504	7.461	11.473
Mathematics	-0.036***	-0.023**	0.009	0.031***	-0.215***
	(0.006)	(0.009)	(0.008)	(0.004)	(0.002)
N	34902	34902	34902	31593	31593
F-Test	31.837	7.109	1.285	75.697	8511.469
Mean	0.745	0.457	0.495	7.287	11.493
Social Studies	-0.035***	0.001	-0.043***	-0.106***	-0.226***
	(0.003)	(0.003)	(0.004)	(0.002)	(0.005)
N	43131	43131	43131	34256	34256
F-Test	114.760	0.085	135.003	3072.837	2388.149
Mean	0.674	0.448	0.506	7.568	11.468
Science	-0.026***	-0.007	-0.005*	0.096***	0.104***
	(0.002)	(0.005)	(0.002)	(0.004)	(0.020)
N	39349	39349	39349	35870	35870
F-Test	253.441	1.959	5.276	698.337	28.237
Mean	0.745	0.454	0.495	7.222	11.453

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table A.2 estimates the effect of failing a portion of the TAKS exam on the number of courses in other subject matters taken the senior year of high school. Column (1) reports art classes taken, Column (2) reports the number of physical education classes taken, Column (3) reports the number of foreign language courses taken, Column (4) reports the total number of classes taken in one of the four TAKS subjects, and Column (5) reports the total number of courses taken the senior year of high school.

Table A.3: First Stage Estimates For Subgroups

	English	Mathematics	Soc. Stud.	Science
Afr. Amer				
Passing Score	0.804***	0.792***	0.765***	0.791***
	(0.008)	(0.007)	(0.007)	(0.007)
Interaction	-0.024	-0.006	-0.006	0.010
	(0.017)	(0.017)	(0.015)	(0.016)
N	32978	42711	52749	49469
F-Test	2.002	0.122	0.148	0.353
R-Squared	0.764	0.711	0.688	0.682
Hispanic				
Passing Score	0.759***	0.755***	0.723***	0.765***
	(0.011)	(0.011)	(0.010)	(0.010)
Interaction	0.064***	0.056***	0.064***	0.048***
	(0.014)	(0.013)	(0.013)	(0.013)
N	32978	42711	52749	49469
F-Test	20.319	17.320	25.183	14.254
R-Squared	0.765	0.712	0.688	0.684

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table A.3 estimates Equation 1.3, the effect of failing a portion of the TAKS exam on the probability of not meeting the standard in the TEA files for specific populations of students: African Americans and Hispanics.

Table A.4: Stability of Specification: Advanced Courses

	Preferred (BW5)	Quadratic	Covariates	BW 7	BW 9
English	-0.030***	-0.045***	-0.027***	-0.020**	-0.037***
	(0.005)	(0.003)	(0.004)	(0.007)	(0.009)
N	35438	50222	32784	50222	65271
F-Test	31.234	216.930	38.174	9.112	17.962
Mean	0.422	0.426	0.412	0.426	0.431
Mathematics	0.032*	0.064***	0.016*	0.004	-0.005
	(0.013)	(0.008)	(0.008)	(0.014)	(0.014)
N	34902	48472	31892	48472	61884
F-Test	6.397	63.802	4.275	0.085	0.146
Mean	0.562	0.559	0.556	0.559	0.553
Social Studies	-0.002	0.028***	-0.004	-0.022**	-0.027*
	(0.003)	(0.002)	(0.003)	(0.008)	(0.011)
N	43131	60182	40255	60182	77001
F-Test	0.295	197.267	1.407	7.087	5.995
Mean	0.423	0.428	0.418	0.428	0.433
Science	0.028***	0.037***	0.028**	0.014**	0.020***
	(0.006)	(0.005)	(0.010)	(0.005)	(0.005)
N	39349	54344	36079	54344	69411
F-Test	23.845	57.496	7.937	8.429	13.519
Mean	0.583	0.573	0.578	0.573	0.562

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table A.4 displays the estimated impact of failing a portion of the exit exam on advanced courses taken senior year using different specifications and bandwidths as a check on the stability of the results. Panel 1 reports advanced courses taken after failing the English portion, Panel 2 reports advanced courses for failing the mathematics portion of the exam, Panel 3 the social studies portion, and Panel 4 the science portion. Column (1) reports the preferred linear specification at a bandwidth of five, Column (2) includes a quadratic term, Column (3) is a linear specification that includes individual controls, Column (4) estimates the impact using a linear specification but a wider bandwidth of seven points, and Column (5) uses a linear specification but a bandwidth of nine.

