UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in Economics of Education and Public Economics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Sieuwerd Jelle Gaastra

Committee in charge:

Professor Roger Gordon, Chair
Professor Julian Betts
Professor Julie Cullen
Professor Ruixue Jia
Professor Krislert Samphantharak

2017

The dissertation of Sieuwerd Jelle Gaastra is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____

Chair

University of California, San Diego

2017

DEDICATION

To my parents, Corrie and Tinus.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Professor Roger Gordon, the chair of my committee, for his invaluable advice and constant support.

I am also very grateful for the help and advice of the other members of my thesis committee: Prof. Julian Betts, Prof. Julie Cullen, Prof. Ruixue Jia and Prof. Krislert Samphantharak. I also benefitted from numerous conversations with Prof. Gordon Dahl and from the valuable comments and suggestions of many seminar participants at the Department of Economics of UC San Diego.

I would also like to thank Andrew Zau for his very extensive help with collecting, cleaning and using data from the San Diego Unified School District. The assistance of Jeffrey Clemons, Christiane Trout-McPhee, Ron Rode and Dr. Karen Volz Bachofer was also much appreciated.

In addition, I would like to thank my peers Vinayak Alladi, Kilian Heilmann, Claudio Labanca, Pablo Ruiz-Junco, Xiaxin Wang and Wei You for many helpful conversations and great company over the years.

Finally, I would like to thank my parents, Corrie and Tinus, my siblings, Ida, Jurjen, Sibrich and Amarins, and my roommates, Amol, Min and Viraj, for their constant support throughout writing this dissertation.

Chapter 1 is currently being prepared for submission for publication of the material. The dissertation author was the sole author of this paper. Gaastra, Sieuwerd.

Chapter 2, in part, has been submitted for publication of the material. The dissertation author was the sole author of this paper. Gaastra, Sieuwerd.

Chapter 3 is currently being prepared for submission for publication of the material. The dissertation author was the sole author of this paper. Gaastra, Sieuwerd.

VITA

2011	B.A. in Mathematics and Economics, Middlebury College
2012	M.A. in Economics, University of California, San Diego
2017	Ph.D. in Economics, University of California, San Diego

PUBLICATIONS

Colander, C., S. Gaastra, and C. Rothschild, "The Welfare Costs of Market Restrictions," *Southern Economic Journal*, 2010, *77* (1), 213-223.

ABSTRACT OF THE DISSERTATION

Essays in Economics of Education and Public Economics

by

Sieuwerd Jelle Gaastra

Doctor of Philosophy in Economics

University of California, San Diego, 2017

Professor Roger Gordon, Chair

This dissertation explores three topics in the economics of education and public economics. In chapter 1, I use a fuzzy regression discontinuity design to estimate how attending an after-school program in San Diego impacts short-run student academic and behavioral outcomes. I find that the after-school program has no impact on academic outcomes in math and English Language Arts for students in K-8. On the other hand, I find suggestive evidence that student behavior is affected, but the direction of the effect varies across years. In chapter 2, I evaluate the potential for changes in expected lifetime income by major that result from changes in the individual income tax law to affect college major choices, using one of the largest federal income tax reforms in recent U.S. history, the 1986 Tax Reform Act (TRA86), as

a case study. Due to the limited differential impact on expected earnings across majors, I find that TRA86 is likely to have had a small impact on the composition of completed college majors after 1986. Lastly, in chapter 3, I study the peer effects of students who speak a language other than English at home (ESL) on native English speaking (ELO) students in grades 2 to 6 in California. I find that higher concentrations of ESL students have no effect on average math test scores and a negative effect on average English test scores of ELO students.

Chapter 1

The Effect of After-School Programs on Short-Run Student Academic and Behavioral Outcomes

Abstract

After-school programs are programs for school-age children that operate during after-school hours, typically from 3 p.m. to 6 p.m., and that offer children the chance to participate in a variety of academic and non-academic extra-curricular activities under the supervision of adults. In 2013-2014, 18 percent of all school-age children in the U.S. participated in an after-school program during an average school week. In this paper, I estimate how attending one such after-school program in San Diego that is typical of many after-school programs offered nationwide impacts short-run student academic and behavioral outcomes. To overcome the usual selection issues associated with students choosing to apply for an after-school program, I use a fuzzy RD design that exploits the structure of the admissions procedure to the after-school program. I find that attending an after-school program for 80 additional days a year (the average change in days of attendance at the cutoff) has no impact

on report card grades and test scores in math and English Language Arts for students in K-8. On the other hand, I find suggestive evidence that student behavior, as measured by various report card grades related to students' in-class behavior, is affected. However, the direction of the effect depends on the particular outcome analyzed. Due to the nature of the RD design, these results apply only to students near the RD threshold. Importantly, these students appear to be relatively more advantaged when compared to the typical after-school program participant.

1.1 Introduction

After-school programs are programs for school-age children that operate during after-school hours, typically from 3 p.m. to 6 p.m., and that offer children the chance to participate in a variety of academic and non-academic extra-curricular activities under the supervision of adults. As a result of increases in public funding and demand, participation in after-school programs has increased dramatically over the last two decades. In 2013-2014, 18 percent of all school-age children in the U.S. participated in an after-school program during an average school week (Afterschool Alliance, 2014), an increase of 7 percentage points since 2004.

The increase in public funding and demand for after-school programs can largely be attributed to three factors (Kugler (2001), Kane (2004), Lauer et al. (2006)). First, an increase in maternal employment has resulted in more children spending time alone at home after school. Second, public spending for after-school programs has increased as a result of the increased accountability in education that followed the passage of the No Child Left Behind Act in 2001. This increased funding reflects the belief that economically disadvantaged children can improve their academic achievement if given more time. Lastly, in light of research showing that juvenile crime peaks during after-school hours (Sickmund et al. (1997); Office of Juvenile Justice and Delinquency Prevention (2014)), there has been an increasing concern among policy makers that unsupervised time after school can lead to negative behaviors among children such as crime and drug abuse. After-school programs address both the need for supervised after-school care and the need for improved academic achievement. They provide a safe environment to school-age children during the time that their parents are still at work and aim to improve these children's academic outcomes by offering various academic activities.

However, despite the popularity of after-school programs and the large amounts of public funding going towards them, there is little rigorous evidence on their causal impact on academic and behavioral outcomes of children. Existing research on the impact of various forms of child care has mostly focused on children up to age 5 and looked at universal child care programs (e.g. Baker et al. (2008), Havnes and Mogstad (2011)) or enriched center-based child care programs such as Head Start (e.g. Barnett (1995), Karoly et al. (1998), Garces et al. (2002)).[1] For school-age children, the literature on the impact of various forms of child care is much smaller. Aizer (2004) evaluates the effect of being supervised by an adult after school, either at home or at any other location such as a child care center, on the behavior of children aged 10-14. She finds that adult supervision leads to less negative behavior such as skipping school or using marijuana or alcohol. The effect of this unspecified adult supervision after school might not be fully comparable to that of after-school programs though, as the counterfactual to after-school programs can include supervision by an adult at home or at a child care center that does not offer enrichment activities.

There are a number of studies that have looked directly at the impact of after-school programs. However, it is hard to draw conclusions about the causal impact of after-school programs based on these studies. First, most of these studies lack exogenous variation in after-school program participation. Because students choose to participate in after-school programs, this means that these studies have trouble separating the impact of factors that are correlated with the application decision from the impact of attending the

[1] Enriched center-based child care programs are generally targeted at a specific population and provide enrichment activities. Universal child care programs on the other hand generally serve everybody and might not provide enrichment activities.

after-school program itself.[2] Second, the studies that are based on randomized control trials are characterized by low levels of participation in the after-school program. In these studies, students in the treatment group on average only attended the after-school program for around 40 out of the roughly 180 school days per year. As a result, it is hard to say whether the lack of significant findings in these studies is due to the after-school program not affecting student outcomes or due to the limited exposure of students to the program.[3]

This paper exploits the admissions procedure to a publicly-funded after-school program in San Diego to investigate the impact of attending this program on the short-run academic and behavioral outcomes of students in K-8. Applicants are admitted based on, among other things, the postmark date of their application in relation to the postmark dates of all other submitted applications. This results in a cutoff application postmark date such that applicants who applied before or on this date are admitted whilst those who applied after this date are wait listed. Importantly, the cutoff date is unknown to the applicants when they apply, allowing me to use a fuzzy regression discontinuity design in which I compare the outcomes of applicants on either side of the cutoff to analyze the causal impact of the program on student outcomes. Students who applied at or just before the cutoff date attend the after-school program on average for close to 80 days more per year than students who applied just

[2]For instance, Le et al. (2011) use a matching design to evaluate an after-school program in a middle school in Oakland that serves mainly immigrant and minority youth. In particular, the students that chose to apply to the after-school program were matched by gender, age and ethnicity to a control group of students who attended the same middle school. In addition, Huang and Wang (2012) use propensity score matching to perform a statewide evaluation of the same after-school program that is studied in this paper. They use 5 school-level characteristics and 11 student-level characteristics to create a control group of non-participating students. Both of these studies rely on the strong assumption that selection into treatment is as good as random conditional on observables.

[3]For instance, James-Burdumy et al. (2005, 2007 and 2008) evaluate the impact of attending a similar after-school program as studied in this paper for students in elementary school using a two-year long randomized control trial. On average, students in the treatment group only attended the after-school program for a total of 81 days during the two years. Similarly, Gottfredson et al. (2010a, 2010b) evaluate an after-school program for middle school students in Baltimore using a one-year randomized control trial. On average, students in the treatment group only attended the after-school program for 35.6 days during the entire school year. Low and irregular attendance of the after-school program could have an effect on the program‘s ability to affect student outcomes.

after the cutoff date. This represents a doubling of the exogenously induced exposure to an after-school program compared to previous studies that are based on RCTs. Due to the nature of the RD design, the causal impact of the after-school program estimated in this paper applies only to students near the RD threshold. These students appear to be relatively more advantaged when compared to the typical after-school program participant. The effect estimated in this paper is therefore likely to differ from the impact of the after-school program on the average participant.

This study adds to the literature mentioned above in various ways. First, it addresses the two main problems of the existing literature on after-school programs, namely limited exposure to the after-school program and a lack of exogenous variation, by using plausibly exogenous variation in after-school program participation in a case where this exogenous variation results in high levels of exposure to the after-school program. In addition, this study also adds to the scarce literature on the impacts of various forms of child care for school-age children. Lastly, this study looks at an after-school program that is implemented on a large scale. This makes it potentially more relevant for policy purposes than the RCTs on after-school programs mentioned before, as these were based on after-school programs implemented on a very small scale or only evaluated after-school programs at a non-random subset of schools[4]

After-school programs can affect students academic and behavioral outcomes in many ways. The academic activities that are part of after-school programs allow students to spend more time on mathematics and English Language Arts (ELA) each day under the supervision of adults and so improve their outcomes in these subjects. At the same time, by allowing parents

[4]In particular, Gottfredson et al. (2010a, 2010b) look at an experimental after-school program that operated at only 5 middle schools. James-Burdumy et al. (2005, 2007 and 2008) look at a large-scale after-school program, the 21^{st} Century Community Learning Centers, but are only able to look at schools that are able to implement random assignment. These after-school programs made less than 2% of all 21^{st} Century Community Learning Centers (see Dynarski et al. (2003)). In contrast, in this study, on average I use data from close to 15% of all after-school programs operated under the supervision of the San Diego Unified School District in a given year.

increased flexibility with work schedules, after-school programs could allow parents to work more. Following research linking family income and children's achievement (Dahl and Lochner (2012)), this could lead to better academic outcomes. In addition, attending after-school programs will keep children off the streets during the hours that juvenile crime peaks and expose them to positive role models in the form of adult staff members. This environment could lead students to improve their behavior during the regular-school day as well.

On the other hand, the additional time that students spend engaged in academic activities at the after-school program each day could lead to decreased effort and increased boredom and fatigue during the regular school day, potentially negatively impacting students' academic performance and behavior. Moreover, the student-to-staff ratios of 20 or 15 to 1 in after-school programs could result in each student receiving little individual attention. Similarly, the often limited communication between regular school day staff and after-school program staff regarding individual student needs (Huang and Wang (2012)) could mean students do not receive attention in areas that they need the most during the after-school program. Both these factors could limit the benefits that students derive from attending an after-school program.

In addition, students are likely to be exposed to different types of peers in the after-school program than if not attending this program. In what direction this will affect student outcomes depends on the peers students would have when not participating in the program. On average, students near the RD threshold appear to be relatively advantaged when compared to typical after-school program participants. For students near the RD threshold, participating in an after-school program could thus mean being exposed to more low-performing and possibly disruptive peers than outside an after-school program. Following extensive research on the peer effects of these types of students[5], we might thus expect peer effects to limit the benefits that the typical student at the threshold derives from the program.

[5] e.g Gaviria and Raphael (2001), Figlio (2007), Carrell and Hoekstra (2010), Lavy et al. (2012a) and Lavy et al. (2012b).

Peers in the after-school program might also impact a student's behavior and academic outcomes by affecting that student's social network outside of the the after-school program. New friendships could be formed during the after-school program that could potentially replace or change the student's friendships with students not in the program. Recent research[6] highlights the importance of the types and number of friends a student has for that student's educational and behavioral outcomes. It is unclear in what direction a student's outcomes will be impacted by a change in his social network due to PrimeTime participation. It is hard to exactly anticipate how this social network will change due to program participation.

In this paper I estimate the net impact of attending an after-school program on student outcomes. This estimated effect will depend on how and with whom students would spend the hours after school if not attending an after-school program. Unfortunately, the data I will be using in this paper do not contain any information on how students who do not attend an after-school program spend their time after school. Since the alternative to after-school programs likely depends on family background, I will therefore also investigate how the impact of attending an after-school program varies along this dimension. In addition, to further understand who is impacted by after-school program attendance, I also estimate treatment effects separately by gender and by students' baseline performance.

I find that students' academic outcomes are on average not affected by attending the after-school program. This result holds whether I look at standardized test scores or school grades in ELA and math. The fact that academic outcomes appear to be unaffected on average does not seem to be a product of the particular grades and years for which data are available, as the various academic outcomes are based on students across all different years and grades for which I have data. Although on average there seems to be no impact, looking at subgroups of students I find suggestive evidence that the after-school program causes academic outcomes to improve for girls and worsen for boys and for students from a single-parent household.

[6] e.g. Patacchini et al. (2011), Fletcher and Ross (2012) and Lavy and Sand (2014).

On the other hand, students' behavioral outcomes, as measured by several report card grades regarding student behavior in K-5, are significantly affected. In particular, attending the after-school program for 80 more days per year (the average change in days of attendance at the cutoff) causes students' assignment completion rate and preparedness for class to decrease by respectively 0.19 and 0.14 standard deviations, and increases students' feeling of responsibility for learning by 0.35 standard deviations. These results are robust to a variety of specifications, the only notable exception being the sensitivity of the negative effects to using a donut RD, and are not specific to a particular subgroup of students. The variable measuring students' feeling of responsibility for learning and the variables measuring students' assignment completion rate and preparedness for class are not available in the same years. The difference in the direction of the effect of the after-school program on these variables could therefore be due to a variety of reasons. I provide suggestive evidence that it is attributable to either changes in the unobservable characteristics of students in my sample over time or changes in the operation of the after-school program at the schools in my sample that were not the result of explicit policy changes. I do so by among other things showing that the coefficients on other behavioral variables that are available in all years change signs across years in ways that are consistent with the signs of the significantly impacted behavioral variables that are available for only a subset of years.

The lack of an impact on academic outcomes and the absence of consistent results on behavioral outcomes is striking, especially given the high cost of the after-school program of up to $1350 per student per year. However, as mentioned before, this study only informs us about the causal impact of the after-school program for the relatively advantaged students at the cutoff. The impact of the after-school program on the typical more disadvantaged participant not near the cutoff might be very different. In addition, due to data limitations, there are many potential effects of after-school program attendance that this paper does not capture. Most importantly, I am unable to measure the effects on students' participation in criminal activity in the short- and long-run and the effect on parental labor supply decisions. More research is needed to obtain a more complete understanding of all the impacts of after-school programs on students and their parents.

This paper proceeds as follows. Section 2 discusses the history, structure and admissions procedure of the after-school program studied in this paper. Section 3 discusses the empirical methodology. Section 4 discusses the data and the implementation of this empirical methodology. Section 5 tests various threats to the identification strategy, presents the results and also tests whether these results are sensitive to a variety of robustness checks. Section 6 concludes with policy implications and briefly discusses the limitations of this study as well as possible directions that future research on after-school programs could take.

1.2 Institutional Background

1.2.1 History

Public funding for after-school programs greatly increased in the early 2000s. For instance, annual federal funding for 21st Century Community Learning Centers, which are after-school programs in low-performing schools and schools in high-poverty areas, grew from $40 million in 1998 to around $1 billion in 2002. States also allocated additional funds to after-school programs. In California, the passing of proposition 49 in 2002 permanently earmarked general state funds for before- and after-school programs operated under the so-called After School Education and Safety (ASES) Program. This increased annual state spending on these programs by around $450 million dollars. Today, these ASES programs are still in place and serve over 400.000 elementary and middle school students daily in 4000 schools around California. Per student, per day funding is currently $5.00 for the ASES program and $7.50 for the ASES after-school program, resulting in an annual cost of around $550 million to the state (California Department of Education (2014)).

In San Diego Unified School District (SDUSD), ASES before- and after-school programs are in place in over 100 schools. In 2014-2015, SDUSD received over \$20 million dollars in state funding to serve close to 15.000 K-8 students, 16% of all such students in that year, each day in after-school programs at 127 different schools. In this paper, I will investigate how attending these SDUSD-based after-school programs affected short-run student outcomes. Unfortunately, I am not able to investigate the impact of before-school programs as attendance data is not available for these programs.[7] All ASES after-school programs in California are by law obliged to meet the same requirements in terms of days and hours of operation, activities offered, student-to-staff ratios, student attendance and staff qualifications as those described for the SDUSD-based ASES after-school programs below. It is therefore likely that the results presented in this paper are generalizable to ASES after-school programs in other school districts in California.

1.2.2 The PrimeTime Extended Day Program

The ASES after-school program in SDUSD is called the PrimeTime Extended Day Program and is offered at most elementary and middle schools. The program aims to provide an educationally enriching and safe environment to children during non-school hours when their parents are still at work. The program is operated at school sites by community-based organizations, and SDUSD provides program management assistance and administrative oversight to these community-based organizations.[8]

School-level eligibility for funding to operate a PrimeTime (PT) program is determined by the state and is primarily a function of the percentage of students eligible for free and reduced price meals. An eligible school can

[7] The admissions procedure for the before-school program is the same as for the after-school program and students can choose to apply to the before-school program, the after-school program or both using one application. The before-school program only runs for around 90 minutes each day, but has the same overall structure as the after-school program, with time being divided between academic activities, physical activity and enrichment activities. In the robustness checks, I drop students who applied to both the before-school program and after-school program to make sure that my results are not biased by students attending before-school programs.

[8] In a few cases, SDUSD provides the program at a school site instead of a community-based organization. The schools at which this occurs are not in my sample.

decide on the number of students it wants to receive funding for, the so-called average daily attendance (ADA), but due to limited state funds a school often receives funding for fewer spots than desired.[9] Importantly, schools need to apply for PrimeTime funding well in advance of the start of the school year and funds are allocated to schools before the start of the student application period. As a result, the ADA of a given school for an upcoming school year cannot be changed in response to the number or composition of applications received during the application period. Once funding has been granted, schools can face funding reductions if the actual daily attendance of the after-school program systematically falls short of the ADA that the school receives funding for. This incentivizes after-school programs to keep up their attendance throughout the school year, for instance by enrolling additional students off the wait list as needed.