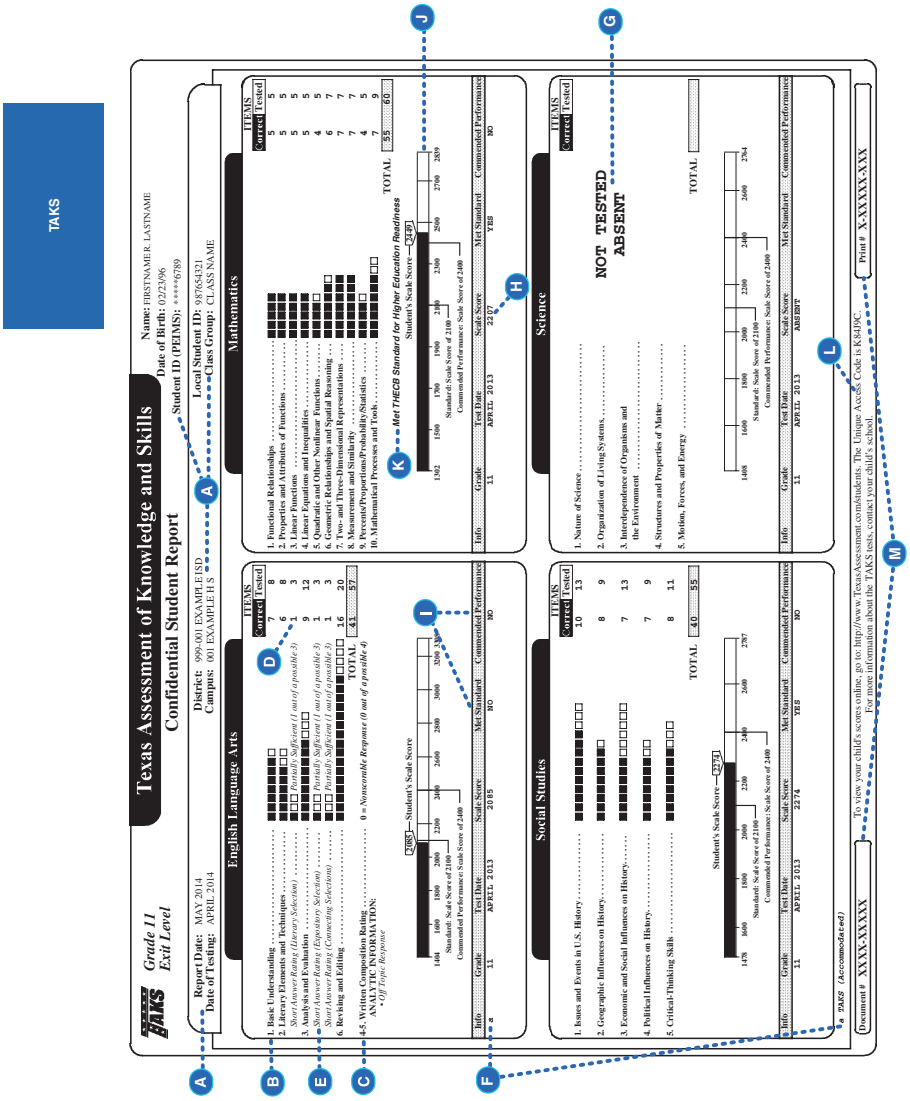
Table A.5: Stability: Dropout

	Linear	Quadratic	Covariates	BW 3	BW 7
English	-0.023***	-0.050***	-0.025***	-0.006	-0.004
	(0.005)	(0.002)	(0.006)	(0.009)	(0.009)
N	43852	62119	36674	62119	80585
F-Test	20.298	641.194	15.097	0.493	0.184
Mean	0.195	0.193	0.221	0.193	0.190
Mathematics	0.036***	0.036***	0.041***	0.036***	0.024***
	(0.001)	(0.000)	(0.002)	(0.002)	(0.006)
N	42711	59310	34651	59310	75704
F-Test	962.318	17865.284	653.557	318.414	18.824
Mean	0.096	0.098	0.109	0.098	0.102
Social Studies	-0.007***	-0.004***	-0.004***	-0.009***	-0.008***
	(0.002)	(0.001)	(0.000)	(0.002)	(0.002)
N	52749	73734	45077	73734	94443
F-Test	19.054	15.035	227.962	26.984	18.919
Mean	0.201	0.199	0.222	0.199	0.197
Science	0.012***	0.015***	0.020***	0.012***	0.008
	(0.000)	(0.001)	(0.001)	(0.003)	(0.005)
N	49469	68360	39376	68360	87067
F-Test	54493.883	258.376	844.394	17.920	2.875
Mean	0.090	0.095	0.103	0.095	0.100

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Table A.5 displays the estimated impact of failing a portion of the exit exam on dropping out of high school using different specifications and bandwidths as a check on the stability of the results. Panel 1 reports dropout after failing the English portion, Panel 2 reports dropout for failing the mathematics portion of the exam, Panel 3 the social studies portion, and Panel 4 the science portion. Column (1) reports the preferred linear specification at a bandwidth of five, Column (2) includes a quadratic term, Column (3) is a linear specification that includes individual controls, Column (4) estimates the impact using a linear specification but a wider bandwidth of seven points, and Column (5) uses a linear specification but a bandwidth of nine.

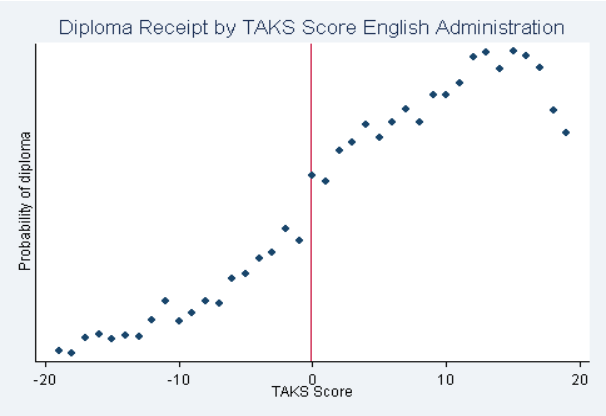
Figure A.1: Example TAKS Score Report



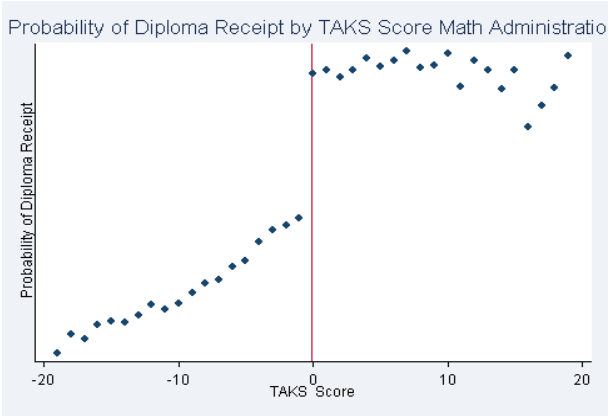
Note: Figure A.1 is an example of the mailing students receive from the TEA, indicating their score and how close they were to the standard.

Figure A.2: High School Diploma Receipt by TAKS Scores by Subject.