PrimeTime programs operate every regular school day. They last from the moment the regular school day ends, typically around 3PM, until 6PM and operate for a minimum of 15 hours per week. Each day, students spend 60 to 90 minutes on various academic activities, 30 minutes on some form of physical activity and the remaining time on various structured enrichment activities such as playing musical instruments, writing and performing plays, participating in athletic leagues and playing cooperative games. Students participate in the various activities in groups that are based on their grade level. Student to staff ratios are 15 to 1 in elementary school and 20 to 1 in middle school.

The 60 to 90 minutes of academic activities that students participate in each day are focused on mathematics, English Language Arts, social science and science. Depending on the day, students either receive homework assistance or classroom instruction in these subjects, or participate in various other academically oriented activities such as hands-on-math games, spelling contests and educational board games. Classroom instruction is provided by credentialed teachers. The school principal plans the classroom instruction at

[9]In addition, schools can also receive funding for fewer spots than desired because there exists a maximum ADA that a school of a given size can apply for.

the beginning of the school year. He selects the particular students that are to receive classroom instruction, identifies the appropriate subject material and curriculum area to be covered based on student needs, and schedules the classroom instruction days throughout the year. When selecting the students that are to receive classroom instruction, priority is given to students performing below grade level and/or qualifying for academic assistance[10], although other students may also be offered classroom instruction if resources are available. All PrimeTime programs are required to spend 8% of their budget on classroom instruction.[11] This allows them to offer around 1.5 hours of classroom instruction by a credentialed teacher per school week per 20 enrolled students.

Homework assistance and the various other academically oriented activities mentioned above are provided by PrimeTime staff members. These staff members are generally not certified teachers, but need to meet SDUSD's criteria for being an instructional aide.[12] PrimeTime staff have access to SDUSD curriculum maps that outline the subject material that is covered in class during the regular school day in a given subject and grade during a particular time of the year. Staff members can use these maps to select academic activities that are appropriate for students throughout the school year and so ensure that the academic activities align with students regular day academic goals.

As mentioned earlier, besides providing academic support, PrimeTime also aims to provide a safe environment to children during non-school hours. To achieve this aim, several so-called respect rules and a student discipline policy are in place to encourage students to take responsibility for their actions and respect others and themselves. Student who violate the student discipline

[10]Unfortunately, I cannot identify these students in the data that I have. Later in this paper, I do however look at heterogeneous treatment effects based on whether the lag of the outcome variable is above or below the in-sample median.

[11]A very small number of schools are allowed to spend less than 8% of their budget on classroom instruction as they have a particular high need for other services that are part of the PrimeTime program.

[12]This means having either 48 units of college credits or passing the district's Classroom Assistant Proficiency Exam. Prior to starting their work at PrimeTime, they also receive a limited amount of training on health and safety standards, dealing with special education students, child abuse reporting, positive behavior management, and on conducting academic and enrichment activities. Unfortunately, there are no exact standards regarding how extensive these trainings should be.

policy can be subject to disenrollment. Staff members also attempt to encourage positive student behavior by using various positive behavior management techniques such as positive reinforcement.

The PrimeTime program has strict attendance requirements. Elementary school pupils are expected to attend every day for the full range of program hours and middle schools pupils are expected to attend at least 3 days a week[13]. Failure to follow this attendance policy can result in disenrollment. As will be illustrated later, this attendance policy results in high rates of attendance of the program.

Students who do not attend a PrimeTime program can potentially spend the hours after school in a variety of ways. For instance, they could be looked after at home by a family member such as a parent or a sibling, or have to look after themselves at home if such family members are not available. They could also be attending a private child care center[14] or be looked after either at home or elsewhere by an unrelated adult such as a neighbor or a babysitter. In addition, they could be participating in individual after-school activities such as sports, music lessons and hobby clubs.[15] Lastly, students could also be participating in a fee-based after-school program. These programs are similar in set-up to the PrimeTime program, but cost around $3000 per year for a student attending five days a week. In contrast to PrimeTime, fee-based programs do not offer academic instruction by credentialed teachers, although they do also offer homework support. In addition, fee-based programs do not have a strict attendance policy. Students can decide how many days a week to attend and for what hours to attend on a particular day. As compared to Primetime, fee-based programs thus require less of a time commitment and offer fewer academic enrichment opportunities.

[13]Conditional on attending at least 50% of all program hours, elementary and middle school students are permitted to leave the PrimeTime program early on some days because of a parallel program, family obligations, medical appointments or transportation needs.

[14]Students are supervised by adults in these programs, but generally do not receive the enrichment opportunities available to them in the PrimeTime program. Depending on the size of the child care center, these centers are located in the homes of the people running the center or at a separate child care facility.

[15]These activities are offered by some schools and a variety of other organizations, but generally run for a much shorter time period than the PrimeTime program.

Fee-based after-school programs are available on some school campuses and at a limited number of off-campus locations belonging to community-based organizations.[16] There are no data on the availability of off-campus fee-based after-school programs to students at various schools, but there are data on the schools at which on-campus fee-based after-school programs are offered in addition to the free PrimeTime program. My results might be weakened if many students in the control group attend a fee-based program, as opposed to making use of the other alternatives to PrimeTime mentioned above, and if these programs are very similar to PrimeTime. Although this is unlikely, as fee-based programs differ in some important ways from the PrimeTime program, one way to ease this worry is to only look at schools that do not offer on-campus fee-based after-school programs. Although some students at these schools could still attend off-campus fee-based programs, on average they should be less likely to attend a fee-based program than students from schools with an on-campus fee based program. Hence, to lessen the worry that my results are weakened by students in the control group attending fee-based programs that could potentially be similar to PrimeTime, in a robustness check later in this paper I estimate the main models using only schools that do not offer an on-campus fee-based program.

SDUSD does not collect data on the ways in which students who do not attend PrimeTime spend their time after school. This means it is not possible to know the exact counter-factual to PrimeTime participation for students in my study. Nevertheless, whether some of the above options are available to a particular student is likely to depend heavily on his family background. I will therefore also check whether the impact of PrimeTime differs depending on whether a student is from a single-parent household and whether his parent(s) have a college degree.

[16]Not all schools offer a fee-based after-school program in addition to the PrimeTime program. Whether or not a fee-based program is offered at a particular school is mostly a function of the demand for such a program at that school. Off-campus fee-based after-school programs generally provide transportation to and from the after-school program to students attending schools in the area surrounding the after-school program.

1.2.3 Admissions Procedure

All PrimeTime programs in SDUSD follow the same admissions procedure in any given school year. During the years considered in this study this admissions procedure remained largely unchanged. On the first or second Monday of March in each year, all K-8 students who are enrolled in a school offering an after-school PrimeTime program receive a paper application for the PrimeTime program during the upcoming school year.[17] [18] If a student is unsure which school he will attend in the following school year, he is advised to submit a separate application to each school that he is considering that offers a PrimeTime program.

From the first day that applications are made available students have 61 days to submit their application if they want to be considered in the initial round of admissions.[19] Students who apply after this so-called initial application timeline will only be considered after this initial round of admissions has been conducted. Applications can only be submitted via U.S. mail as the postmark date of the application plays an important role in the admissions procedure.[20] Since applications cannot be postmarked on Sundays, this means that there are a total of 53 days during the initial application timeline on which applications can be postmarked.[21] In the analysis later in this paper, differences in postmark dates between applications will refer to differences in the number of days in this 53 day window. That is, for the purpose of the data analysis applications postmarked on a Saturday and on the first Monday following this Saturday are considered to have a postmark date that is 1 day

[17] As all middle schools offer PrimeTime programs, applications are also distributed to students who currently attend an elementary school that does not offer a PrimeTime program but who will be attending a middle school in the following school year.

[18] To ensure all students have an equal opportunity to apply, all parents in the school district receive an email when this application is distributed to the students and paper applications are also made available on the school district‘s website in English, Spanish, Tagalog and Vietnamese.

[19] As mentioned before, applications are generally made available on the first or second Monday in March. Hence, 60 days means that students can submit their application during the nine school weeks following the day applications are made available.

[20] Applications that are faxed, e-mailed or hand delivered are not accepted.

[21] This includes 6 possible days in each of the first 8 weeks and 5 days in the last week of the initial application timeline.

apart.

All applications received by a PrimeTime program at a given school during the initial application timeline are ranked based on two criteria. First, each application is assigned a priority score ranging from 0 to 4, with a higher priority score being preferred. Each applicant's priority score is based on whether or not he meets 4 different criteria. Each criterion is worth 1 point. A student receives a point if he qualifies for academic assistance[22], comes from a single-parent household[23], has parents that are both either full-time employed or full-time students[24], and if he was on the wait list for or participated in a PrimeTime program in the previous school year.[25] [26] Once all applications have been ranked based on their priority scores, they are further ranked by their postmark date to distinguish between applications with the same priority score. In this second step, applications with an earlier postmark date are preferred to those with a later postmark date.

Having ranked all applications based on their priority score and postmark date, in late June a PrimeTime program offers admission to individuals following the ordering of the ranking and up to the capacity of the program.[27] This capacity is partially determined by the program itself, as a program has some flexibility in the number of students it admits during the initial timeline.[28] This raises the worry that this capacity could be set endogenously by a

[22]This is determined by the school district and is based on the Standards Based Report Card at the elementary level and course grades at the middle level.

[23]Students from a two-parent household in which one of the parents is on military deployment also qualify.

[24]Parents need to provide their employer's or school's phone number on the PrimeTime application such that the employment or school enrollment can be verified.

[25]To ensure parents truthfully report the information requested on the applications, 5% of all received applications are randomly selected and verified. Falsifying information on the application can disqualify a student from participating in a PrimeTime program.

[26]From 2010-2011 to 2012-2013, the priority score ranged from 0 to 5. In 2010-2011 and 2011-2012, students received 2 points instead of 1 if they participated in a PrimeTime program in the previous year. In 2012-2013 students received an additional point if they qualified for free or reduced-price lunch.

[27]That is, a program with a capacity of 75 students will admit every student ranked 75th or above, and waitlist all students with a rank of 76 or lower.

[28]The number of students a PrimeTime program admits during the initial application timeline is not determined by the school district. Rather, it is left to the discretion of the individual PrimeTime program providers. On the one hand, programs need to admit a sufficient number of students to meet the daily attendance, the so-called ADA, that they

program and depend on the type of students at the margin of being admitted. I explore this possibility in detail in section 1.5.1 and find no evidence for any such endogenous capacity selection.

Students who are not admitted to a PrimeTime program during this initial round of admissions are placed on a waiting list that preserves the application ranking order. These waitlisted students might still be offered admission to the PrimeTime program during the upcoming school year if space opens up at this time as a result of enrolled students dropping out of the program. For an initially waitlisted student to be offered admission at this later time though, he needs to still be at the top of the waiting list at the time that the other students drop out. This might not be the case as applicants who apply after the initial application timeline are placed on the same waiting list as the applicants who applied during the initial application timeline. These later applicants might be higher ranked on the wait list if their priority score exceeds that of the individuals who applied during the initial timeline. That is, after the initial round of admissions, individuals‘ ranking on the wait list can fall if individuals with a higher priority score (but later postmark date) apply. As a result, many of the students at the top of the waiting list at the end of the initial round of admissions will not be offered admission to the PrimeTime program during the next school year.

Students who are initially waitlisted can also be admitted to the PrimeTime program as a result of a request by the principal. Each year, up to 10% of the slots of a particular PrimeTime program are left open during the initial application timeline. At the beginning of the school year, the school's principal can offer these slots to individuals who were not admitted during the initial application timeline or applied after the initial application timeline, and who

receive funding for. On the other hand, the number of students they admit is bounded above as the program needs to keep spots open for the students admitted by the principal and also needs to meet the required student to staff ratios. In addition, a program might not want to admit the maximum number of students that student to staff ratios allow to be able to better serve those students it does admit. As a result, it is likely that programs have a target number of admissions that they can exceed slightly whilst still meeting the desired student to staff ratios. These programs will have some flexibility in how many students they admit during the initial application timeline.

the principal identifies as having a great need for the PrimeTime program. As most of these spots are generally offered to students who applied after the initial application timeline, few students who were waitlisted during the initial application timeline get admitted in this way.

Importantly, a cutoff rank will exist for each program that receives a higher number of applications than its capacity during the initial timeline. Students whose applications rank below this cutoff will be waitlisted whereas students with applications ranked at or above the cutoff will be admitted. Corresponding to this cutoff rank will be a cutoff priority score and cutoff application postmark date. Students with a priority score equal to this cutoff score and application postmark date after the cutoff date will be placed on the waitlist. On the other hand, students with a priority score equal to the cutoff score and application postmark date at or before the cutoff application postmark date will be admitted.[29] Importantly, the cutoff priority score and application postmark date will depend on the quantity, priority scores and postmark dates of all applications submitted during the initial timeline, making them impossible to perfectly predict for any individual applicant. As a result, the fundamental identifying assumption of this paper that applicants should be as good as randomly assigned near the cutoff application postmark date is likely to hold.

The ranking procedure of applications is described in detail on each year's paper application. Parents appear to be well aware of this procedure as most submit their child's application during the initial weeks of the application timeline. In figure 1.1a, I plot the distribution of postmark dates of all applicants with a priority score equal to the cutoff priority score in schools and years that had to waitlist applicants. Most applications are submitted during the initial two weeks of the application timeline. In particular, 42.6% of all applications are submitted during the first week and a further 15.0% are

[29]To give a concrete example, say a PrimeTime program admits 75 students during the initial application timeline and the applicant ranked 75^{th} has a priority score of 2 points and submitted his application on March 31^{st}. In this case, any applicant with a priority score of 2 points who submitted his application on or before March 31^{st} will be admitted and any applicant with a priority score of 2 points who submitted his application after March 31^{st} will be waitlisted.

submitted during the second week.

In this paper, I only consider students who applied to a PrimeTime program during the initial application timeline. Moreover, I focus on PrimeTime programs in which the highest ranked student to be waitlisted and the lowest ranked student to be admitted have the same priority score. Using a fuzzy regression discontinuity design, I then compare the short-run behavioral and academic outcomes of the lowest ranked students to be admitted and the highest ranked students to be waitlisted with the same priority score. The lowest ranked students to be admitted applied on or just before the cutoff application postmark date and the highest ranked students to be waitlisted applied just after this date. On average, the highest ranked students to be waitlisted will end up attending the PrimeTime program for fewer days during the upcoming school year. As explained above, they are less likely to attend the PT program as they are less likely to be admitted, and if they are admitted they are likely to be admitted off the waitlist during the school year, resulting in fewer days of program attendance. Besides their participation in a PrimeTime program however, the lowest ranked students to be admitted and the highest ranked students to be waitlisted should be very similar, allowing us to attribute any difference in outcomes between them to the causal impact of increased PrimeTime program attendance.

1.3 Empirical Framework

Ordinary least squares estimates of the causal impact of PrimeTime program attendance on short-run student outcomes for applicants to the PrimeTime program suffer from selection bias. For instance, using OLS we can identify the difference in average short-run outcomes between applicants who do and do not attend a PrimeTime program. However, this difference equals the causal effect of PrimeTime program attendance on short-run student outcomes across applicants who attend a PrimeTime program plus a selection bias term. The former is the term we want to identify and that is important from a policy perspective as it captures the effect of PrimeTime program attendance on the short-run outcomes of applicants to the PrimeTime program. The latter selection bias term is the difference in average potential short-run outcomes between applicants who do and do not attend the PrimeTime program if they both would not be able to attend the PrimeTime program. This term captures the influence of other determinants of student outcomes that are potentially correlated with PT program attendance such as parental involvement and students' academic histories.

The fuzzy regression discontinuity design outlined in the previous section allows us to separately identify the causal effect of PrimeTime program attendance on short-run student outcomes from the selection bias term. Namely, as we restrict students to be within a narrow window of the cutoff application postmark date, differences in the other determinants of student outcomes that are captured by the selection bias term should tend to zero. Formally, we assume the average treatment effect and the selection bias term vary continuously with the postmark date of the application at the cutoff postmark date. Under this assumption, the average causal effect of one more day of PrimeTime program attendance at the cutoff application postmark date can then be identified by the ratio of the reduced form effect of applying on time on short-run student outcomes over the first stage effect of applying on time on the number of days of PrimeTime program attendance.

This fuzzy regression discontinuity design can be captured in the following two-equation system (Imbens and Lemieux (2008), Lee and Lemieux (2010) and Dahl et al. (2014)):

$$p_{ist} = \alpha_1 + 1[t_{ist} \leq c_{st}]\left(g_l(c_{st} - t_{ist}) + \lambda\right) + 1[t_{ist} > c_{st}]g_r(t_{ist} - c_{st}) + \epsilon_{1ist} \quad (1.1)$$

$$y_{ist} = \alpha_2 + \beta\hat{p}_{ist} + 1[t_{ist} \leq c_{st}]f_l(c_{st} - t_{ist}) + 1[t_{ist} > c_{st}]f_r(t_{ist} - c_{st}) + \epsilon_{2ist} \quad (1.2)$$

p_{ist} indicates how many days student i attended the PrimeTime after-school program at school s during school year t. t_{ist} is the postmark date of the application of this individual to this PrimeTime program and c_{st} is the cutoff application postmark date in this school and year. y_{ist} is a short-run behavioral or academic outcome. g_l, g_r, f_l and f_r are unknown functions in the individual's application postmark date relative to the cutoff. The indicator for having an application postmark date at or before the cutoff, $1[t_{ist} \leq c_{st}]$, in the first-stage equation 1.1 is used as an instrument for days of participation in the PrimeTime program in the second-stage equation 1.2, $\hat{p}_{ist}$. The coefficient λ on $1[t_{ist} \leq c_{st}]$ in equation 1.1 can be interpreted as the jump in the average number of days of PT attendance at the cutoff application postmark date. The 2SLS estimate of β in equation 1.2 then captures how, on average, one more day of PrimeTime program attendance affects short-run student outcomes.[30] Importantly, to consistently estimate the effect of PrimeTime attendance via 2SLS, we need to assume that the only channel for individuals to be affected by the initial admissions decision is through changing the average number of days of PT attended. In addition, two stage least squares also requires the monotonicity assumption that being admitted to a PrimeTime program during the initial application timeline does not cause any individuals to attend the PrimeTime program for fewer days.

As an alternative to the 2SLS approach, I will also estimate the following reduced-form equation:

$$y_{ist} = \gamma_1 + 1[t_{ist} \leq c_{st}]\left(h_l(c_{st} - t_{ist}) + \pi\right) + 1[t_{ist} > c_{st}]h_r(t_{ist} - c_{st}) + \eta_{ist} \quad (1.3)$$

As before, h_l and h_r are unknown functions in the individual's application postmark date relative to the cutoff. The coefficient of interest is π. If the usual RD assumptions hold[31], this coefficient consistently captures the effect

[30]Formally, the jump in average days of PT program attendance at the cutoff is a results of two factors: (1) an increase in the likelihood of attending the PT program, and (2) an increase in the number of days of program attendance conditional on attending a positive number of days. Specifically, the likelihood of PT program attendance increases from around 45% to 85% at the cutoff and the average number of days attended conditional on attendance increases from around 120 to 155 at the cutoff. The average marginal effect of 1 more day of PT program attendance identified in this paper is thus a weighted average of (1) the average marginal effect of going from 0 to 155 days of PT attendance for the 40% of people at the cutoff who would not have attended PT if they had applied just after the cutoff, and (2) the average marginal effect of going from 120 to 155 days of PT attendance for the 45% of people at the cutoff who would only have attended the program for 120 days if they had applied just after the cutoff.