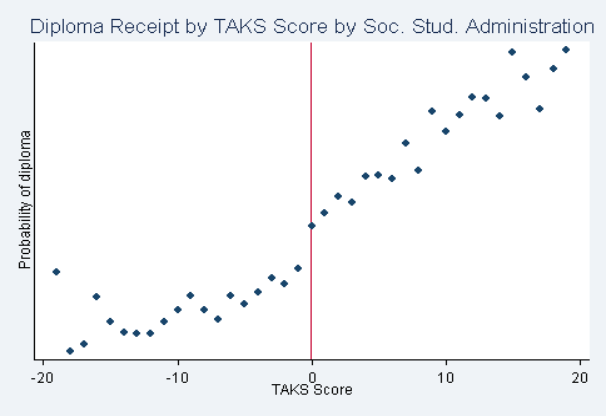
(a) English



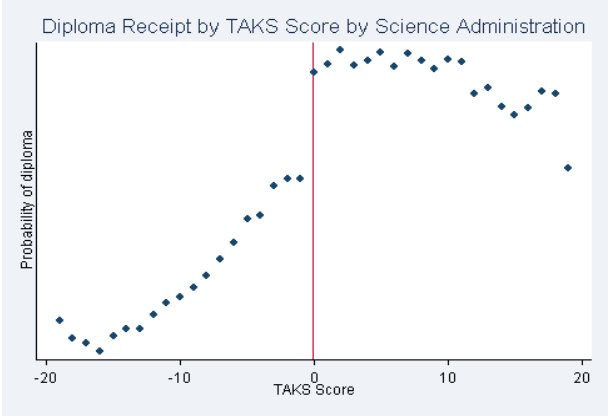
(b) Math



(c) Social Studies



(d) Science



Note: Figure A.2 displays the percentage of students earning a high school diploma for students in the "last chance" exam the spring of their senior year by score on a given portion of the TAKS exam. Panel (a) displays high school diploma receipt for the English portion of the exam, Panel (b) displays high school diploma receipt for the mathematics portion of the exam, Panel (c) displays the social studies portion of the exam, and Panel (d) the science portion of the exam. Vertical line indicates the minimum score required to "pass."

Bibliography

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 2010.

Texas Education Agency. Texas public school districts categorized by type, 2007-08, January 2011. URL http://tea.texas.gov/acctres/analyze/0708/district0708.html.

Texas Education Agency. *Directions for District Coordinators, Campus Coordinators, and Test Administrators-TAKS, TAKS (Accomodated), and TAKS-M 2013.*, 2013.

Andreas Ammermueller. Pisa: What makes the difference? In *The Economics of Education and Training*, pages 241–265. Springer, 2008.

Audrey L Amrein and David C Berliner. High-stakes testing & student learning. *Education Policy Analysis Archives*, 10:18, 2002.

Martha J Bailey. More power to the pill: the impact of contraceptive freedom on women's life cycle labor supply. *The Quarterly Journal of Economics*, 121(1):289–320, 2006.

Olesya Baker and Kevin Lang. The effect of high school exit exams on graduation, employment, wages and incarceration. Working Paper 19182, National Bureau of Economic Research, June 2013. URL http://www.nber.org/papers/w19182.

Gary S Becker. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press, 2009.

William E Becker and Sherwin Rosen. The learning effect of assessment and evaluation in high school. *Economics of Education Review*, 11(2):107–118, 1992.

Kelly Bedard. Human capital versus signaling models: university access and high school dropouts. *Journal of Political Economy*, 109(4):749–775, 2001.

No Child Left Behind. No Child Left Behind Act. *Washington, DC: US Department of Education*, 2002.

Roland Bénabou and Jean Tirole. Self-confidence and social interactions. Technical report, National Bureau of Economic Research, 2000.

Julian R Betts and Jeff Grogger. The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22(4):343–352, 2003.

John Bishop. High school exit examinations: When do learning effects generalize? *Yearbook of the National Society for the Study of Education*, 104(2):260–288, 2005.

John H Bishop and Ferran Mane. The impacts of career-technical education on high school labor market success. *Economics of Education Review*, 23(4):381–402, 2004.

Sandra E Black. Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, 114(2):577–599, 1999.

Sandra E Black, Paul J Devereaux, and Kjell G Salvanes. Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *The American Economic Review*, 95 (1):437–449, 2005.

Francine D Blau and Lawrence M Kahn. Do cognitive test scores explain higher US wage inequality? *Review of Economics and Statistics*, 87(1):184–193, 2005.

Alan S Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.

Melinda A Bone. *The Tennessee Education Lottery Scholarships' Impact on Tennessee's Public Higher Education Institutions*. ProQuest, 2008.

Vani K Borooah and Sriya Iyer. The decomposition of inter-group differences in a logit model: Extending the Oaxaca-Blinder approach with an application to school enrolment in India. *Journal of Economic and Social Measurement*, 30(4):279–293, 2005.