[31]In particular, this refers to there being no manipulation in the assignment variable as well as there not being any other factors that evolve discontinuously at the cutoff date.

of being admitted to a PrimeTime program during the initial application timeline on short-run student behavioral and academic outcomes. Importantly, to consistently estimate the coefficient of interest in this reduced-form framework does not require the strong assumptions necessary for the 2SLS framework.

The baseline regression specifications in this paper use applications submitted within a 5 week window of the cutoff application postmark date, include separate linear trends in the postmark date relative to the cutoff on either side of the cutoff and employ triangular weights. The triangular weights and bandwidth of 5 weeks ensure that I do not rely on data far away from the cutoff to identify local effects. To increase precision, I also include predetermined control variables. Specifically, I control flexibly for the student's gender, age, grade, ethnicity, English learner status, Special Education status and parental level of education using various indicators. I also include indicators for each of the 4 criteria on which a student‘s priority score is based, an indicator for whether a student applied to a before-school PrimeTime program as well and, when available, the lag of the dependent variable. Lastly, I include school by year fixed effects. This prevents differences in baseline outcomes across schools and years from biasing the estimate of the effect of the after-school program. The standard errors are clustered at the school by year level. Importantly, in the robustness section I show that the results of this paper are not sensitive to variations in the particular regression specification that I use.

Some students who apply to a PrimeTime program at a particular school do not end up attending a school in SDUSD for the entire school year. I do not observe outcome data for these students. If being waitlisted causes some students to attend a school outside of SDUSD, my results might be biased. In the next section, I check whether the probability of observing students‘ outcome data varies discontinuously across the threshold. Importantly, I find no evidence of this being the case.

Furthermore, some students who attend a school in SDUSD for the entire school year do not attend the school of the PrimeTime application for the entire year. These students might attend another school in SDUSD for the entire school year or only attend the school of the PrimeTime application for part of the school year. If being waitlisted caused some students to attend another school in SDUSD, the causal impact of attending a PrimeTime program identified in this paper could potentially also be partially attributable to attending a school that is different from the school of the PrimeTime application. In section 1.5.1, I show evidence that the probability of attending the school of the PrimeTime application for the entire year varies discontinuously across the threshold for one of the two providers.[32] Restricting the sample even further by dropping students who do not attend the school of the PrimeTime application for the entire school year could thus bias my results. In the main analysis I therefore use the full sample of students who have non-missing outcome data. In section 1.5.4, I show that the main results hold once I restrict the sample to students who attend the school of the PrimeTime application for the entire year for the provider where differential attrition to other schools inside of SDUSD is not an issue. Restricting the sample to these students leads to increased precision and highlights that the causal effects estimated in this paper are due to the PrimeTime program and not due to waitlisted individuals attending different schools than accepted individuals.

1.4 Data and Implementation

In the analysis I use two different administrative datasets. Individuals can be linked across these datasets based on unique individual identifiers. The first dataset consists of application and attendance data to PrimeTime programs. The application data contain the content of applications submitted during the initial application timeline. It includes the name of the school, the

[32]Students are notified of PrimeTime admission decisions in late June. At this point in time, students have limited options to change the school in which they plan to enroll during the upcoming school year if they wish to stay inside of SDUSD. In late June, some students will not be scheduled to attend their neighborhood school in the upcoming year as they are scheduled to attend a different school inside SDUSD through the School Choice Program. Except for students who move during the school year, only these students could easily switch schools in SDUSD as they still have the option to enroll in their neighborhood school during the summer and during the first few weeks of the upcoming school year.

priority score of each application, the individual priority score criteria that the applicant received points for, the postmark date of the application and whether the applicant was admitted or waitlisted for the PrimeTime program after the initial application timeline. The attendance data have information on the total number of days that an individual attended an after-school PrimeTime program at a particular school during the school year.

For many schools and years, parts of the application data that are necessary to conduct the regression discontinuity analysis are missing. The application and attendance data that are complete cover the years 2010-2011 to 2014-2015 and are from two large providers. Together, these providers are responsible for operating over half of all the close to 130 PrimeTime after-school programs in SDUSD.[33] Complete data for both providers are present in 2012-2013 and 2014-2015. For the other years, complete data are only present for one of the providers.[34]

To implement the regression discontinuity design, I restrict the sample to PrimeTime programs that received more applications than their capacity during the initial timeline. I also drop PrimeTime programs where the lowest ranked student to be admitted and the highest ranked student to be waitlisted have different priority scores. This results in observations from 90 different school by year combinations that are divided roughly equally between the two providers[35]. For the analysis, I pool all the available data from all years and providers. I only keep those applicants who have a priority score equal to the cutoff priority score in their school and year. In some cases, capacity constraints force a provider to waitlist one applicant at the cutoff whilst offering admission to another. For instance, it could be that a provider has the

[33]In particular, in 2014-2015, provider 1 operated 23 out of the 127 programs and provider 2 operated 46 out of the 127 programs.

[34]It is unclear how the providers with complete data for some years differ from the other providers that did not have complete data for any year. It could be that the administration of these providers is more organized, as they kept complete data on the admissions process. It is unclear however how this degree of organization affects the impact that the after-school program has on student outcomes. This paper can therefore only tell us something about the impact of the after-school program on students near the cutoff for providers 1 and 2, and not for the other providers operating in SDUSD.

[35]48 of these school by year combinations are from provider 1 and 42 are from providers 2.

capacity to admit 50 students during the initial timeline and that the applicants ranked 50^{th} and 51^{st} have identical postmark dates and priority scores. In these cases, the provider is given the discretion to decide which applicant to admit. Since these admissions decisions might be based on unobservable student characteristics that are correlated with short-run student outcomes, I drop all individuals that are at the cutoff if there are both waitlisted and admitted students at the cutoff in a school and year.

Figure 1.2 shows the distribution of priority scores of all applications submitted during the initial application timeline to schools in my sample. As priority scores ranged from 0 to 5 up to 2012-2013 and from 0 to 4 afterwards, I show a separate distribution for the years 2013-2014 and 2014-2015. In all years, the mode priority score equals 3 and very few applications have a priority score of 0, 1 or 5. In 63.2% of all schools and years in my sample the cutoff priority score equals 2 and in 24.2% it equals 1. Furthermore, 63.8% of all applicants had a priority score that exceeded the cutoff priority score. These applicants were admitted to the PrimeTime program regardless of when they submitted their applications. On the other hand, 24.9% of applicants had a priority score equal to the cutoff score and 11.3% had priority score that was lower than the cutoff score. The latter group of applicants were all waitlisted. These statistics suggest that the students who are in the cutoff priority score group and who are considered in this paper fall in the bottom third of all students in terms of their need for the program as measured using priority scores.

Figure 1.1b displays the distribution of cutoff application postmark dates for the schools and years in my sample. Most cutoff application postmark dates occur early on during the initial application timeline. This is not surprising given the evidence shown earlier in figure 1.1a that most applications are submitted during the initial weeks of the application timeline.

The application and attendance data are merged with administrative data from SDUSD that provides a longitudinal panel of students‘ school records for as long as they are enrolled in a school within SDUSD. These school records contain students' personal characteristics such as gender, ethnicity, English-Learner status, parental level of education and special education status, as well as students' report cards, standardized testing scores and school attendance records.

The short-run academic outcomes studied in this paper are standardized test scores, report card grades in K-5 and middle school GPA's, and mostly focus on math and ELA (English Language Arts) performance. The standardized test scores are from the California Standards Test (CST) in ELA and mathematics. Until 2012-2013 all students in grades 2 to 11 took these tests towards the end of each school year. For the analysis, I normalize test scores using districtwide means and standard deviations for each subject, grade and year.[36] The report card grades are from students' Standards Based Report Cards (SBRC). These report card grades cover math, reading and writing, and are available in all years for students in K-5. Up to 2013-2014, the report card grades used in this paper represent the average grade on a scale from 1 ("Below Basic") to 4 ("Advanced") that a student received during the school year in a given subject. In 2014-2015, the grading standards and scale changed to be in line with the Common Core. For this year, the report card grades used in this paper still represent a student's average grade in a given subject on a scale from 1 to 4, but a 1 now means "Beginning progress towards grade level expectations" and a 4 means "Exceeding grade level expectations". Since report card grades in 2014-2015 might not be comparable to those in earlier years, I perform the analysis separately for report card grades in 2014-2015. For the analysis, I normalize the average report card grades using districtwide means and standard deviations for each subject and grade. As grading standards might not be comparable across schools and years, I also include school by year fixed effects in all regressions that I run in this paper. Lastly, I also look at middle school students' overall GPA as well as their GPA in English and Math separately. These measures are available for students in grades 6 to 8 in all years and are normalized using districtwide means and standard

[36]The ELA CST test is grade-specific for all years and grades. The math CST test is grade-specific up to grade 6 in each year. After grade 6, students can choose which version of the math CST test to take. Hence, for students in grade 7 and 8, math CST scores are standardized by grade, year and the version of the math test that is taken. All regressions will control for differences in the version of the math test that students take by including separate dummy variables for each version.

deviations for each grade.

Most of the short-run behavioral outcomes studied in this paper are from students‘ Standards Based Report Cards. There are six key teacher-reported measures of student behavior on these report cards: student's level of interest in learning, student's level of respect for others in class, student's preparedness for class, student's assignment completion rate, student's level of critical thinking and student's feeling of responsibility for learning. Each measure is based on a scale from 1 ("Rarely") to 3 ("Consistently") and is available for students in K-5. Student interest level and level of respect for others in class are available for all years. Student's preparedness for class and assignment completion rate are available only until 2013-2014, and student's level of critical thinking and feeling of responsibility for learning are only available for 2014-2015. For the analysis, these behavioral grades are normalized using districtwide means and standard deviations for each measure and grade.

In addition, I also look at a student‘s citizenship GPA in middle school. In middle school, students receive a citizenship grade for each course that they take. This citizenship grade aims to measure a student‘s overall classroom behavior in a class and ranges from 0 ("Unsatisfactory") to 4 ("Excellent"). A student's average citizenship grade across all courses taken in a given year is available for all students in grades 6 to 8 in all years. These citizenship GPAs are normalized using districtwide means and standard deviations for each grade. Lastly, I also look at the percentage of school days that students are absent. This measure is available for students in all grades and years.

Table 3.1 shows summary statistics for applicants with the cutoff priority score, who constitute my sample, and for applicants in the same schools and years with priority scores above the cutoff priority score, who were admitted to the PrimeTime program regardless of the postmark date of their application and make up the bulk of after-school program participants. The fourth column shows results from two-sample t-tests that test whether the mean of the relevant variable is the same for both groups. Applicants with the cutoff priority score are more likely to be White and less likely to be Hispanic or Black than applicants with higher priority scores. They are also less likely to be English Learners and more likely to have a parent with a college degree than applicants with higher priority scores. Furthermore, applicants with the cutoff priority score are also much less likely to come from a single parent household and slightly less likely to have parents who are both full-time employed. As the behavioral and academic outcomes are standardized using districtwide means and standard deviations, the table also shows that students in my sample perform better at baseline on all these measures than the average district student in their grade and year. On the other hand, students in higher priority score groups perform worse at baseline than the average district student on all these measures. The difference in average baseline performance between students in my sample and students with higher priority scores is large, and varies from around 0.2 to 0.4 standard deviations. It is not surprising then that students with higher priority scores are 30 percentage points more likely to qualify for academic assistance than students in my sample. The differences in predetermined characteristics and baseline outcomes between students in the cutoff priority score group and students with higher priority scores suggest that these groups differ greatly. Students in my sample appear to be relatively advantaged when compared to students with higher priority scores who make up the bulk of after-school program participants. It is therefore unlikely that the local average treatment effect identified in this paper generalizes to all students who were admitted to a PrimeTime program during the initial application timeline.[37]

[37]Recall that students' priority scores are based on participating in PrimeTime in the previous year, coming from a single parent household, having parents who are both full-time employed and qualifying for academic assistance. Since the students in my sample have lower priority scores, the differences I find in table 3.1 for these characteristics are not surprising.

1.5 Results

1.5.1 Threats to Identification

The RD design is not valid if students can manipulate the assignment variable. A student could potentially manipulate the relative postmark date of his application if he knew the exact capacity of the PrimeTime program and the number of applications submitted at any given time with a priority score equal to or exceeding his own priority score. It is very unlikely that a student would have this information. Among others, it would require access to the content of all the applications submitted by any given date.[38] Potentially more worrisome is that the PrimeTime program staff who conduct the admissions process could actively decide upon the location of the cutoff application postmark date. As discussed in section 1.2.3, programs have some flexibility in the number of students that they admit during the initial application timeline. As a result, it would for instance theoretically be possible for staff members to actively decide whether the cutoff application postmark date should be on March 31st or April 1st based on the students who applied on April 1st. This would be troublesome if staff members would make such a decision based on characteristics of these applicants that are correlated with these students' future outcomes.

Figure 1.1c shows the distribution of postmark dates relative to the cutoff in schools that admitted all students with the cutoff priority score and

[38]Manipulation by students would also be possible if cutoff priority scores and postmark dates are highly correlated over time for a given school and if previous years' cutoffs are known to students. Students are not informed of the cutoff score and postmark date in a particular year, as they are only told whether they are waitlisted or admitted. This makes it unlikely that they would know previous years' cutoffs. In addition, knowing previous years' cutoffs is not necessarily informative to students. For instance, 8 schools are in my sample in both 2013-2014 and 2014-2015, and 9 schools are in my sample in both 2011-2012 and 2012-2013. Only 11 of these 17 schools have the same cutoff priority score in both years. For the other schools, the cutoff priority score is either higher or lower in the later year. In addition, for these 11 schools with equal cutoff priority scores, the absolute difference in cutoff application postmark dates between the two years is on average 9.5 days, with the cutoff date being either earlier or later in the later year. As a result, applying just before or at last year's cutoff application postmark date is unlikely to cause an individual to systematically be located to the left of or at this year's cutoff.

cutoff postmark date[39]. The general shape of the distribution of postmark dates relative to the cutoff is a result of the nature of the admissions procedure. First, for a given cutoff application postmark date, not all possible relative postmark dates can be observed. For instance, if the cutoff postmark date is day 15, the possible relative postmark dates vary from -15 to 37. As a result, the density of relative postmark dates declines as we move away from the cutoff. In addition, most applications are submitted during the early weeks of the initial application timeline, causing the density of relative postmark dates to decline as we move away from the cutoff to the right. Lastly, most cutoff application postmark dates occur during the early weeks of the application timeline, causing the density to decline as we move away from the cutoff to the left.

There are more students with a postmark date equal to the cutoff postmark application date than students with a postmark date just after the cutoff date. That is, the density is higher at zero than at 1. A formal McCrary-test (see McCrary (2008)) for the presence of a discontinuity in the density of application postmark dates at the cutoff also confirms this. The estimated log-difference between the frequency of applications on either side of the cutoff is 0.182 with a standard error of 0.096 and is statically different from 0 at the 10% level of significance.[40] This observed distribution could be due to some students being able to manipulate the assignment variable, although I explained earlier that this is very unlikely. It could also be a result of PrimeTime staff members actively deciding upon the location of the cutoff application postmark date. For instance, suppose some PrimeTime programs have a policy of either admitting or waitlisting everybody who applies on the same

[39]Remember that some schools had to waitlist some individuals with the cutoff priority score and cutoff postmark date due to capacity constraints. See section 1.4 for a discussion of this issue. In the regressions, I drop all individuals with the cutoff postmark date if the school both admitted and waitlisted individuals with the cutoff priority score and cutoff postmark date.

[40]This test was performed using a bin size of 1 and a bandwidth of 12 days. The positive coefficient indicates that there are more applications with a postmark date at or before the cutoff than applications with a postmark date after this cutoff.

day.[41] In addition, they might decide where to locate the cutoff application postmark date based on the presence of at least one high-need student among the students who applied on a given day. Since such a high-need student is more likely to have applied on days on which many people apply, this could result in a higher density of applications at the cutoff than just to the right of it.

However, for a given school and year, the cutoff application postmark date is defined as the application postmark date of the lowest ranked student that is admitted. By definition, every school and year will have students with a relative postmark date of 0. However, schools do not necessarily receive applications on each day of the initial application timeline. As a result, there is not always a student with a relative postmark date of 1. That is, it is for example possible that the cutoff postmark date of a given school is March 31^{st} and that the earliest application received after this date has a postmark date of April 4^{th}. In this case, there are no applications with relative postmark dates of 1, 2 or 3. This could also explain the discontinuity in the distribution of relative postmark dates at the cutoff observed in figure 1.1c. To correct for this feature of the data, in figure 1.1d, I plot the distribution of "ranked" relative application postmark dates in which days on which no applications were submitted are dropped. That is, in the above example, the "ranked" relative application postmark date of the application with postmark date of April 4^{th} is 1. In the case of manipulation by staff members, we should observe a discontinuity at the cutoff in this distribution as well. In an earlier example, I gave the example of staff members potentially deciding between a cutoff application postmark date of March 31^{st} and April 1^{st}. This is not possible if no applications were submitted on April 1^{st}. Instead, in the above example the relevant comparison would be between March 31^{st} and April 4^{th}. To look at evidence for manipulation of the assignment variable by PrimeTime staff members, we therefore need to look at the distribution of "ranked" relative postmark dates as given in figure 1.1d. As can be seen in this figure, there is

[41] That is, these programs do not waitlist some students who applied on the cutoff application postmark date.

no apparent discontinuity at the cutoff in this distribution. In fact, there are slightly more applications with a ranked relative application postmark date of 1 (157) than 0 (141). A formal McCrary-test confirms this. The estimated log-difference between the frequency of applications on either side of the cutoff is -0.101 with a standard error of 0.103 and is not statistically significant.[42]

Although I just argued that manipulation by either students or PrimeTime staff members is unlikely, if it did happen we would expect the distribution of predetermined student characteristics to not be smooth around the cutoff as treatment was assigned non-randomly. In table 1.2, I test whether baseline student characteristics change at the cutoff application postmark date. To do so, for each pre-determined student characteristic, I estimate equation 1.3 with that characteristic as the dependent variable and using the baseline regression specification that I outlined in the previous section [43]. I include all students regardless of whether they attended a school in SDUSD for the entire year. As can be seen in the table, none of the estimated coefficients are significant at the 10% level. Both the student background characteristics and the lagged values of the outcome variables thus seem to vary continuously across the threshold. This makes it unlikely that students or PrimeTime staff members were actively manipulating the relative postmark date of applications.[44] However, as I cannot fully exclude the possibility of manipulation, I will also estimate donut RD specifications in which I exclude students who applied within a one-day window on either side of the cutoff application postmark date. The intuition behind these donut RDs is that manipulation should

[42]This test was performed using a bin size of 1 and a bandwidth of 6 days. The negative coefficient indicates that there are less applications with a "ranked" relative postmark date at or before the cutoff than applications with a "ranked" relative postmark date after this cutoff.

[43]I do not include the other predetermined characteristics as control variables in these regressions.

[44]Admissions are conducted by each provider separately. Since I pool data from two providers, the insignificant results might mask opposing manipulation practices by providers. To make sure this is not the case, I also estimated all regressions in table 1.2 separately for each provider. For provider 1, 3 out of 27 variables were significant at the 10% level and for provider 2, 2 out of 27 variables were significant at the 10% level. This is what we would expect purely due to chance if the predetermined characteristics are not correlated within students.

occur right out around the threshold. Hence, by dropping applications submitted close to the cutofff date, I should eliminate all potentially manipulated data.

I only observe outcome data for PT applicants who are enrolled in a school in SDUSD for the entire year. If being waitlisted causes some students to attend a school outside of SDUSD, my results might be biased. To make sure this is not the case, I estimate equation 1.3 with a dummy for being enrolled in a school in SDUSD for the entire year as the dependent variable and using the baseline regression specification[45] This results in an estimated coefficient of 0.0054 with a standard error of 0.025, a statistically insignificant and very small effect given that over 91% of all applicants are enrolled in SDUSD for the entire year. Hence, differential attrition across the cutoff is not likely to be an issue in my study.[46]

As discussed in section 1.3, some of the students who attend a school in SDUSD for the entire year do not attend the school of their PrimeTime application for the entire school year. For one of the providers, I find that

[45]I do not include the predetermined control variables in this regression. I did however also run this regression including these variables and doing so does not change the results.

[46]Most of the outcome variables studied in this paper are available only for a subsample of students who are in certain grades and years. The insignificant overall result might hide the differential attrition that could be taking place for some of these particular subsamples of students. To make sure this is not the case, I first define six possibly overlapping groups of students based on the years and grades for which particular outcome variables are available. Then, for each of these groups, I estimate whether the probability of observing the outcome of interest varies continuously across the cutoff and whether the predetermined characteristics of students with non-missing outcome data vary smoothly across the cutoff. I perform these estimations for each provider separately as well as for the combined sample. In general, whether I combine data from both providers or not, I do not find evidence of differential attrition across the threshold. The only exception occurs when looking at the standardized behavioral scores for provider 1 in K-5 before 2014-2015. In this case, those at the cutoff are 7.9 percentage points more likely to be in SDUSD all year than those who are to the right of the cutoff. This does not occur in the other provider, so looking at this provider only in the robustness checks to my main specification will help alleviate concerns that my results are biased because of differential attrition. Moreover, predetermined student characteristics are balanced in nearly all subsamples, with at most 3 out of 27 variables being significant at the 10% level. Importantly, for all of the subsamples, the lagged values of the outcome variable is never significant at the 10% level. The only exception here occurs when looking at middle school GPA's in provider 2, in which case 5 out of 27 variables are significant at the 10% level. Again, I will also look at results for the other provider, in which none of the variables are significant at the 10% level, to make sure the results for middle school GPA's are not biased by differential attrition.

being waitlisted causes students to not attend the school of their PrimeTime application for the entire year. In particular, I estimated equation 1.3 with a dummy for being enrolled in the school of the PrimeTime application for the entire year as the dependent variable and using the baseline specification. For one of the two providers, this resulted in a statistically significant coefficient of 0.13, a large effect given that 84% of all applicants attend the school of their PrimeTime application for the entire school year. As discussed earlier, restricting the sample even further by dropping students who do not attend the school of the PrimeTime application for the entire school year could thus bias my results. I therefore use the full sample of students who attended a school in SDUSD for the entire year. In section 1.5.4, I show that the main results hold once I restrict the sample to students who attend the school of their PrimeTime application for the entire year for the provider where differential attrition to other schools inside of SDUSD is not an issue. This highlights that the causal effects estimated in this paper are due to the PrimeTime program and not due to waitlisted individuals attending different schools than accepted individuals.

1.5.2 Graphical Results

Figure 1.3 shows the effect of the application postmark date on PT attendance. In each graph, an observation is the unweighted average of the outcome variable in a 6-day bin, where the bin is based on the relative postmark date of individuals' applications. As applications cannot be postmarked on Sunday, 6-day bins contain all applications submitted during a one-week interval. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using separate linear trends on either side of the cutoff. Figure 1.3a shows that there is a sharp jump in the average number of days an individual attends a PT program at the cutoff, with the average days of attendance rising from roughly 50 to 130 days. This increase in the average number of days attended is a result of two factors. First, as can be seen in figure 1.3b, the likelihood of attending a PrimeTime program for any positive number of days increases from around 0.45 to 0.85 at the cutoff. At the same time, as can be seen in figure 1.3c, conditional on attending any positive number of days, the average number of days attended increases from around 120 to around 155 at the cutoff. Applicants who are initially waitlisted are often admitted off the wait list after the start of the school year. As a result, conditional on attending, these students still have lower average attendance than those students who were initially admitted. Overall, these graphs provide strong support for the relevance of the instrument. Applying at or before the cutoff application postmark date resulted in a large increase in PT attendance.

Figures 1.4 and 1.5 illustrate the reduced form relationship between the various academic and behavioral outcomes and the relative postmark date of the application. Again, each observation is the unweighted average of the outcome variable in a 6-day bin and the straight lines are regression lines estimated using separate linear trends on either side of the cutoff. Figure 1.4 reveals no clear discontinuities at the cutoff for almost all the academic outcomes. Both the estimated regression lines and the raw data averages line up very closely on either side of the cutoff in almost all graphs. The only exception is the SBRC writing score in 2014-2015. Individuals who applied just before or on the cutoff date have a SBRC writing score score that is on average around 0.25 standard deviations higher than that of students who applied just after the cutoff. However, based on the confidence intervals, it is hard to tell whether this difference will turn out to be statististically significant in the regression analysis presented in the next section.

The behavioral outcomes shown in figure 1.5 present a different picture. Although school attendance does not jump at the threshold, most behavioral report card grades in K-5 do. In particular, students' level of preparation for class and assignment completion rate is on average around 0.2 standard deviations lower to the left of the cutoff. On the other hand, students' level of critical thinking and feeling of responsibility for their own learning is around 0.2 standard deviations higher to the left of the cutoff. Although these two results appear contradictory, it is hard to compare the various graphs as they are based on different students. Students' level of preparation for class and assignment completion rate is available for K-5 students from 2010-2011 to 2013-2014. Students' level of critical thinking and feeling of responsibility for their own learning is only available for K-5 students in 2014-2015. The Prime-Time program remained largely unchanged across all these years, suggesting that differences in the type of students and schools at the cutoff might explain the difference in results across years. I look further into this potential explanation in section 1.5.6. In short, the reduced-form graphs thus suggest that PT attendance has an impact on behavioral but not on academic student outcomes, and that this impact might be heterogeneous across students and schools.

In many of the graphs, the relationship between the relative postmark date of an application and the outcome variable is of opposite sign on either side of the cutoff. Although this does not invalidate the empirical approach of this paper, we would a priori expect them to be the same. In appendix 1.7.1 I show that these opposite signs can largely be attributed to changes in observable student characteristics as we move away from the cutoff. Appendix figures 1.7, 1.8 and 1.9 replicate respectively figures 1.3, 1.4 and 1.5 using residuals of the relevant variable instead of the variable itself. These residuals are generated from a regression of the relevant variable on the predetermined control variables that are included in the baseline regression specification and school by year fixed effects. Using these residuals eliminates the effects on the figures of changes in observable characteristics of students as we move away from the cutoff. The observed discontinuities at the cutoff when using residuals are similar to those observed when using the actual values of the variables. This is not surprising given that I showed earlier that student co-variates evolve smoothly around the cutoff. Importantly, in contrast to the original figures, the relationship between the postmark date of an application and the outcome variable is generally of the same sign on either side of the cutoff when using residuals. Moreover, in cases where it is still of opposite sign, it is generally much weaker than when using the actual variables.

1.5.3 Regression Results

Table 1.3 shows the baseline estimates for the effect of PT program attendance on student outcomes. The effect on academic outcomes in math and ELA is reported in the top part of the table, and the effect on behavioral outcomes is reported in the bottom part. Column 1 of table 1.3 estimates the first stage equation 1.1 and corresponds to figure 1.3a. This column shows that applying at or just before the cutoff application postmark date on average resulted in 71 to 86 more days of PT attendance, depending on the particular outcome studied. This is a large effect given that those who applied just after the cutoff application postmark date on average only attended the PT program for around 50 days. Importantly, the F-values reported in column 4 show that the instrument is relevant.

The second column of table 1.3 shows the reduced form estimates and corresponds to figures 1.4 and 1.5. The third column of table 1.3 shows the 2SLS estimate of the causal effect of attending one more day of PT and can be obtained by dividing the reduced form coefficient in column 2 by the first stage coefficient reported in column 1. Looking first at academic outcomes, none of the estimated coefficients in columns 2 and 3 are statistically significant at the 10% level. This is not surprising given that we did not see very clear discontinuities at the cutoff in figure 1.4. The fact that academic outcomes appear to be unaffected does not seem to be a product of the particular grades and years for which data are available, as the various academic outcomes are based on students across all different years and grades for which I have data. It thus appears that for students at the cutoff an increase in PrimeTime attendance does not affect academic outcomes in ELA and Math. This is somewhat surprising given that the academic component of the PrimeTime program specifically aims to improve student outcomes in these subjects. It is important to note however that the zeroes are imprecisely estimated due to the relatively small sample size of this study. The coefficients have confidence intervals that contain large effects. For the most precisely estimated coefficients on SBRC grades in math, reading and writing in K-5 up to 2013-2014 the confidence intervals contain effect sizes of up to 0.15 standard deviation in either direction as a result of 80 more days of PT attendance.[47]

On the other hand, behavioral student outcomes do appear to be affected by PrimeTime program attendance. In particular, in 2010-2011 to 2013-2014, students in K-5 who applied just at or before the cutoff on average received a 0.2 standard deviation lower grade on a measure of their assignment completion rate and a 0.15 standard deviation lower grade on a measure of their preparedness for class. These estimated coefficients are significant at the 5% and 10% level, respectively. For the average student who applied just before or at the cutoff and attended the PT program for 155 days, this translates into a 0.40 standard deviation lower grade on the measure of their assignment completion rate and a 0.29 standard deviation lower grade on the measure of their preparedness for class than if he did not attend PT at all. On the other hand, in 2014-2015, students in K-5 who applied just at or before the cutoff on average received a 0.34 standard deviation higher grade on a measure of taking responsibility for their learning. This coefficient is significant at the 10% level. For the same average student at the cutoff as mentioned above, this would translate into a 0.64 standard deviation higher grade on this measure than if he did not attend PT at all. The other behavioral outcomes, including middle school citizenship GPA and the fraction of days that a student is absent, appear to be unaffected by PT attendance. The coefficients on these behavioral outcomes are small and statistically insignificant.

So far, the results thus indicate that for students at the cutoff PT program attendance does not affect academic outcomes, but does influence in-class behavior.[48] As the laws regulating how after-school programs such

[47]To place these effect sizes into context, average standardized baseline SBRC grades in math, reading and writing in K-5 equal 0.40 for students in the cutoff group. Assuming these standardized grades follow a standard normal distribution, an increase in these scores by 0.15 standard deviations would move a student from the 66^{th} to the 71^{st} percentile. A decrease of 0.15 standard deviations would move a student from to the 66^{th} to 60^{th} percentile.

[48]These results are very different when estimating equation 1.2 using OLS instead. In appendix table 1.60, I show the results from estimating equation 1.2 using OLS and by not controlling for the relative postmark date of the application. The sample for these regressions only includes students who applied on or at most 29 days before the cutoff postmark date and who attended the school of the PrimeTime application for the entire year. All students

as PrimeTime have to be administered remained unchanged throughout my sample period, it is not immediately clear why PT attendance seems to have opposite effects on student behavior in K-5 across the different measures of student behavior that are available in different years. I explore several explanations for these results in more detail in section 1.5.6 after performing robustness checks of the baseline results and looking at how the estimated effects vary across subgroups of students.

1.5.4 Robustness Checks

In this section, I check the robustness of the baseline results using a variety of specifications that are common in RD studies. Although I perform these robustness checks for all the outcome variables listed in table 1.3, for the sake of brevity I only show robustness tables for the three statistically significant behavioral outcome measures discussed above. The robustness tables for the other variables can be found in appendix section 1.7.2.

Each column of each robustness table shows the estimated 2SLS coefficient from a variation of the baseline specification[49]. The corresponding first stage F-statistic is listed at the bottom of each column. For comparison, I show the estimated coefficient under the baseline specification in column 1 of each table. In column 2, I drop all control variables and the school by year fixed effects. If the control variables are balanced across the cutoff, dropping them should not affect the value of the discontinuity at this cutoff. In column 3, I do not include triangular weights. I include separate quadratic and cubic trends in the postmark date relative to the cutoff on each side of the cutoff in

in this sample are to the left of the cutoff and are thus admitted to PrimeTime. The degree of participation is therefore completely determined by students and their parents, making selection an important issue. As is clear from looking at table 1.60, the OLS regressions show that PT attendance has a positive effect on ELA test scores, MS ELA GPAs and MS overall GPAs and a negative effect on math test scores and possibly report card grades in reading. In addition, students' behavioral report card grades are negatively or not affected by PT attendance and school attendance improves due to increased PT attendance. The difference between these results and the baseline RD results presented in table 1.3 emphasizes the importance of accounting for selection using the RD approach.

[49]I multiplied this coefficient by 80 in each column to correspond to the average change in average days of PT attendance at the cutoff.

columns 4 and 5, respectively. In columns 6 and 7, I check whether the results are sensitive to the chosen bandwidth. In particular, in columns 6 and 7 I include applications submitted within a respectively 3 and 7 week window of the cutoff application postmark date, as opposed to the 5 week window used in the baseline specification. In column 8 I run a donut RD and exclude all applications submitted within a 1-day window of the cutoff application postmark date[50]. Assuming that potential perfect manipulation of the assignment variable occurs right around the threshold, excluding these observations should eliminate all potentially manipulated data.

In addition, in column 9, I drop all students who also applied to the before-school PrimeTime program. As explained earlier on, admissions to the before-school PrimeTime program are conducted in the same way as admissions to the after-school program. Including students who applied to the before-school program could bias my estimates if student outcomes are affected by the before-school program and if the cutoff application postmark date and priority score for the before-school program are heavily correlated with the cutoff application postmark date and priority score of the after-school program. In this case, the jump in student outcomes at the cutoff might be partially attributable to attendance of a before-school program. In column 10, I drop all schools that have an on-campus fee-based after-school program. If many students in the control group end up attending fee-based programs that are possibly very similar to the PrimeTime program, my estimates might be weakened. In schools without on-campus fee-based programs, a smaller fraction of students will attend a fee-based program, reducing this potential effect. Lastly, in columns 11 and 12 I estimate the baseline specification separately for each provider. As discussed in footnote 46, there is suggestive evidence of differential attrition around the cutoff for one of the providers for some outcome variables. Hence, by estimating the baseline specification separately for each provider, I can check that the estimates are not driven by this differential attrition. In addition, similar estimates across providers would also suggest

[50]When excluding applications submitted within a 1-day window of the cutoff application postmark date, I exclude applications with a relative postmark date of 0 or 1.

that the estimated impact of the PrimeTime program is not very sensitive to the particular organization providing the program.

The standard errors are clustered at the school by year level in all specifications to account for possible correlation in the error terms within a given school and year. However, there are very few clusters for some outcome variables and specifications. This small number of clusters can cause the estimated standard errors to be biased towards zero (Bertrand et al. (2004)). To make sure that the significance of the estimated effects does not crucially depend on this limited number of clusters, I use a Wild bootstrap procedure (Davidson and MacKinnon (2010))[51] to calculate a new p-value for the estimated 2SLS coefficient in each specification. I report this new p-value at the bottom of each column.

Across all the robustness checks and outcome variables, having an application postmark date at or before the cutoff is a relevant instrument for average days of PT attendance. The F-statistic exceeds 10 in nearly all specifications and is always larger than 9.

The estimated effects of days of PT attendance on academic student outcomes are largely insensitive to the particular regression specification used. As can be seen in tables 1.11 through 1.18 of the appendix, standards based report card grades in reading, writing and math as well as standardized test scores in math and ELA consistently appear to be unaffected by days of PT attendance. Across virtually all robustness checks, the coefficients are close to zero and highly statistically insignificant.[52] When looking at middle school overall, math and ELA GPAs, a similar picture arises. The estimated effects are close to zero and statistically insignificant in almost all specifications. As

[51] I thank Claudio Labanca for helping me implement this procedure.

[52] Only a few of the estimated effects are statistically significant. However, these effects change sign and size and are highly insignificant in other specifications, making it hard to draw any conclusions based on them. First, when looking at standards based report card grades in math up to 2013-2014, the estimated effect is negative and significant at the 10% level when introducing a cubic control in the relative postmark date of applications. However, in other specifications the estimated effect is positive or close to zero and highly statistically insignificant. Lastly, the effect on standardized test scores in math is significant at the 5% level and negative when introducing a cubic control in the relative postmark date of applications, but virtually zero with a p-value of over 0.5 in 8 other specifications.

discussed in footnote 46 however, there is some evidence of outcome data being differentially missing across the cutoff for provider 2 in the case of these middle school outcomes. Column (10) of tables 1.19 to 1.21 in the appendix shows results based on the baseline specification using only data for provider 1 for which there was no evidence of such differential attrition. These results show that attending the PT program for 80 more days per school year results in an increase in overall and math GPAs of close to 0.3 standard deviations. These effects are significant at respectively the 10 and 15% level, but are based on only 7 school by year combinations and 160 students. This small sample and the lack of significant results when data from both providers are used mean it is hard to say whether this finding is generalizable to a larger sample of schools as well.

The estimated effects of PT attendance on behavioral student outcomes are also largely robust to the regression specification used. In particular, as can be seen in tables 1.22 through 1.25 in the appendix, the percentage of school days that students are absent as well as their level of interest in learning, their level of respect for others in class and their level of critical thinking are all consistently unaffected by days of PT attendance. As can be seen in table 1.26 in the appendix, middle school citizenship GPA is also almost always unaffected. Again, when looking at provider 1 only for this middle school outcome to deal with the potential differential attrition in provider 2, there seems to be a slight hint of a positive effect of PT attendance. However, the p-value corresponding to the estimated coefficient of 0.19 is 0.21.

The robustness checks for student's assignment completion rate, student's preparedness for class and student's feeling of responsibility for learning can be found in respectively tables 1.4, 1.5 and 1.6. The coefficient on student assignment completion rate is around -0.2 in most specifications and statistically significant at the 10% level. However, the coefficient goes to zero and the significance disappears when using a 1-day donut.[53] The coefficient on student preparedness for class is around -0.15 in most specifications and significant at

[53] In particular, when using a 1-day donut the coefficient changes to -0.029 with a standard error of 0.11.

the 10% level. However, this coefficient is more sensitive to changes in the specification. In particular, the coefficient goes to zero and the significance disappears when not including triangular weights, when using a wider bandwidth and when using a 1-day donut. For both variables, the results that use data from only one of the two providers are similar to the pooled baseline result but are less precisely estimated. The estimated effect on student's feeling of responsibility for learning is around 0.30 and almost always significant at the 5 or 10% level. The significance of this coefficient, but not the size, disappears when including a cubic trend in the relative postmark date and when dropping the control variables and school by year fixed effects. In contrast to the other two variables, the estimated effect remains significant at the 10% level or very close to the 10% level when only using data from one provider. Overall, the estimated positive effect on student's feeling of responsibility for learning in 2014-2015 is very robust to the specification chosen, whilst the negative effect on student's assignment completion rate up to 2013-2014 and, in particular, student's preparedness for class up to 2013-2014 seems to be more sensitive to the regression specification. In particular, the fact that both these negative effects are sensitive to using a 1-day donut RD suggests that PrimeTime staff members might have set the application cutoff date endogenously up to 2013-2014 to include applicants that needed help in improving their behavioral outcomes.