John Bound, Michael F. Lovenheim, and Sarah Turner. Why have college completion rates declined? an analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2(3):129–57, 2010. doi: 10.1257/app.2.3.129. URL http://www.aeaweb.org/articles.php?doi=10.1257/app.2.3.129.

Donald J Bruce and Celeste K Carruthers. Jackpot? the impact of lottery scholarships on enrollment in Tennessee. *Journal of Urban Economics*, 81:30–44, 2014.

Stephen V Cameron and James J Heckman. The dynamics of educational attainment for black, hispanic, and white males. *Journal of political Economy*, 109(3):455–499, 2001.

Pedro Carneiro, James J Heckman, and Dimitriy V Masterov. Labor market discrimination and racial differences in premarket factors*. *Journal of Law and Economics*, 48(1):1–39, 2005.

Martin Carnoy and Susanna Loeb. Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4):305–331, 2002.

Center on Education Policy. State high school exit exams: A policy in transition, September 2012. URL http://www.cep-dc.org/displayDocument.cfm?DocumentID=408.

Donghun Cho. The role of high school performance in explaining women's rising college enrollment. *Economics of Education Review*, 26(4):450–462, 2007.

Damon Clark and Paco Martorell. The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):pp. 282–318, 2014. ISSN 00223808. URL http://www.jstor.org/stable/10.1086/675238.

Charles T Clotfelter, Helen F Ladd, Jacob L Vigdor, and Roger Aliaga Diaz. Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, 23(2):251–271, 2004.

Charles T Clotfelter, Helen F Ladd, and Jacob L Vigdor. Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6): 673–682, 2007.

Charles T Clotfelter, Helen F Ladd, and Jacob L Vigdor. The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, 91(2):398–419, 2009.

Charles T Clotfelter, Helen F Ladd, and Jacob L Vigdor. Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3):655–681, 2010.

Timothy G Conley and Christopher R Taber. Inference with difference in differences with a small number of policy changes. *The Review of Economics and Statistics*, 93(1):113–125, 2011.

Christopher Cornwell, Kyung Hee Lee, and David B. Mustard. the effects of merit-based financial aid on course enrollment, withdrawal, and completion in college.. *Journal of Human Resources*, 40(4):895–917, 2005.

Christopher Cornwell, David B Mustard, and Deepa J Sridhar. The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE program. *Journal of Labor Economics*, 24 (4):761–786, 2009.

Christopher Cornwell, David B Mustard, and Jessica Van Parys. Noncognitive skills and the gender disparities in test scores and teacher assessments: evidence from primary school. *Journal of Human Resources*, 48(1):236–264, 2013.

Thomas S Dee and Linda A Jackson. Who loses HOPE? attrition from Georgia's college scholarship program. *Southern Economic Journal*, 66 (2):379–390, 1999.

Thomas S Dee and Brian A Jacob. Do high school exit exams influence educational attainment or labor market performance? Technical report, National Bureau of Economic Research, 2006.

Thomas S Dee, Brian Jacob, Justin McCrary, and Jonah Rockoff. Rules and discretion in the evaluation of students and schoools: The case of the New York Regents examinations. 2011.

David J Deming, Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. School accountability, postsecondary attainment and earnings. Working Paper 19444, National Bureau of Economic Research, September 2013. URL http://www.nber.org/papers/w19444.

John DiNardo, Nicole M Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica: Journal of the Econometric Society*, pages 1001–1044, 1996.

Sharon Dixon Gentry. An assessment of the impact of the Tennessee Education Lottery Scholarship program on access to higher education. (UMI No. 3260217), 2007.

Will Dobbie and Roland G Fryer Jr. Are high-quality schools enough to increase achievement among the poor? evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, pages 158–187, 2011.

Susan Dynarski. Hope for whom? financial aid for the middle class and its impact on college attendance. *National Tax Journal*, 53(3):2–61, 2000.

Susan Dynarski. The behavioral and distributional implications of aid for college. *The American Economic Review*, 92(2):279–285, 2002.

Susan Dynarski. Building the stock of college-educated labor. *Journal of Human Resources*, 43 (3):576–610, 2008.

Robert W Fairlie. An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of economic and social measurement*, 30(4):305–316, 2005.

David N Figlio and Maurice E Lucas. Do high grading standards affect student performance? *Journal of Public Economics*, 88(9):1815–1834, 2004.

Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Decomposition methods in economics. *Handbook of labor economics*, 4:1–102, 2011.

Brigham R Frandsen. Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. *Unpublished manuscript, Brigham Young University*, 2013.

Roland G Fryer and Steven Levitt. An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–40, 2010.

Roland G Fryer and Steven D Levitt. Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2):447–464, 2004.

Roland G Fryer and Paul Torelli. An empirical analysis of acting white. *Journal of Public Economics*, 94(5):380–396, 2010.

Patricia Gandara, Russell Rumberger, Julie Maxwell-Jolly, and Rebecca Callahan. English learners in California schools: Unequal resources,'unequal outcomes. *education policy analysis archives*, 11:36, 2003.

Andrew Gelman and Guido Imbens. Why high-order polynomials should not be used in regression discontinuity designs. Working Paper 20405, National Bureau of Economic Research, August 2014. URL http://www.nber.org/papers/w20405.

Michael Geruso. Black-white disparities in life expectancy: How much can the standard ses variables explain? *Demography*, 49(2):553–574, 2012.

William T Gormley and Ted Gayer. Promoting school readiness in Oklahoma: an evaluation of Tulsa's pre-k program. *Journal of Human Resources*, 40(3):533–558, 2005.

Jeff Groger and Stephen J Trejo. *Falling Behind or Moving Up?: The Intergenerational Progress of Mexican Americans*. Public Policy Institute of California San Francisco, CA, 2002.

Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

Daniel S Hamermesh. Labor demand and the structure of adjustment costs. *The American Economic Review*, 1989.

Daniel S Hamermesh and Stephen J Trejo. The demand for hours of labor: Direct evidence from california. *Review of Economics and Statistics*, 82(1):38–47, 2000.

Karsten T Hansen, James J Heckman, and Kathleen J Mullen. The effect of schooling and ability on achievement test scores. *Journal of econometrics*, 121(1):39–98, 2004.

Eric A Hanushek and Steven G Rivkin. School quality and the black-white achievement gap. Technical report, National Bureau of Economic Research, 2006.

Eric A Hanushek, John F Kain, and Steven G Rivkin. New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383, 2009.

James Heckman and Edward Vytlacil. Identifying the role of cognitive ability in explaining the level of and change in the return to schooling. *Review of Economics and Statistics*, 83(1):1–12, 2001.

James J Heckman and Paul A LaFontaine. The American high school graduation rate: Trends and levels. *The review of economics and statistics*, 92(2):244–262, 2010.

James J Heckman, Lance J Lochner, and Petra E Todd. Earnings functions and rates of return. *Journal of Human Capital*, 2(1):1–31, 2008.

Gary T Henry, Ross Rubenstein, and Daniel T Bugler. Is HOPE enough? impacts of receiving and losing merit-based financial aid. *Educational Policy*, 18(5):686–709, 2004.

Thomas Hungerford and Gary Solon. Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, pages 175–177, 1987.

Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

Brian A Jacob. Getting tough? the impact of high school graduation exams. *Educational evaluation and policy analysis*, 23(2):99–121, 2001.

Robin Tepper Jacob, Pei Zhu, Marie-Andrée Somers, and Howard S Bloom. *A Practical Guide to Regression Discontinuity*. MDRC, 2012.

Ben Jann. A stata implementation of the Blinder-Oaxaca decomposition. *Stata Journal*, 8(4): 453–479, 2008.

Christopher Jepsen, Peter Mueser, and Kenneth Troske. Labor-market returns to the GED using regression discontinuity analysis. 2012.

William R Johnson, Derek Neal, et al. Basic skills and the black-white earnings gap. *The Black-White test score gap*, pages 480–497, 1998.

Chinhui Juhn, Kevin M Murphy, and Brooks Pierce. Wage inequality and the rise in returns to skill. *Journal of political Economy*, pages 410–442, 1993.

Patrick Kline. Oaxaca-blinder as a reweighting estimator. *The American Economic Review*, pages 532–537, 2011.

John M Krieg and Paul Storer. How much do students matter? applying the Oaxaca decomposition to explain determinants of adequate yearly progress. *Contemporary Economic Policy*, 24(4): 563–581, 2006.