In short, few of the estimated baseline effects are drastically affected by variations in the regression specification. Academic outcomes almost always remain unaffected, although there is some hint that middle school academic outcomes could be positively affected at a small subset of PT programs. In addition, the sign and size of the coefficients on the three impacted behavioral outcomes does not vary by provider.[54] This suggests that I am capturing the effect of attending a PT program more generally and not just the effect of attending a PT program of a specific provider. As a result, the effects estimated

[54]For the other variables, there is also no systematic difference in the estimated coefficients between providers.

here are likely to also hold for PT programs offered by other providers.[55]

When performing multiple hypothesis tests simultaneously, there is a high chance that at least one of the estimated coefficients will be significant even if the null hypothesis holds for all variables. One way of addressing this problem is the Benjamini-Hochberg method (Benjamini and Hochberg (1995)). There is no standard way to implement this method in this paper, as I have variables covering multiple domains (academic and behavioral), variables that are available for different subsamples of students[56] and variables that are heavily correlated[57]. One commonly used way to implement this method however is to consider the behavioral and academic variables separately (see What Works Clearinghouse (2014)). The academic variables will all remain insignificant as they were insignificant in the baseline specification. A conservative way to look at the behavioral variables is to consider them all at once, even though some of the variables are heavily correlated. When doing so, and using the p-values reported in column (1) of the robustness tables, student's level of responsibility for learning and student's assignment completion rate are both statistically significant at the 12% level. Student's level of preparedness for class is only significant at the 22% level. Conservatively accounting for multiplicity thus suggests that student's level of responsibility and student's assignment completion rate are affected by PT attendance, but does cast some doubt on the impact of PT on student's level of preparedness for class.

Thus far, all regressions have included all individuals who are enrolled in a school in SDUSD for the entire school year. Some of these individuals do not attend the school of the PrimeTime application for the entire year. As a result, the causal impact of attending a PrimeTime program identified in this

[55]In addition to the above robustness checks, I also estimated a specification in which I excluded individuals who appeared in the sample multiple times in a given year. Individuals can appear in the sample multiple times in a given year if they apply to PrimeTime programs at multiple schools in a given year. This additional restriction decreased the sample size on average by less than 1% and therefore did not significantly affect the baseline results.

[56]For instance, student's level of interest is available for students in all years in K-5 whilst student's feeling of responsibility for learning is only available for students in K-5 in 2014-2015.

[57]For instance, the correlation between student's level of preparedness for class and student's assignment completion rate is 0.76

paper could potentially also be partially attributable to attending a school that is different from the school of the application. In section 1.5.1, I showed that the probability of attending the school of the PrimeTime application for the entire year did not vary discontinuously across the threshold for one of the two providers. As a result, I can restrict the sample even further for this provider by dropping students who do not attend the school of the PrimeTime application for the entire school year without risking biasing my results. In appendix table 1.27, I show that the main results hold once I drop these students and only use data from this provider. If anything, as is expected due to the increased precision, the size of the estimated effects increases due to this additional sample restriction.[58] This highlights that the causal effects estimated in this paper are due to the PrimeTime program and not due to waitlisted individuals attending different schools than accepted individuals.

Lastly, I also estimate local linear regressions. These regressions avoid the problem of identifying local effects using variation far away from the cutoff. In particular, for each outcome variable I estimate local linear regressions with bandwidths of varying size, including the optimal bandwidth for a fuzzy RD design specified in Calonico et al. (2016).[59] The results of these regressions are displayed in table 1.7 for the three behavioral variables that were impacted by PrimeTime attendance in the baseline specification and in appendix tables 1.28 through 1.43 for the other variables. Regardless of the choice of bandwidth, the results are in line with the baseline regression results discussed earlier. None of the academic outcomes are affected by PT attendance and the behavioral outcomes that are impacted are those that were found to be impacted under the baseline specification as well. In fact, the estimated impact

[58]When using data from only one provider, the number of school by year combinations is often low. P-values of the 2SLS coefficient based on a Wild bootstrap procedure are therefore shown in column (6) of the table. The three behavioral variables that were significant before have p-values ranging from around 0.05 to 0.12. The only coefficient that is very different from the baseline results for provider 2 is the coefficient on SBRC Reading scores in 2014-2015. This coefficient is now statistically significant, whereas it was not statistically significant before. The other 10 academic outcomes are still not affected significantly though, so it is hard to attach too much weight to this one coefficient.

[59]These local linear regressions use a uniform kernel and do not include school by year fixed effects or control variables.

on these outcomes is slightly larger using local linear regressions than using the baseline regression specification.

1.5.5 Heterogeneous Impacts

In this section, I analyze whether the effect of PT attendance varies across subgroups of the student population. In particular, I estimate the baseline specification separately for male and female students, for students from a single-parent household and students from a two-parent household, and for students with a parent with a college degree and students without a parent with a college degree. Among other things, the alternative to PrimeTime might vary across these subgroups. For instance, students from a single-parent household might be more likely to have to care for themselves after school if PrimeTime is not available. Similarly, students without a parent with a college degree, an indicator of socioeconomic status, might not be able to attend paid enrichment after-school activities if not attending PrimeTime. Different counterfactuals to PrimeTime could affect the impact that PrimeTime has on student outcomes. In addition, when possible, I also estimate the baseline specification separately for students with a lagged outcome variable above and below the in-sample median. Students performing below the median in the previous year might stand to gain more from the academic component of PrimeTime and the emphasis that PrimeTime places on encouraging positive behavior.

The heterogeneity analyses are presented in tables 1.8 through 1.10 for the three behavioral variables that were impacted by PrimeTime attendance in the baseline specification and in appendix tables 1.44 through 1.59 for the other variables. Looking first at the academic outcomes, a few results stand out. First, the lack of an effect on academic outcomes does not depend on whether or not a student's parent has a college degree or on the student's lagged value of the outcome variable. None of the coefficients on the academic outcomes are statistically significant when looking separately at students with lagged outcome values above and below the median. Similarly, both for students with and for students without a parent with a college degree only 1 out of 11 coefficients is statistically significant at the 5% level, a result that could be purely due to chance.

There is some suggestive evidence that girls' academic outcomes improve by attending PrimeTime, whilst boys' academic outcomes worsen due to PrimeTime attendance. In particular, attending PrimeTime for 80 more days per year causes standardized test scores in ELA to increase by 0.18 standard deviations for girls, an effect that is statistically significant at the 5% level, but does not affect boys' scores. At the same time, attending PrimeTime for 80 more days per year causes standardized test scores in math to decrease by 0.22 standard deviations for boys, an effect that is statistically significant at the 5% level, but does not affect girls' scores. These patterns appear the same when looking at report card grades in math, reading and writing in K-5, but in these cases the coefficients are not statistically significant. However, these patterns do not hold for middle school GPAs. In fact, girls' math GPA is estimated to decrease by 0.20 standard deviations in response to 80 more days of PrimeTime attendance, an effect that is marginally significant at the 10% level.

In addition, I also find suggestive evidence that PT attendance causes the academic outcomes of students from single-parent households to worsen, but does not affect the outcomes of students from two-parent households. For instance, 80 more days of PrimeTime attendance is estimated to decrease middle school ELA GPA by 0.37 standard deviations and overall middle school GPA by 0.45 standard deviations for students from a single-parent household. Also, standardized test scores in math are estimated to decrease by 0.30 standard deviations as a result of 80 more days of PrimeTime attendance for students from a single-parent household. All these effects are statistically significant at the 10% level. This pattern is however not present for report card grades in math, reading and writing in K-5, as none of these outcomes are significantly affected by PT attendance for students from a single-parent household.

When looking at the behavioral outcomes, a more consistent pattern emerges. I find little evidence of differential effects on behavioral outcomes across subgroups. The behavioral report card grades that were not impacted by PrimeTime in the baseline analysis are also not impacted when looking at particular subgroups of students. Similarly, as can be seen in tables 1.8 through 1.10, the behavioral report card grades that were significantly affected by PrimeTime are in most cases also significantly affected when looking at subgroups of students. For all three behavioral variables, the sign of the coefficient for a particular subgroup is always the same as the overall baseline coefficient. The precision of the coefficients varies across subgroups, but the magnitudes are generally close to the baseline coefficient. The only clear exception occurs when looking at the coefficients on students' level of preparedness for class and students' assignment completion rate for students whose lagged value is below the median. In this case, the estimated coefficients are two to three times larger in magnitude than the baseline coefficient. Lastly, looking at the percentage of school days that students are absent in appendix table 1.55, we find some evidence that 80 more days of PT attendance decreases this percentage by 0.74 points for girls and by 0.71 point for students whose parents do not have a college degree. This corresponds to around 1.4 more days of regular school day attendance per school year.

Overall, looking across behavioral and academic outcomes, I find some suggestive evidence that in K-5 girls' outcomes improve from attending PrimeTime, whilst they worsen for boys. In addition, I also find suggestive evidence that academic outcomes worsen for students from single-parent households and that behavioral outcomes are more negatively impacted for students with lagged outcome values below the median. These last two results suggest that students at the cutoff who could potentially gain the most from PrimeTime attendance, as they have lower baseline scores or are less likely to have a parent at home during the after-school hours, in fact seem to gain the least from attending PrimeTime.[60]

[60]As I mentioned in the introduction, the observed negative coefficient on some of the behavioral variables could be due to students being exposed to more low-performing and possibly disruptive peers in PrimeTime than outside the program. To naively look at the importance of such peer effects, one can estimate the baseline regressions separately for two groups of schools. In the first group of schools, students attending the program are more disadvantaged than students not attending the program. In the second group of

1.5.6 Impact on Student Behavior

In the above analysis, I found that PrimeTime attendance impacted students' assignment completion rate and level of preparedness for class negatively and students' feeling of responsibility for learning positively. In this section, I try to better understand what is driving these apparently contradictory results.

The variable measuring students' feeling of responsibility for learning is available only for 2014-2015, whilst the other two variables are only available up to 2013-2014. The apparently contradictory effects of PrimeTime on these variables could therefore be the result of differences in the schools and students in my sample in each year. To look at the effect of differences in schools, I re-estimated the baseline specification for students' assignment completion rate and level of preparedness for class using only data from the schools that are also present in my sample in 2014-2015. This leaves both coefficients virtually unaffected. To look at the effect of differences in students across years, it is important to note that I cannot test whether differences in unobservable student characteristics can explain the above results. To check whether differences in observable student characteristics matter, I could first calculate how the compliers - PrimeTime applicants for whom days of attendance is affected by the initial admission decision - in 2014-2015 differ from those in earlier years in terms of observable characteristics (see Angrist and Pischke (2008)). Such an analysis would for instance tell me what share of compliers in each year come from a single-parent household or have a parent with a college degree. Using the subgroup-specific treatment effects reported in table **??**, I could then cal-

schools, students attending the program are less disadvantaged. Doing so using the share of students who have a parent with a college degree as a measure of advantage, I find no difference between the estimated impact of PT across the two subgroups of schools. This suggests that peer effects are not important. However, the composition of students in a particular PrimeTime program relative to other students in the same school is endogenous. For instance, a very well-run program is likely to be highly popular with students. This popularity could lead to a high cutoff priority score, meaning that students in the program will appear relatively more disadvantaged when compared to students not attending the program. In this case, the potential negative peer effects would be offset by the positive effect of attending a very well-run PT program. To isolate the peer effects, one needs exogenous variation in the composition of students in the program, something that I lack in this study.

culate a weighted average treatment effect for students' feeling of responsibility for learning where the weights would reflect the composition of compliers in the earlier years (see Angrist and Fernandez-Val (2013)). That is, knowing for instance what share of compliers are from single-parent households in earlier years, I could weight the estimated coefficients reported in table **??** to get one coefficient on students' feeling of responsibility for learning that is not affected by changes over time in the share of compliers that are from single-parent households.[61] For each subgroup of the population that I looked at, the coefficient on students' feeling of responsibility for learning was positive. Hence, the new weighted average treatment effect would also be positive. Similarly, the subgroup-specific coefficients on students' assignment completion rate and level of preparedness for class are all negative. Correcting the overall coefficient on these variables to take into account changes in observable student characteristics over time would thus also not change the sign of these coefficients. That is, changes in observable student characteristics over time cannot explain why the coefficients on the behavioral variables are of opposing signs in different years.

An alternative explanation is that the various variables are not comparable and measure different dimensions of student behavior. If different dimensions of student behavior are impacted differently by PrimeTime attendance, the above results would not actually contradict each other. Although I cannot test this directly, looking at the correlations between the variables suggests that if all variables were to be available in the same year, we would expect them all to be impacted by PrimeTime in the same direction. The correlation between students' feeling of responsibility for learning and these students' lagged 2013-2014 assignment completion rate and level of preparedness for class is respectively 0.48 and 0.44. This is very close to the correlation

[61]The small number of observations in 2014-2015 means that I can only take one observable student characteristics into account when calculating a new weighted coefficient for students' feeling of responsibility for learning. Taking into account more characteristics at the same time would require me to calculate heterogeneous treatment effects based on even smaller subgroups of the 2014-2015 student population. Given that the first-stage is now already weak in some cases when estimating heterogeneous treatment effects, this would not be possible to do accurately.

between students' assignment completion rate and its own one-year lag of 0.44 and to the correlation between students' level of preparedness for class and its own one-year lag of 0.45. This suggests that students' level of feeling of responsibility for learning is closely related to these other two variables.

In addition, two other behavioral variables, students' level of respect for others and students' level of interest in learning, are available in all years. In figure 1.6 I show the coefficient and corresponding confidence intervals on these variables when estimating the baseline equation separately for the years up to 2013-2014 and for the year 2014-2015. The coefficients are negative up to 2013-2014 and positive in 2014-2015, with the difference between the two coefficients being statistically significant for students' level of interest in learning. The correlation between these two behavioral variables that are available in all years and the behavioral variables that are available for a subset of years is around 0.5. Hence, the opposing coefficients on students' level of interest and respect for others over time suggest that the apparently contradictory results are due to factors that affect the coefficients on all behavioral variables analyzed here more broadly and cannot be explained by changes in the definition of particular variables.

Overal, I find evidence suggesting that the opposing signs on the behavioral variables that were significantly impacted by PrimeTime attendance in the baseline analysis cannot be explained by changes in the schools or observable characteristics of students in my sample over time, or changes in the behavioral variables over time. This suggests that changes in the unobservable characteristics of students in my sample over time or changes in the operation of PrimeTime at the schools in my sample that were not the result of explicit policy changes[62] are responsible for the opposing signs on the behavioral variables. Differences in unobservable characteristics of students at the cutoff over time could for instance arise if PrimeTime staff members set the application cutoff date endogenously in some years to include applicants that needed help

[62]For instance, it is possible that the staff members of the PrimeTime programs changed over time. Huang and Wang (2012) mention that high staff turnover is a large problem at after-school programs. Unfortunately, I do not have data on the staff members of each PrimeTime program at each school.

in improving their behavioral outcomes. The fact that the negative coefficients are sensitive to using a 1-day donut RD is in line with this hypothesis. Importantly however, the changes in the unobservable characteristics of students in my sample over time or changes in the operation of PrimeTime that were not the result of explicit policy changes did not seem to have affected how PrimeTime affects academic outcomes. As mentioned before and as can be seen in table 1.3, just as in earlier years PrimeTime attendance has no effect on report card grades in math, reading and writing in K-5 in 2014-2015.

1.5.7 Comparison to previous literature

The results of this paper are in line with previous studies on after-school programs. The RCTs on after-schools programs in elementary schools (see James-Burdumy et al.(2005, 2007 and 2008)) and middle schools (see Gottfredson et al.(2010a, 2010b)) also found that after-school program attendance did not impact academic outcomes. In addition, they also found that student behavior was unaffected in middle school, but negatively affected in elementary schools, with participating students in elementary schools displaying lower levels of effort in class as judged by their teachers. Students in my study participated in the after-school program for more days than in the RCTs. The results of this paper therefore suggest that the intensity of program participation is not related to student outcomes. However, the students in this study seem to come from less disadvantaged backgrounds than in the RCTs. It is therefore still possible that a very high level of participation in after-school programs is beneficial for students with disadvantaged backgrounds.

The results of this paper are also in line with some of the literature on the effects on child development of large-scale publicly subsidized child care programs for younger children that find no effect on children's cognitive development and negative effects on non-cognitive development (e.g. Baker et al. (2008)) . On the other hand, a large literature on targeted intensive pre-school and child care programs for disadvantaged children finds that these programs can have positive effects on children's academic and behavioral outcomes (e.g.

Barnett (1995), Karoly et al. (1998) and Garces et al. (2002)). Again, this suggests that after-school programs could be beneficial for student outcomes when targeted at the right population.[63]

Lastly, the results of this paper are not surprising given the large literature on peer effects of low-achieving and disruptive students that finds that these students negatively impact the academic performance and behavior of average performing students (e.g. Gaviria and Raphael (2001), Figlio (2007), Carrell and Hoekstra (2010), Lavy et al. (2012a) and Lavy et al. (2012b).). As was shown in table 3.1, baseline behavioral and academic outcomes for students in the cutoff priority score group are on average 0.2 to 0.4 standard deviations higher than for students in higher priority score groups, who are admitted regardless of the postmark date of their application and make up the bulk of program participants.[64] For students in the cutoff priority score group, participating in the PrimeTime program could thus mean being exposed to more low-performing and possibly disruptive peers than outside the program. The negative peer effects from these students could limit the benefits that students in my sample derive from program participation and explain why I fail to find consistent evidence of positive impacts of program participation. In addition, the finding that boys academic outcomes worsened due to PrimeTime attendance whilst they improved for girls is also in line with the above-mentioned studies on peer effects, as these generally find that boys are more negatively impacted by low-performing or disruptive peers than girls.

[63]In addition, Aizer (2004) finds that being supervised by an adult after school, either at home or at any other location such as a child care center, has a positive impact on the behavior of children aged 10-14. The lack of clear positive findings in this study therefore seem to indicate that students in my study who are not admitted to PrimeTime are not left to take care of themselves after school and do have access to other forms of adult supervision. Again, positive effects might still be expected for students from more disadvantaged backgrounds without access to adult supervision after school. However, it is unclear whether after-school programs are necessary for such children to improve their outcomes. If positive effects are just due to being supervised, which is what the results of Aizer (2004) combined with the results of this paper suggest, general child care programs that do not provide enrichment activities could be as beneficial to these children as after-school programs such as PrimeTime.

[64]In fact, students in the cutoff group have higher baseline outcomes than their average district peer whereas students in higher priority score groups perform worse than the average district student.

1.6 Conclusion

In this paper, I used a fuzzy RD design based on the admissions procedure to an after-school program to robustly estimate the effect that attending this after-school program has on students' short-run academic and behavioral outcomes. I find that students' academic outcomes are on average not affected by attending the after-school program. This result holds whether I look at standardized test scores or school grades in ELA and math for students in K-8. The fact that academic outcomes appear to be unaffected on average does not seem to be a product of the particular grades and years for which data are available, as the various academic outcomes are based on students across all different years and grades for which I have data. Although on average there seems to be no impact, looking at subgroups of students I find suggestive evidence that the after-school programs cause academic outcomes to improve for girls and worsen for boys and students from a single-parent household.

On the other hand, students' behavioral outcomes, as measured by several report card grades regarding student behavior in K-5, were significantly affected. In particular, attending the after-school program for 80 more days per year (the average change in days of attendance at the cutoff) caused students' assignment completion rate and preparedness for class to decrease by respectively 0.19 and 0.14 standard deviations, and increased students' feeling of responsibility for learning by 0.35 standard deviations. These results are robust to a variety of specifications, the only notable exception being the sensitivity of the negative effects to using a donut RD, and are not specific to a particular subgroup of students. I provided suggestive evidence that the difference in the direction of the effect of PrimeTime on student behavior across variables is attributable to either changes in the unobservable characteristics of students in my sample over time or changes in the operation of PrimeTime at the schools in my sample that were not the result of explicit policy changes. Overall, it is important to keep in mind that the RD design only informs us about the causal impact of the after-school program for the relatively advantaged students at the cutoff. The impact of the after-school program on the typical more disadvantaged participant not near the cutoff might be very different.