Victor Lavy, Avraham Ebenstein, and Sefi Roth. The long run human capital and economic consequences of high-stakes examinations. Working Paper 20647, National Bureau of Economic Research, October 2014. URL http://www.nber.org/papers/w20647.

Edward Lazear. Educational production. *Quarterly Journal of Economics*, 116:777–803, 2001.

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

David S Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

Jason M Lindo, Nicholas J Sanders, and Philip Oreopoulos. Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, 2 (2):95–117, 2010.

Robert L Linn, Eva L Baker, and Damian W Betebenner. Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6):3–16, 2002.

Mark C Long, Dylan Conger, and Patrice Iatarola. Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, 49(2):285–322, 2012.

Ferran Mane. Trends in the payoff to academic and occupation-specific skills: the short and medium run returns to academic and vocational high school courses for non-college-bound students. *Economics of education review*, 18(4):417–437, 1999.

Dave E Marcotte. High school dropout and teen childbearing. *Economics of Education Review*, 34:258–268, 2013.

Francisco Martorell. Do high school graduation exams matter? a regression discontinuity approach. *Job Market Paper*, 2004.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714, 2008.

Richard J. Murnane. U.S. high school graduation rates: Patterns and explanations. *Journal of Economic Literature*, 51(2):370–422, 2013. doi: 10.1257/jel.51.2.

370. URL http://www.aeaweb.org.ezproxy.lib.utexas.edu/articles.php?doi=10.1257/jel.51.2.370.

Richard J Murnane, John B Willett, and Frank Levy. The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, 77(2):251–66, 1995.

Georgia Student Finance Commission (n.d.). Georgia's HOPE Program. Online, 6 August, 2012. http://www.gsfc.org/gsfcnew/HopeProgramm.CFM?sec=3.

Derek Neal. Why has black–white skill convergence stopped? *Handbook of the Economics of Education*, 1:511–576, 2006.

Derek A Neal and William R Johnson. The role of premarket factors in black-white wage differences. *The Journal of Political Economy*, 104(5):869–895, 1996.

Erik C Ness and Richard Tucker. Eligibility effects on college access: Under-represented student perceptions of Tennessee's merit aid program. *Research in Higher Education*, 49(7):569–588, 2008.

Austin Nichols. rd 2.0: Revised Stata module for regression discontinuity estimation, 2011. URL http://ideas.repec.org/c/boc/bocode/s456888.html.

Sharon L Nichols, Gene V Glass, and David C Berliner. High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14:1, 2006.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709, 1973.

Dongshu Ou. To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam. *Economics of Education Review*, 29(2):171–186, 2010.

Lindsay C Page, Richard J Murnane, and John B Willett. Trends in the black-white achievement gap: Clarifying the meaning of within-and between-school achievement gaps. Technical report, National Bureau of Economic Research, 2008.

Amanda Pallais. Taking a chance on college is the Tennessee Education Lottery Scholarship program a winner? *Journal of Human Resources*, 44(1):199–222, 2009.

John P Papay, Richard J Murnane, and John B Willett. The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, 32(1):5–23, 2010.

John P Papay, Richard J Murnane, and John B Willett. How performance information affects human-capital investment decisions: the impact of test-score labels on educational outcomes. Technical report, National Bureau of Economic Research, 2011a.

John P Papay, John B Willett, and Richard J Murnane. High-school exit examinations and the schooling decisions of teenagers: A multi-dimensional regression-discontinuity analysis. Technical report, National Bureau of Economic Research, 2011b.

John P Papay, Richard J Murnane, and John B Willett. High-school exit examinations and the schooling decisions of teenagers: Evidence from regression-discontinuity approaches. *Journal of Research on Educational Effectiveness*, 7(1):1–27, 2014.

David Penn and Reuben Kyle. The Tennessee Education Lottery Scholarship: A reward for past achievement or motivator for future performance? *Middle Tennessee State University*, 2007.

Chester Polson. Taks-ing students? texas exit exam effects on human capital formation. *Working Paper*, 2015.

Sean F Reardon and Claudia Galindo. The hispanic-white achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, 46(3):853–891, 2009.

Sean F Reardon and Joseph P Robinson. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5(1):83–104, 2012.