The results from this paper suggest that changing the capacity of PrimeTime after-school programs in schools with wait lists is unlikely to significantly impact the short-run academic and behavioral outcomes of the students affected by such a policy. This is important in light of the high costs of operating after-school programs such as PrimeTime. The California state government currently spends $7.50 per student per day on after-school programs such as PrimeTime, resulting in a maximum annual cost per student of $1350 for a student attending all possible 180 school days.[65] To put this into perspective, this is 11.9% of the $11,340 that was spent on average per pupil in San Diego Unified School District in 2014-2015 (California Department of Education (2016)).

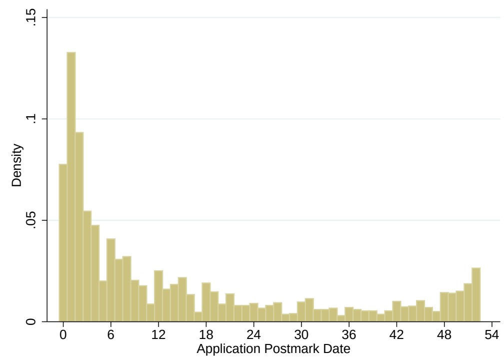
However, there are many potential effects of PrimeTime attendance that this paper cannot capture. Most importantly, by providing children a safe environment after school, PrimeTime is very likely to influence parents' labor supply decisions. Such labor supply decisions not only affect the economic well-being of these parents and their children, but also influence the amount of income taxes that the state and federal government receive. This change in income tax revenue could potentially offset the costs of PrimeTime. In addition, I only have data on a limited number of student outcomes. For example, I lack data on students' involvement in criminal activity. Removing students from the street during the after-school hours during which juvenile crime peaks is one of the main reasons behind funding after-school programs such as PrimeTime. I may thus be capturing only part of the impact of PrimeTime on students' short-run outcomes. Similarly, I am not able to study the long-run effect of participation in the PrimeTime program. By keeping children of the street during after-school hours and so reducing their exposure to

[65] The actual cost per student per year is likely to be lower as most students do not attend the full 180 days per year and some PrimeTime program admit more students than they formally receive funding for.

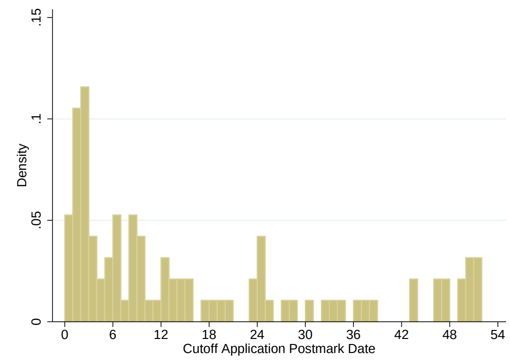
crime at a young age, it is very likely that PrimeTime decreases the likelihood that youth participate in crime at a later age (see Damm and Dustmann (2014)).

Further research on after-school programs is needed. To better understand the full impact of after-school programs such as PrimeTime on students and their parents, it would be useful to investigate the impact of PrimeTime on a wider variety of short- and long-run student outcomes and on the labor supply decisions of parents. In addition, more research is needed on the impact of after-school programs on the outcomes of students that are not near the admissions cutoff. As mentioned before, these students appeared to have a higher need for the program and the impact of the after-school program could therefore be very different for these students. Furthermore, to better understand the results of this study, additional information on the ways that students not attending PrimeTime spend their time after school would be beneficial. For instance, knowing whether these students spend this time engaged in academic activities would allow us to better judge the effectiveness of the after-school program in improving student academic outcomes. Also, in the absence of information on other student outcomes and parental labor supply, knowing about the counter-factual would allow us to better understand how likely it is that these other outcomes are affected by after-school programs. Lastly, after-school programs consist of many different components, and their administration is likely to differ widely across schools. Understanding better how after-school programs are administered at individual schools and comparing the impact of the after-school programs based on these characteristics could help in understanding the mechanisms driving the estimated impact and in designing effective after-school programs that improve student outcomes.

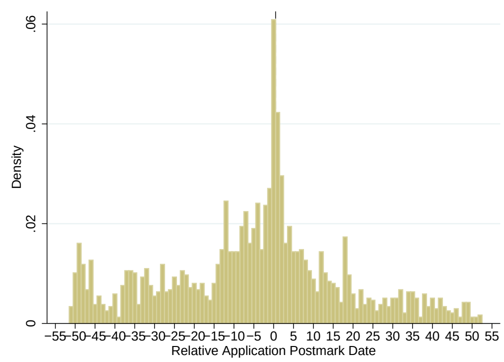
Chapter 1 is currently being prepared for submission for publication of the material. The dissertation author was the sole author of this paper. Gaastra, Sieuwerd.



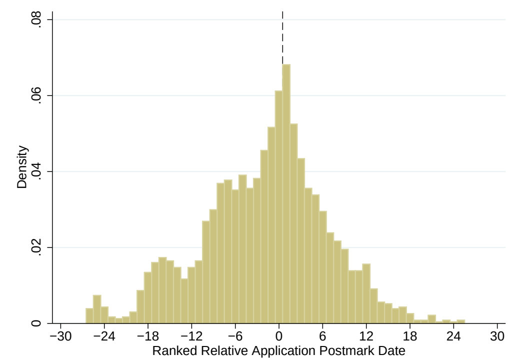
(a) Distribution of All Application Postmark Dates



(b) Distribution of Cutoff Application Postmark Dates



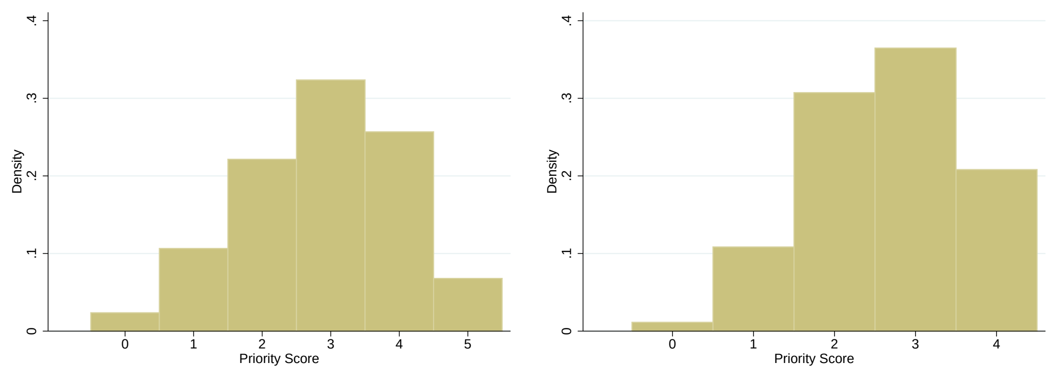
(c) Distribution of Relative Application Postmark Dates



(d) Distribution of Ranked Relative Application Postmark Dates

Figure 1.1: Distribution of Application Postmark Dates

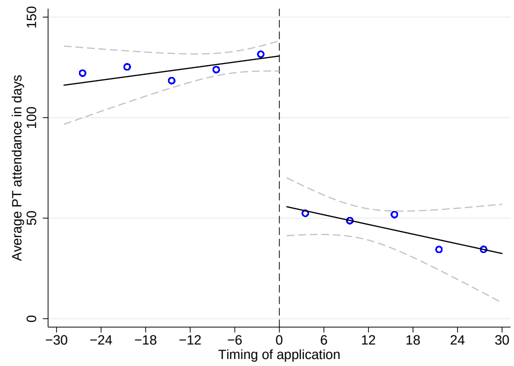
Notes: Figure 1.1a shows the distribution of postmark dates of all applicants with a priority score equal to the cutoff priority score in schools and years that had to waitlist applicants. Figure 1.1b shows the distribution of cutoff application postmark dates in these schools and years. In these figures, postmark dates are relative to the start of the initial application timeline (day 0). Figure 1.1c shows the distribution of postmark dates relative to the cutoff date in schools and years in my sample that did not waitlist any applicants at the cutoff. Since applications cannot be postmarked on Sundays, there are only 53 possible postmark dates during the 61-day initial application timeline. For the purposes of these graphs and the data analysis in this paper, applications postmarked on a Saturday and on the first Monday following this Saturday are considered to have a postmark date that is 1 day apart. Figure 1.1d shows the distribution of "ranked" relative application postmark dates in which days on which no applications were submitted are dropped.



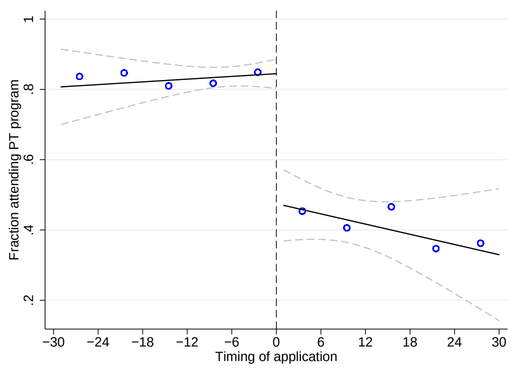
(a) 2010-2011 to 2012-2013 (b) 2013-2014 and 2014-2015

Figure 1.2: Distribution of Priority Scores

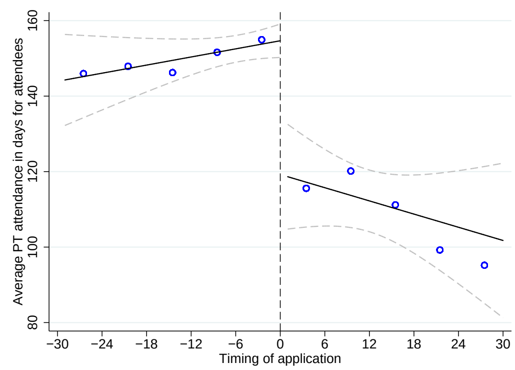
Notes: This figure shows the distribution of priority scores of all applications submitted during the initial application timeline to schools in my sample. Priority scores ranged from 0 to 5 up to 2012-2013 and from 0 to 4 afterwards.



(a) Average PT Attendance in Days



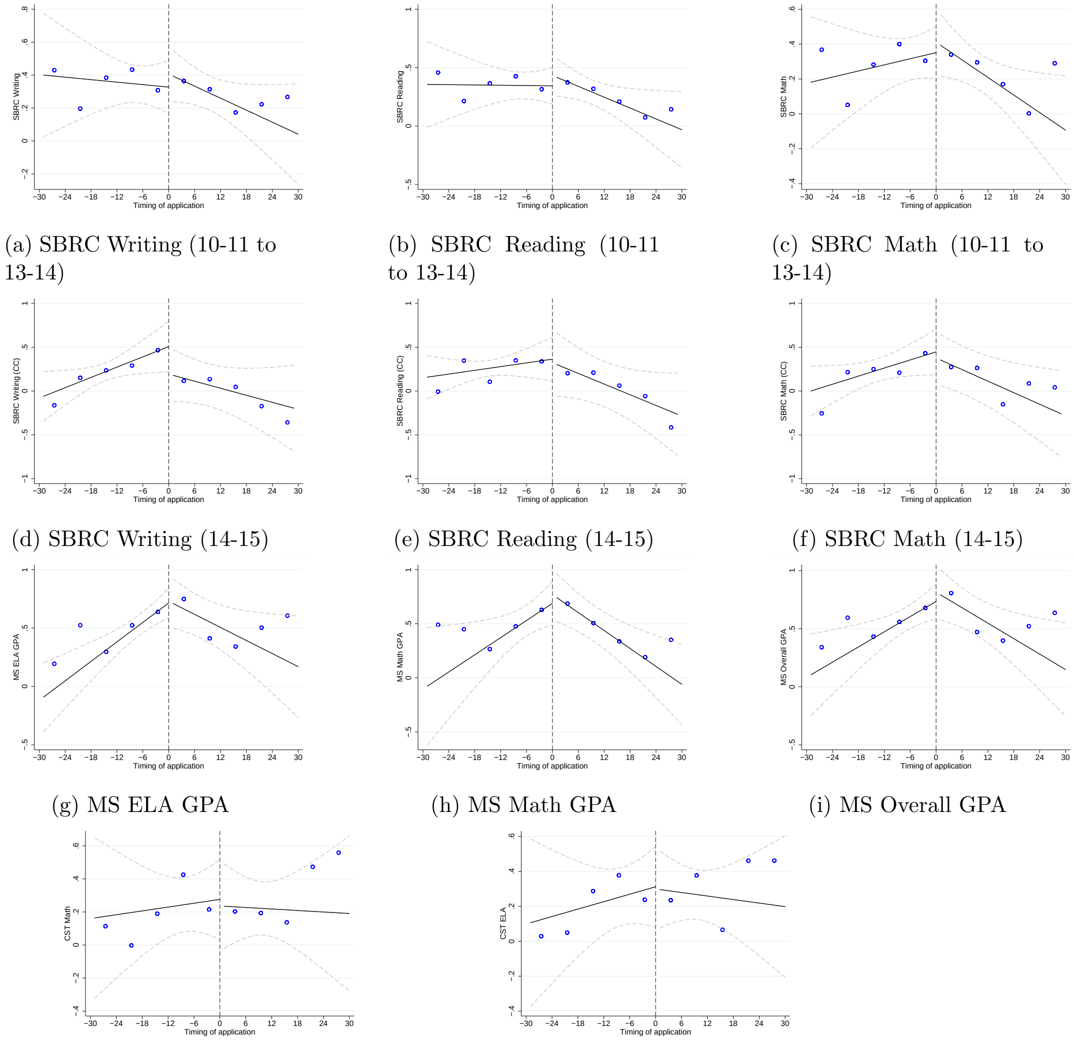
(b) Fraction attending PT program



(c) Average Attendance for Attendees

Figure 1.3: Effect of Timing of Application on PT Attendance

Notes: Each observation is the unweighted average of the outcome variable in a 6-day bin, where the bin is based on the relative postmark date of individuals' applications. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using triangular weights and separate linear trends on either side of the cutoff. The dashed lines surrounding these estimated regression lines mark the 95% confidence intervals.



(a) SBRC Writing (10-11 to 13-14)

(b) SBRC Reading (10-11 to 13-14)

(c) SBRC Math (10-11 to 13-14)

(d) SBRC Writing (14-15)

(e) SBRC Reading (14-15)

(f) SBRC Math (14-15)

(g) MS ELA GPA

(h) MS Math GPA

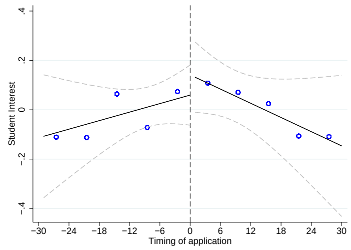
(i) MS Overall GPA

(j) CST Math Score

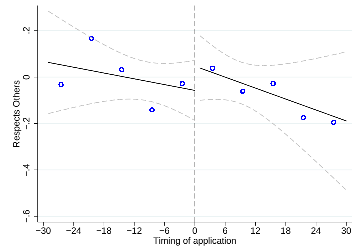
(k) CST ELA Score

Figure 1.4: Academic Outcomes

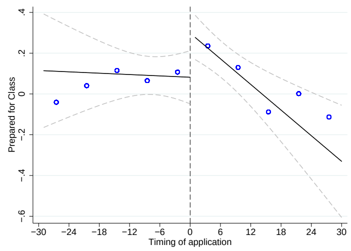
Notes: Each observation is the unweighted average of the outcome variable in a 6-day bin, where the bin is based on the relative postmark date of individuals' applications. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using triangular weights and separate linear trends on either side of the cutoff. The dashed lines surrounding these estimated regression lines mark the 95% confidence intervals. The graphs are not all based on the same sample of students as different variables are available for different grades and years.



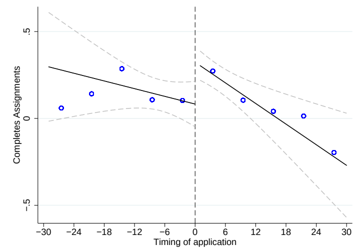
(a) Student shows interest in learning



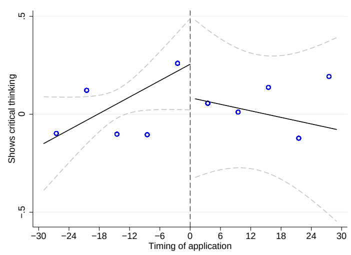
(b) Student respects others



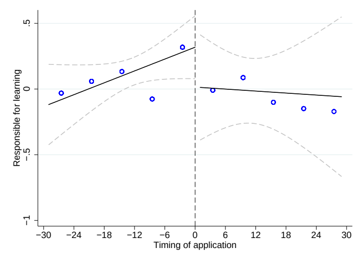
(c) Student prepares and organizes



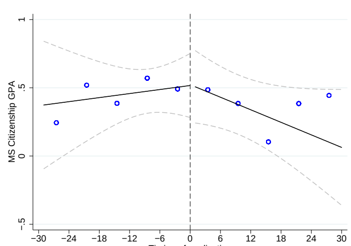
(d) Student completes assignments when due



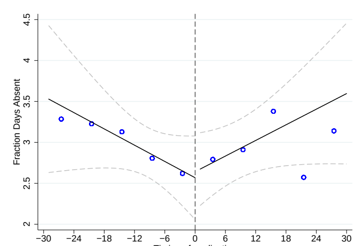
(e) Student demonstrates critical thinking



(f) Student takes responsibility for learning



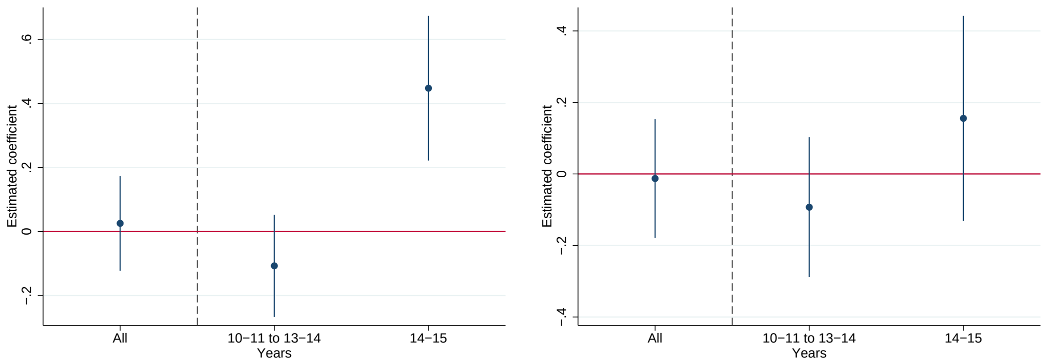
(g) MS Citizenship GPA



(h) Percentage of school days absent

Figure 1.5: Behavioral Outcomes

Notes: Each observation is the unweighted average of the outcome variable in a 6-day bin, where the bin is based on the relative postmark date of individuals' applications. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using triangular weights and separate linear trends on either side of the cutoff. The dashed lines surrounding these estimated regression lines mark the 95% confidence intervals. The graphs are not all based on the same sample of students as different variables are available for different grades and years.



(a) Student shows interest in learning (b) Student respects others

Figure 1.6: Impact of PT on Selected Behavioral Outcomes over Time

Notes: The above graphs show the 2SLS coefficients, multiplied by 80, that are obtained by estimating equation 1.2 using data only from the years that are listed on the x-axis. Other than this restriction, the sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 1.3. Standard errors are clustered at the school by year level. 95% confidence intervals for the estimated coefficients are indicated by the blue lines.