Sean F Reardon, Allison Atteberry, Nicole Arshan, and M Kurlaender. *Effects of the California high school exit exam on student persistence, achievement, and graduation*. S. Reardon, 2009.

Sean F Reardon, Nicole Arshan, Allison Atteberry, and Michal Kurlaender. Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation. *Educational Evaluation and Policy Analysis*, 32(4):498–520, 2010.

Douglas S Reed. Is there an expectations gap? educational federalism and the demographic distribution of proficiency cut scores. *American Educational Research Journal*, 46(3):718–742, 2009.

Steven G Rivkin, Eric A Hanushek, and John F Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goekin, Matthew B. Schroeder, and Matthew Sobek. (2010). integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. *Minneapolis: University of Minnesota*, 2010.

Guido Schwerdt and Martin Raymond West. The effects of early grade retention on student outcomes over time: Regression discontinuity evidence from florida. 2012.

Judith Scott-Clayton. On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3):614–646, 2011.

Mathias Sinning, Markus Hahn, Thomas K Bauer, et al. The Blinder-Oaxaca decomposition for nonlinear regression models. *Stata Journal*, 8(4):480–492, 2008.

David L Sjoquist and John V Winters. Building the stock of college-educated labor revisited. *Journal of Human Resources*, 47(1):270–285, 2012.

David L Sjoquist and John V Winters. State merit-based financial aid programs and college attainment. *Journal of Regional Science*, 2014.

James P Smith. Assimilation across the Latino generations. *American Economic Review*, pages 315–319, 2003.

Gary Solon, Steven J. Haider, and Jeffrey Wooldridge. What are we weighting for? Working Paper 18859, National Bureau of Economic Research, February 2013. URL http://www.nber.org/papers/w18859.

Tennessee Higher Education Commission. Tennessee Education Lottery Scholarship Program Annual Report (Outcomes Through Fall 2010). 2011.

Texas Education Agency. *Texas Assessment of Knowledge and Skills Information Booklet: Mathematics Exit Level*, 2007.

Texas Education Agency. Academic excellence indicator system, 2012. URL http://ritter.tea.state.tx.us/perfreport/aeis/.

Texas Education Agency. Interpreting assessment reports, 2014a. URL http://www.tea.state.tx.us/student.assessment/interpguide/.

Texas Education Agency. TAKS Test Administration Manuals, Feb 2014b. URL http://www.tea.state.tx.us/student.assessment/taks/manuals/.

Texas Education Agency. STAAR Resources, Sept 2014c. URL http://www.tea.state.tx.us/student.assessment/staar/.

Petra E Todd and Kenneth I Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, 1(1):91–136, 2007.

Sergio Urzua. Racial labor market gaps the role of abilities and schooling choices. *Journal of Human Resources*, 43(4):919–971, 2008.

US News and World Report. National University Rankings. Online, 27 Nov. 2011. http://colleges.usnews.rankingsandreviews.com/best-colleges/rankings/national-universities/spp

Alan Vanneman, Linda Hamilton, Janet Baldwin Anderson, and Taslima Rahman. Achievement gaps: How black and white students in public schools perform in mathematics and reading on the national assessment of educational progress. statistical analysis report. nces 2009-455. *National Center for Education Statistics*, 2009.

Jedrzej Yendrick Zieleniak. Jumping through hoops: Student achievement and the California high school exit exam. 2013.

Vita

Chester William Polson II was born in Austin, Texas, on July 16th, 1987, to Lee Riddell Polson and Melissa Curtin Polson. With an exception of a five-year stint in the Washington, D.C. metro area, he was raised in Austin. He attended Vanderbilt University where he earned a Bachelor of Arts in Economics and English in 2009. Upon graduating from Vanderbilt, he spent 13 months in the greater metropolitan area of Quito, Ecuador, engaged in holistic community development with Manna Project International, a 501(c)(3) non-profit. During his time there Chester taught adult English classes, worked with microfinance and small business classes, taught a children's art class, and staffed a community center and library. He entered graduate study at the University of Texas at Austin Department of Economics in August, 2010. Upon graduating, Chester plans on being very successful.

Permanent address: 5900 Cary Dr.
Austin, Texas 78757

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.