Table 1.1: Descriptive Statistics

	Cutoff Group	Higher Priority Score	
CST Math (t-1)	0.32	-0.067	0.38^{***}
	(0.030)	(0.015)	(0.033)
CST ELA (t-1)	0.28	-0.057	0.34^{***}
	(0.028)	(0.015)	(0.032)
Interest Level (t-1)	0.21	-0.11	0.32^{***}
	(0.018)	(0.015)	(0.024)
Respects Others (t-1)	0.062	-0.17	0.23^{***}
	(0.022)	(0.016)	(0.027)
Completes Assignments (t-1)	0.20	-0.089	0.29^{***}
	(0.019)	(0.015)	(0.024)
Prepared for Class (t-1)	0.19	-0.097	0.29^{***}
	(0.020)	(0.015)	(0.025)
Academic Assistance Status	0.18	0.49	-0.30^{***}
	(0.0071)	(0.0057)	(0.0091)
Participated in PT in t-1	0.59	0.78	-0.20^{***}
	(0.0090)	(0.0047)	(0.010)
Single Parent HH	0.44	0.68	-0.24^{***}
	(0.0091)	(0.0053)	(0.011)
Parents FT Employed	0.85	0.93	-0.076^{***}
	(0.0065)	(0.0029)	(0.0071)
Parent has BA	0.53	0.34	0.19^{***}
	(0.0097)	(0.0058)	(0.011)
English Learner	0.24	0.33	-0.085^{***}
	(0.0078)	(0.0054)	(0.0095)
Female	0.51	0.50	0.014
	(0.0092)	(0.0057)	(0.011)
White	0.24	0.17	0.077^{***}
	(0.0079)	(0.0043)	(0.0090)
Black	0.15	0.18	-0.030^{***}
	(0.0066)	(0.0044)	(0.0079)
Hispanic	0.35	0.47	-0.12^{***}
	(0.0088)	(0.0057)	(0.010)
Observations	2991	7674	

Notes: Summary statistics are based on data from two large providers for the years 2010-2011 to 2014-2015. The averages for outcome variables are based on students who attended a school in SDUSD. Standard errors of the mean are reported in parentheses. The fourth column shows results from two-sample t-tests that test whether the mean of the relevant variable is the same for both groups. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.2: Balance of predetermined student characteristics at the threshold

(a) Background Characteristics

	(1) $1[t \leq c]$	(2) P-value	(3) N
Applied to Before School PT	0.024	0.54	2161
	(0.038)		
Female	0.0062	0.86	2146
	(0.035)		
Age	-0.040	0.71	2156
	(0.11)		
Grade	-0.035	0.73	2161
	(0.10)		
Hispanic	0.059	0.20	2135
	(0.046)		
Parent has BA	-0.058	0.15	1843
	(0.040)		
English Learner	0.045	0.16	2149
	(0.032)		
Fluent English Proficient	0.0024	0.92	2149
	(0.025)		
Special Education Student	-0.016	0.47	1973
	(0.023)		
Academic Assistance Status	-0.029	0.34	2154
	(0.030)		
Parents FT Employed	-0.0087	0.78	2154
	(0.030)		
Single Parent HH	-0.065	0.16	2154
	(0.046)		
Participated in PT in t-1	0.054	0.18	2154
	(0.040)		
Enrolled in SDUSD in t-1	-0.020	0.38	2161
	(0.023)		

(b) Lagged Outcomes

	(1) $1[t \leq c]$	(2) P-value	(3) N
Fraction Days Absent	0.0015	0.67	1916
	(0.0034)		
Completes Assignments	-0.039	0.59	1401
	(0.072)		
Prepared for Class	0.013	0.81	1401
	(0.055)		
Interest Level	0.056	0.47	1401
	(0.077)		
Respects Others	0.070	0.46	1401
	(0.094)		
SBRC Math	0.096	0.25	1572
	(0.083)		
SBRC Reading	0.077	0.42	1534
	(0.096)		
SBRC Writing	0.030	0.73	1535
	(0.088)		
CST Math	0.075	0.58	829
	(0.14)		
CST ELA	-0.0038	0.98	831
	(0.13)		
MS Math GPA	-0.18	0.30	201
	(0.17)		
MS ELA GPA	-0.012	0.93	201
	(0.14)		
MS Overall GPA	-0.14	0.19	201
	(0.10)		
MS Citizenship GPA	0.21	0.29	201
	(0.19)		

Notes: Column (1) reports the coefficient on $1[t \leq c]$ that is obtained by estimating equation 1.3 with the variable in the first column as the dependent variable. Each regression uses all applications submitted within a 5-week window of the cutoff application postmark date for which the outcome variable is not missing and which have a priority score equal to the cutoff priority score. Each regression includes separate linear trends in the postmark date relative to the cutoff on either side of the cutoff and school by year fixed effects. Each regression also employs triangular weights. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The p-value corresponding to each estimated coefficient is reported in column (2). The number of observations is reported in column (3) and varies across regressions as some variables are only available for a subsample of the observations. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.3: Baseline Regression Results

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
Academic Outcomes:					
CST Math	84.9***	-0.0032	-0.000038	114.0	620
	(7.96)	(0.096)	(0.0010)		
CST ELA	85.3***	0.090	0.0010	122.4	619
	(7.72)	(0.100)	(0.0011)		
SBRC Math	78.3***	-0.027	-0.00034	125.1	1111
	(7.00)	(0.061)	(0.00073)		
SBRC Reading	78.2***	-0.027	-0.00034	119.7	1100
	(7.14)	(0.073)	(0.00087)		
SBRC Writing	78.0***	-0.0063	-0.000080	118.9	1100
	(7.15)	(0.080)	(0.00097)		
SBRC Math (14-15)	81.5***	0.052	0.00064	33.2	353
	(14.2)	(0.18)	(0.0020)		
SBRC Reading (14-15)	81.5***	-0.078	-0.00096	33.2	353
	(14.2)	(0.20)	(0.0022)		
SBRC Writing (14-15)	81.5***	0.17	0.0021	33.0	352
	(14.2)	(0.21)	(0.0023)		
MS Overall GPA	72.0***	-0.023	-0.00032	27.1	415
	(13.8)	(0.11)	(0.0015)		
MS ELA GPA	71.9***	-0.021	-0.00029	26.8	415
	(13.9)	(0.13)	(0.0016)		
MS Math GPA	72.2***	-0.066	-0.00092	26.7	410
	(14.0)	(0.13)	(0.0016)		
Behavioral Outcomes:					
Fraction Days Absent	76.2***	-0.10	-0.0013	148.8	1978
	(6.25)	(0.25)	(0.0031)		
Student Interest	79.1***	0.025	0.00032	150.0	1473
	(6.46)	(0.078)	(0.00095)		
Respects Others	79.1***	-0.013	-0.00016	151.0	1473
	(6.44)	(0.088)	(0.0011)		
Prepared for Class	78.1***	-0.14*	-0.0018*	123.6	1111
	(7.03)	(0.086)	(0.0010)		
Completes Assignments	78.1***	-0.18**	-0.0023**	122.6	1111
	(7.06)	(0.080)	(0.00094)		
Shows critical thinking	83.0***	0.18	0.0022	34.1	362
	(14.2)	(0.18)	(0.0020)		
Responsible for learning	83.0***	0.35*	0.0042**	34.1	362
	(14.2)	(0.18)	(0.0017)		
MS Citizenship GPA	73.1***	-0.051	-0.00069	27.0	415
	(14.1)	(0.099)	(0.0012)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1.1 and the reduced form equation 1.3 for the variable given in the first column. Column (3) reports the 2SLS coefficient from equation 1.2. Each regression uses all applications submitted within a 5-week window of the cutoff application postmark date for which the outcome variable is not missing and which have a priority score equal to the cutoff priority score. Each regression includes separate linear trends in the postmark date relative to the cutoff on either side of the cutoff and school by year fixed effects. Each regression also controls flexibly for the student's gender, age, grade, ethnicity, English learner status, Special Education status and parental level of education using various indicators. Indicators are also included for each of the 4 criteria on which a student's priority score is based and for whether a student applied to a PrimeTime program as well. When available, the lagged value of the outcome variable is also included as a control variable. Each regression employs triangular weights. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5) and varies across regressions as some variables are only available for a subsample of the observations.* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.4: Robustness Checks - Completes Assignments when Due

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.19**	-0.26***	-0.11	-0.25***	-0.37***	-0.22***	-0.12	-0.029	-0.27***	-0.21**	-0.21*	-0.16
	(0.075)	(0.092)	(0.081)	(0.080)	(0.13)	(0.075)	(0.077)	(0.11)	(0.097)	(0.094)	(0.12)	(0.11)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	55	55	55	55	55	54	55	53	53	43	30	25
First-stage F stat.	122.6	47.3	101.8	62.1	18.8	115.0	111.9	67.7	88.9	110.1	91.9	47.5
P-value Wild BS	0.016	0.0020	0.13	0.0020	0.0060	0.012	0.12	0.43	0.0020	0.086	0.094	0.29
Observations	1111	1111	1111	1111	1111	919	1317	944	680	758	515	596

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1.1 and 1.2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.5: Robustness Checks - Prepared for Class

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.14*	-0.24**	-0.074	-0.19*	-0.21	-0.16*	-0.098	-0.053	-0.17	-0.17*	-0.13	-0.13
	(0.082)	(0.094)	(0.085)	(0.097)	(0.17)	(0.089)	(0.082)	(0.085)	(0.11)	(0.10)	(0.11)	(0.11)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	55	55	55	55	55	54	55	53	53	43	30	25
First-stage F stat.	123.6	47.3	102.6	62.5	18.8	116.5	113.7	69.1	93.2	110.5	93.3	47.4
P-value Wild BS	0.082	0.012	0.37	0.032	0.16	0.070	0.24	0.53	0.15	0.11	0.31	0.34
Observations	1111	1111	1111	1111	1111	919	1317	944	680	758	515	596

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1.1 and 1.2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.6: Robustness Checks - Takes Responsibility for Learning

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	0.34**	0.26*	0.29**	0.38**	0.32	0.33**	0.30**	0.25**	0.27**	0.39**	0.41**	0.37***
	(0.14)	(0.14)	(0.14)	(0.18)	(0.28)	(0.15)	(0.13)	(0.11)	(0.12)	(0.19)	(0.18)	(0.13)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	20	20	19	15	10	10
First-stage F stat.	34.1	35.9	47.4	14.4	12.6	18.1	40.4	41.6	27.0	16.0	16.2	10.9
P-value Wild BS	0.030	0.15	0.080	0.054	0.25	0.046	0.058	0.036	0.034	0.082	0.12	0.028
Observations	362	362	362	362	362	265	399	332	222	231	147	215

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1.1 and 1.2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.7: Local Linear Regressions

(a) Completes Assignments when Due

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	70.8***	-0.19**	-0.0027**	68.6	1111
	(8.55)	(0.082)	(0.0011)		
18 days	77.3***	-0.26***	-0.0034***	51.0	919
	(10.8)	(0.088)	(0.0013)		
12 days	74.5***	-0.26***	-0.0034**	27.5	732
	(14.2)	(0.097)	(0.0014)		
8 days	66.4***	-0.30***	-0.0045*	15.0	590
	(17.1)	(0.11)	(0.0023)		

(b) Prepared for Class

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	70.8***	-0.16**	-0.0022*	71.1	1111
	(8.40)	(0.081)	(0.0012)		
18 days	77.3***	-0.24***	-0.0031***	48.7	919
	(11.1)	(0.077)	(0.0012)		
12 days	74.5***	-0.23**	-0.0030**	27.5	732
	(14.2)	(0.10)	(0.0015)		
9 days	67.3***	-0.24*	-0.0036	17.2	635
	(16.3)	(0.13)	(0.0024)		

(c) Takes Responsibility for Learning

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	93.7***	0.28*	0.0030*	46.5	362
	(13.7)	(0.17)	(0.0016)		
18 days	101.1***	0.36*	0.0036*	27.5	265
	(19.3)	(0.22)	(0.0020)		
12 days	99.8***	0.44*	0.0044*	21.2	223
	(21.7)	(0.25)	(0.0025)		
11 days	101.7***	0.41	0.0040	21.4	211
	(22.0)	(0.28)	(0.0027)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.8: Heterogeneous Treatment Effects - Completes Assignments when Due

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	78.1***	-0.18**	-0.0023**	122.6	1111
	(7.06)	(0.080)	(0.00094)		
1. Male	73.0***	-0.21	-0.0028	55.1	547
	(9.84)	(0.16)	(0.0019)		
Female	83.4***	-0.12	-0.0015	71.1	564
	(9.89)	(0.11)	(0.0012)		
2. Single-parent HH	94.7***	-0.21	-0.0022	79.4	444
	(10.6)	(0.16)	(0.0015)		
Two-parent HH	67.5***	-0.21**	-0.0032**	43.3	660
	(10.3)	(0.10)	(0.0015)		
3. Parent has BA	70.8***	-0.071	-0.00100	72.1	488
	(8.33)	(0.11)	(0.0015)		
Parent has no BA	79.0***	-0.29**	-0.0037**	40.9	505
	(12.4)	(0.13)	(0.0016)		
4. Lag above median	71.4***	-0.17*	-0.0023**	50.8	444
	(10.0)	(0.094)	(0.0011)		
Lag below median	63.7***	-0.31	-0.0049*	14.9	356
	(16.5)	(0.22)	(0.0027)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1.1 and the reduced form equation 1.3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 1.2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 1.3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.9: Heterogeneous Treatment Effects - Prepared for Class

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	78.1***	-0.14*	-0.0018*	123.6	1111
	(7.03)	(0.086)	(0.0010)		
1. Male	73.5***	-0.17	-0.0023	57.5	547
	(9.70)	(0.16)	(0.0020)		
Female	83.4***	-0.13	-0.0015	70.3	564
	(9.95)	(0.11)	(0.0012)		
2. Single-parent HH	94.5***	-0.31*	-0.0033*	79.6	444
	(10.6)	(0.19)	(0.0017)		
Two-parent HH	67.4***	-0.13	-0.0019	43.6	660
	(10.2)	(0.12)	(0.0017)		
3. Parent has BA	71.6***	-0.12	-0.0016	78.2	488
	(8.09)	(0.13)	(0.0017)		
Parent has no BA	79.2***	-0.028	-0.00036	42.1	505
	(12.2)	(0.15)	(0.0017)		
4. Lag above median	70.5***	-0.12	-0.0017	37.5	459
	(11.5)	(0.10)	(0.0013)		
Lag below median	63.9***	-0.42*	-0.0066**	14.4	341
	(16.8)	(0.25)	(0.0033)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1.1 and the reduced form equation 1.3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 1.2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 1.3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

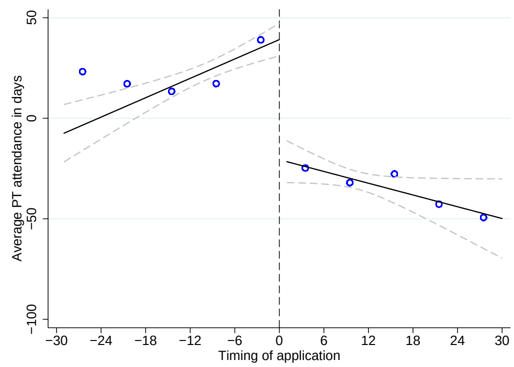
Table 1.10: Heterogeneous Treatment Effects - Takes Responsibility for Learning

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	83.0***	0.35*	0.0042**	34.1	362
	(14.2)	(0.18)	(0.0017)		
1. Male	66.5**	0.25	0.0037	6.35	176
	(26.4)	(0.28)	(0.0031)		
Female	101.3***	0.40**	0.0040***	26.8	186
	(19.6)	(0.16)	(0.0012)		
2. Single-parent HH	76.6***	0.28*	0.0036**	15.7	182
	(19.3)	(0.16)	(0.0017)		
Two-parent HH	68.5**	0.35*	0.0051***	6.12	180
	(27.7)	(0.19)	(0.0018)		
3. Parent has BA	85.1***	0.40*	0.0047**	17.0	177
	(20.6)	(0.22)	(0.0020)		
Parent has no BA	70.2***	0.30	0.0042*	18.8	166
	(16.2)	(0.20)	(0.0023)		

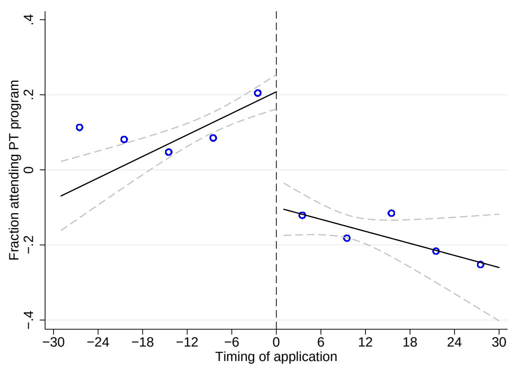
Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1.1 and the reduced form equation 1.3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 1.2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 1.3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

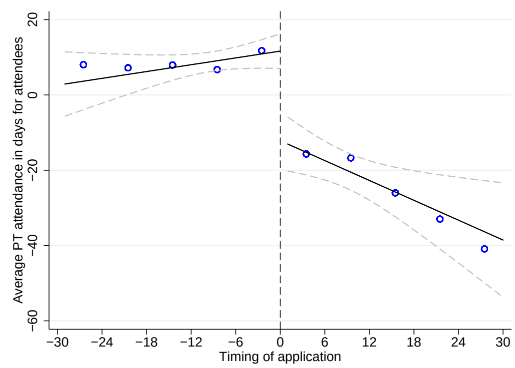
1.7 Appendix

1.7.1 Figures



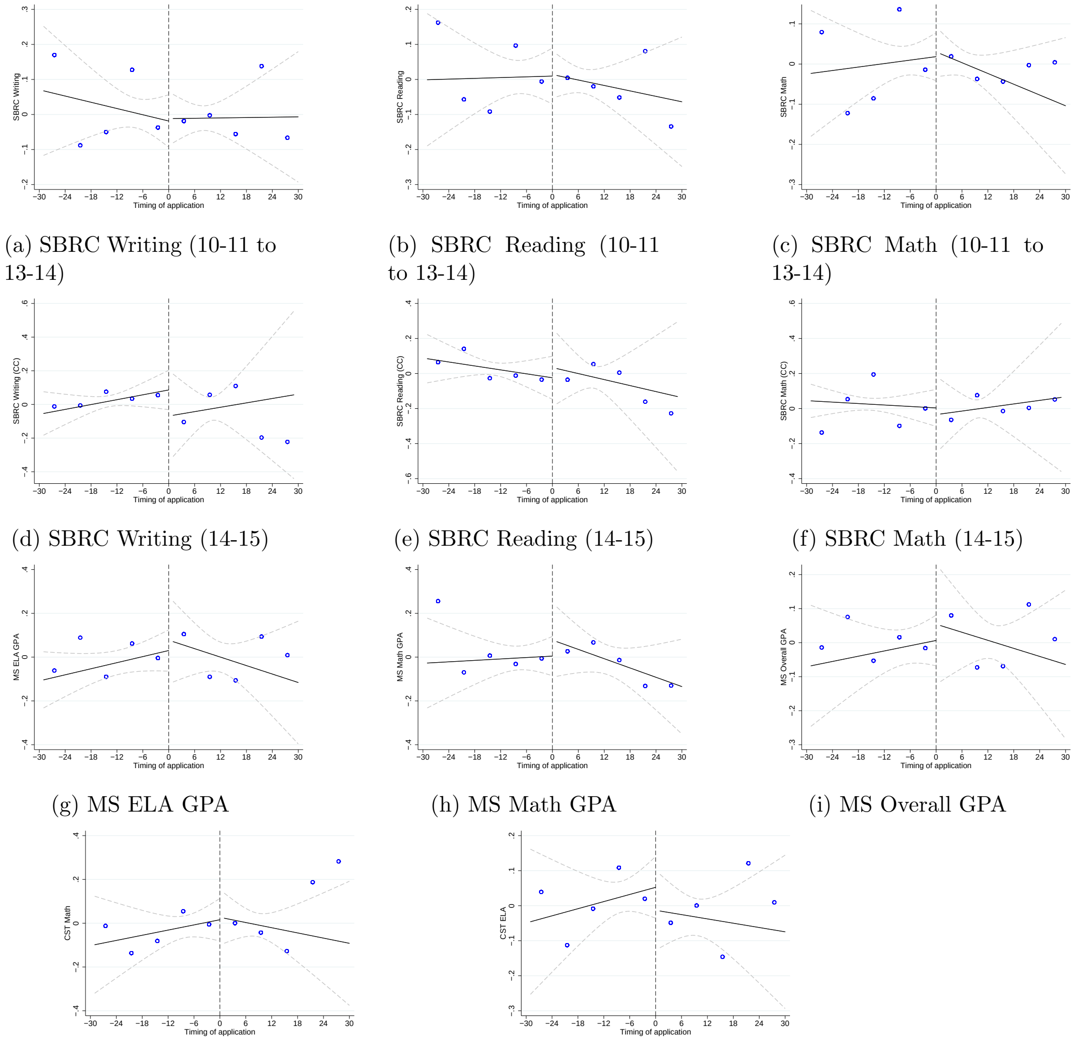
(a) Average PT Attendance in Days





(b) Fraction attending PT program (c) Average Attendance for Attendees

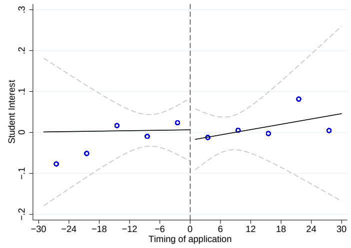
Figure 1.7: Residual Plot - Effect of Timing of Application on PT Attendance
Notes: Each observation is the unweighted average residual in a 6-day bin, where the bin is based on the relative postmark date of individuals' applications. The residuals are generated from a regression of the outcome variable on the predetermined control variables that are included in the baseline regression specification and school by year fixed effects. Triangular weights are not used when running this regression. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using triangular weights and separate linear trends on either side of the cutoff. The dashed lines surrounding these estimated regression lines mark the 95% confidence intervals.



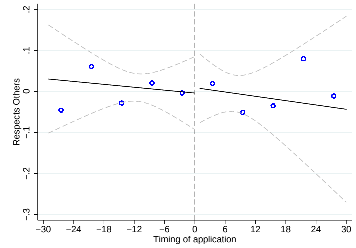
(a) SBRC Writing (10-11 to 13-14)
(b) SBRC Reading (10-11 to 13-14)
(c) SBRC Math (10-11 to 13-14)
(d) SBRC Writing (14-15)
(e) SBRC Reading (14-15)
(f) SBRC Math (14-15)
(g) MS ELA GPA
(h) MS Math GPA
(i) MS Overall GPA
(j) CST Math Score
(k) CST ELA Score

Figure 1.8: Residual Plot - Academic Outcomes

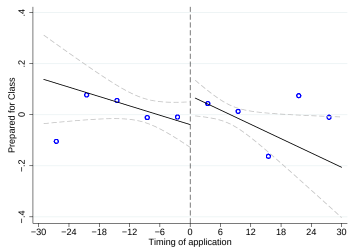
Notes: Each observation is the unweighted average residual in a 6-day bin, where the bin is based on the relative postmark date of individuals‘ applications. The residuals are generated from a regression of the outcome variable on the predetermined control variables that are included in the baseline regression specification and school by year fixed effects. Triangular weights are not used when running this regression. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using triangular weights and separate linear trends on either side of the cutoff. The dashed lines surrounding these estimated regression lines mark the 95% confidence intervals. The graphs are not all based on the same sample of students as different variables are available for different grades and years.



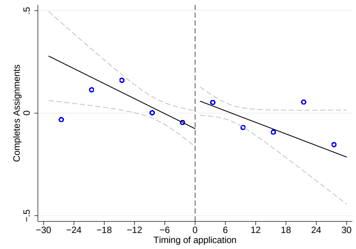
(a) Student shows interest in learning



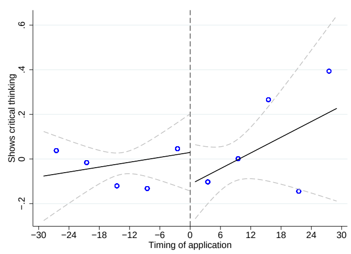
(b) Student respects others



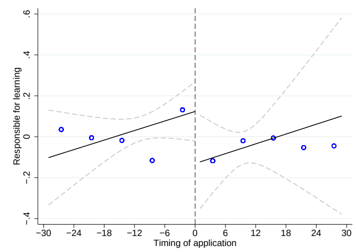
(c) Student prepares and organizes



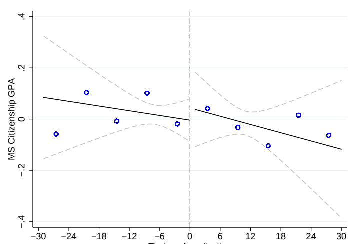
(d) Student completes assignments when due



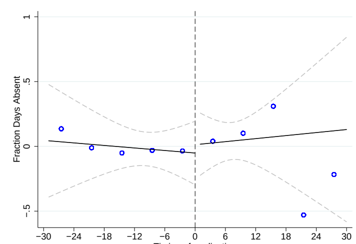
(e) Student demonstrates critical thinking



(f) Student takes responsibility for learning



(g) MS Citizenship GPA



(h) Percentage of school days absent

Figure 1.9: Residual Plot - Behavioral Outcomes

Notes: Each observation is the unweighted average residual in a 6-day bin, where the bin is based on the relative postmark date of individuals' applications. The residuals are generated from a regression of the outcome variable on the predetermined control variables that are included in the baseline regression specification and school by year fixed effects. Triangular weights are not used when running this regression. The dashed vertical line denotes the normalized cutoff application postmark date and the straight lines are estimated regression lines using triangular weights and separate linear trends on either side of the cutoff. The dashed lines surrounding these estimated regression lines mark the 95% confidence intervals. The graphs are not all based on the same sample of students as different variables are available for different grades and years.

1.7.2 Robustness Tables

Table 1.11: Robustness Checks - CST Math

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.0030	0.039	0.062	-0.17	-0.34**	-0.13	0.0045	0.063	0.036	-0.022	-0.041	0.088
	(0.083)	(0.17)	(0.083)	(0.12)	(0.16)	(0.094)	(0.076)	(0.083)	(0.085)	(0.085)	(0.11)	(0.12)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	48	48	48	48	48	48	49	46	47	40	35	13
First-stage F stat.	114.0	96.2	92.5	54.5	20.5	74.9	120.1	68.4	80.1	79.4	70.2	82.3
P-value Wild BS	0.76	0.76	0.38	0.55	0.29	0.50	0.71	0.28	0.52	0.90	0.80	0.84
Observations	620	620	620	620	620	484	765	536	434	504	428	192

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.12: Robustness Checks - CST ELA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	0.084	0.012	0.099	0.027	-0.0059	0.028	0.11	0.058	0.14	0.12	0.084	0.13
	(0.087)	(0.17)	(0.080)	(0.12)	(0.15)	(0.10)	(0.081)	(0.088)	(0.089)	(0.094)	(0.11)	(0.12)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	48	48	48	48	48	48	49	46	47	40	35	13
First-stage F stat.	122.4	96.2	98.8	58.3	22.0	80.1	127.8	61.4	74.4	82.7	75.5	100.0
P-value Wild BS	0.28	0.92	0.090	0.77	0.79	0.74	0.17	0.27	0.082	0.29	0.24	0.64
Observations	619	619	619	619	619	483	765	535	433	503	427	192

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.13: Robustness Checks - SBRC Math (up to '13-'14)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.027	-0.065	0.0037	-0.065	-0.15	-0.049	0.0050	0.013	0.0059	-0.0045	-0.041	0.11
	(0.059)	(0.11)	(0.064)	(0.066)	(0.096)	(0.057)	(0.059)	(0.081)	(0.069)	(0.072)	(0.056)	(0.095)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	55	55	55	55	55	54	55	53	53	43	30	25
First-stage F stat.	125.1	47.3	104.9	63.1	19.2	118.7	114.2	70.3	95.4	117.5	97.2	47.2
P-value Wild BS	0.86	0.59	0.70	0.34	0.10	0.55	0.81	0.70	0.46	0.98	0.84	0.44
Observations	1111	1111	1111	1111	1111	919	1317	944	680	758	515	596

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.14: Robustness Checks -SBRC Reading (up to '13-'14)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.027	-0.097	-0.025	-0.025	-0.093	-0.040	-0.0059	-0.037	-0.066	-0.056	-0.055	0.017
	(0.070)	(0.12)	(0.070)	(0.082)	(0.11)	(0.070)	(0.070)	(0.092)	(0.080)	(0.085)	(0.072)	(0.14)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	55	55	55	55	55	54	55	53	53	43	30	25
First-stage F stat.	119.7	46.4	100.4	61.5	18.7	114.6	110.2	65.2	97.5	106.3	95.6	42.2
P-value Wild BS	0.81	0.41	0.93	0.82	0.51	0.74	0.92	0.89	0.85	0.74	0.93	0.88
Observations	1100	1100	1100	1100	1100	911	1290	933	680	747	514	586

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.15: Robustness Checks - SBRC Writing (up to '13-'14)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.0064	-0.086	-0.015	-0.029	-0.11	-0.025	0.0024	0.013	-0.13	0.0020	-0.083	0.12
	(0.078)	(0.11)	(0.081)	(0.088)	(0.11)	(0.078)	(0.075)	(0.10)	(0.10)	(0.091)	(0.086)	(0.14)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	55	55	55	55	55	54	55	53	53	43	30	25
First-stage F stat.	118.9	46.4	100.0	60.9	18.6	112.1	110.0	64.5	95.4	104.9	95.8	43.4
P-value Wild BS	0.90	0.42	1	0.62	0.34	0.72	1	0.90	0.56	0.83	0.84	0.93
Observations	1100	1100	1100	1100	1100	911	1292	933	680	747	514	586

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.16: Robustness Checks - SBRC Math ('14-'15)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	0.051	0.048	0.068	0.059	0.13	0.063	0.017	0.036	0.012	0.19	-0.023	0.040
	(0.16)	(0.15)	(0.12)	(0.26)	(0.36)	(0.23)	(0.13)	(0.11)	(0.16)	(0.18)	(0.26)	(0.20)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	20	20	19	15	10	10
First-stage F stat.	33.2	34.5	46.2	13.4	10.4	17.2	40.0	41.8	27.8	15.0	17.2	9.71
P-value Wild BS	0.74	0.74	0.58	0.77	0.69	0.73	0.83	0.77	0.90	0.39	0.98	0.85
Observations	353	353	353	353	353	259	390	325	219	224	142	211

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.17: Robustness Checks - SBRC Reading ('14-'15)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.077	0.032	-0.076	0.0039	0.11	-0.0067	-0.071	-0.073	-0.044	-0.042	0.099	-0.31
	(0.18)	(0.18)	(0.15)	(0.25)	(0.34)	(0.23)	(0.16)	(0.12)	(0.19)	(0.21)	(0.27)	(0.21)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	20	20	19	15	10	10
First-stage F stat.	33.2	34.5	46.2	13.4	10.4	17.2	39.9	41.8	27.8	15.0	17.2	9.71
P-value Wild BS	0.76	0.82	0.66	0.96	0.71	0.97	0.71	0.60	0.86	0.93	0.82	0.26
Observations	353	353	353	353	353	259	389	325	219	224	142	211

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.18: Robustness Checks - SBRC Writing ('14-'15)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	0.17	0.26	0.16	0.35	0.46	0.29	0.13	0.0077	0.045	0.28	0.43	-0.011
	(0.19)	(0.18)	(0.14)	(0.28)	(0.34)	(0.26)	(0.15)	(0.14)	(0.19)	(0.22)	(0.31)	(0.21)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	20	20	19	15	10	10
First-stage F stat.	33.0	34.3	45.7	13.3	10.1	17.1	39.6	41.8	27.8	14.9	17.2	9.18
P-value Wild BS	0.56	0.23	0.41	0.26	0.21	0.42	0.62	0.94	0.84	0.37	0.39	0.98
Observations	352	352	352	352	352	258	388	325	219	223	142	210

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.19: Robustness Checks - MS Overall GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.026	-0.094	-0.038	-0.041	-0.032	-0.045	-0.053	-0.061	0.023	-0.027	0.27***	-0.19
	(0.12)	(0.12)	(0.10)	(0.13)	(0.12)	(0.12)	(0.11)	(0.12)	(0.11)	(0.12)	(0.094)	(0.13)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	22	20	18	18	7	13
First-stage F stat.	27.1	29.1	37.0	13.1	8.75	17.1	31.6	12.8	34.1	27.0	20.9	16.0
P-value Wild BS	0.83	0.44	0.73	0.67	0.64	0.79	0.66	0.86	0.82	0.86	0.062	0.37
Observations	415	415	415	415	415	337	477	379	321	398	159	256

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.20: Robustness Checks - MS ELA GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.023	-0.018	-0.027	-0.067	-0.070	-0.059	-0.034	-0.059	0.024	-0.046	0.22	-0.16
	(0.13)	(0.13)	(0.12)	(0.14)	(0.12)	(0.13)	(0.12)	(0.15)	(0.13)	(0.13)	(0.15)	(0.16)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	22	20	18	18	7	13
First-stage F stat.	26.8	29.1	36.4	13.2	8.94	17.2	31.0	12.4	33.9	26.9	20.3	16.3
P-value Wild BS	0.95	0.92	0.85	0.79	0.72	0.86	0.79	0.94	0.89	0.87	0.36	0.53
Observations	415	415	415	415	415	337	477	379	321	398	159	256

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.21: Robustness Checks - MS Math GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.073	-0.096	-0.13	0.086	0.050	0.027	-0.12	-0.11	-0.027	-0.086	0.24	-0.22
	(0.13)	(0.10)	(0.12)	(0.17)	(0.17)	(0.15)	(0.12)	(0.17)	(0.13)	(0.13)	(0.17)	(0.18)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	22	20	18	18	7	13
First-stage F stat.	26.7	28.1	36.4	13.1	9.25	17.0	31.3	12.2	33.8	26.6	19.7	15.8
P-value Wild BS	0.50	0.36	0.28	0.77	0.92	0.88	0.35	0.53	0.84	0.49	0.12	0.30
Observations	410	410	410	410	410	333	472	374	321	393	159	251

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.22: Robustness Checks - Fraction Days Absent

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.11	-0.078	-0.19	0.15	0.32	0.032	-0.11	-0.29	-0.100	0.19	-0.43	0.21
	(0.25)	(0.27)	(0.25)	(0.26)	(0.32)	(0.23)	(0.24)	(0.32)	(0.29)	(0.27)	(0.35)	(0.31)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	90	90	90	90	90	90	90	88	89	73	48	42
First-stage F stat.	148.8	87.9	162.0	72.8	31.1	108.3	165.1	102.4	145.6	115.5	147.6	51.0
P-value Wild BS	0.97	0.81	0.72	0.48	0.32	0.74	0.98	0.74	0.81	0.22	0.47	0.38
Observations	1978	1978	1978	1978	1978	1603	2304	1729	1265	1472	859	1119

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.23: Robustness Checks - Student Interest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	0.026	-0.075	0.049	0.017	-0.0034	0.0046	0.027	0.071	0.082	0.015	0.073	-0.076
	(0.076)	(0.086)	(0.074)	(0.11)	(0.17)	(0.093)	(0.072)	(0.083)	(0.10)	(0.078)	(0.11)	(0.089)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	75	75	75	75	75	74	75	73	72	58	40	35
First-stage F stat.	150.0	72.0	139.9	73.6	25.8	121.1	152.4	103.5	116.6	135.6	122.5	60.7
P-value Wild BS	0.75	0.39	0.49	0.85	0.91	0.92	0.69	0.63	0.36	0.75	0.47	0.55
Observations	1473	1473	1473	1473	1473	1184	1716	1276	902	989	662	811

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.24: Robustness Checks -Respects Others

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.013	-0.10	0.010	-0.056	-0.14	-0.034	-0.012	0.033	-0.11	0.042	-0.044	0.013
	(0.085)	(0.090)	(0.088)	(0.099)	(0.15)	(0.093)	(0.083)	(0.095)	(0.10)	(0.093)	(0.13)	(0.12)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	75	75	75	75	75	74	75	73	72	58	40	35
First-stage F stat.	151.0	72.0	140.6	75.0	26.7	122.5	153.5	103.8	115.2	137.1	122.4	61.7
P-value Wild BS	0.92	0.28	0.64	0.90	0.84	0.91	0.81	1	0.50	0.60	0.86	0.78
Observations	1473	1473	1473	1473	1473	1184	1716	1276	902	989	662	811

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.25: Robustness Checks - Shows critical thinking

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	0.18	0.14	0.18	0.21	0.12	0.22	0.15	0.082	0.044	0.19	0.051	0.34
	(0.16)	(0.13)	(0.12)	(0.21)	(0.30)	(0.23)	(0.13)	(0.14)	(0.14)	(0.20)	(0.35)	(0.21)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	20	20	19	15	10	10
First-stage F stat.	34.1	35.9	47.4	14.4	12.6	18.1	40.4	41.6	27.0	16.0	16.2	10.9
P-value Wild BS	0.36	0.33	0.19	0.39	0.69	0.40	0.36	0.63	0.78	0.41	0.93	0.17
Observations	362	362	362	362	362	265	399	332	222	231	147	215

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.26: Robustness Checks -MS Citizenship GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PT Attendance in Days	-0.055	-0.0069	-0.018	-0.099	-0.086	-0.11	-0.013	-0.16*	0.0097	-0.079	0.19**	-0.20**
	(0.097)	(0.12)	(0.078)	(0.12)	(0.12)	(0.11)	(0.089)	(0.087)	(0.091)	(0.097)	(0.094)	(0.095)
Controls and FEs	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic control	No	No	No	Yes	Yes	No	No	No	No	No	No	No
Cubic control	No	No	No	No	Yes	No	No	No	No	No	No	No
Triangular weights	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	18	42	30	30	30	30	30
Donut RD	No	No	No	No	No	No	No	1-day	No	No	No	No
Before-school applicants	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Fee-based ASPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Provider	Both	Both	Both	Both	Both	Both	Both	Both	Both	Both	1 Only	2 Only
Number of clusters	20	20	20	20	20	20	22	20	18	18	7	13
First-stage F stat.	27.0	29.1	36.1	14.1	9.72	18.7	31.1	12.1	32.8	27.3	22.3	18.0
P-value Wild BS	0.75	1	0.54	0.94	0.82	1	0.56	0.52	0.61	0.85	0.21	0.59
Observations	415	415	415	415	415	337	477	379	321	398	159	256

Notes: The reported coefficient is the 2SLS coefficient that is obtained by estimating equations 1 and 2. Standard errors are clustered at the school by year level. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. *Wild bootstrap performed as in Davidson and MacKinnon (2010) using 999 repetitions.*

Table 1.27: Baseline Results - All Applicants Attend School of Application (Provider 2 Only)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N	(6) P-value Wild BS
Academic Outcomes:						
CST Math	114.0***	0.0043	0.000038	138.9	182	0.68
	(9.67)	(0.13)	(0.00092)			
CST ELA	112.5***	0.14	0.0013	122.2	182	0.88
	(10.2)	(0.18)	(0.0013)			
SBRC Math	68.7***	0.069	0.0010	32.1	557	0.62
	(12.1)	(0.074)	(0.0010)			
SBRC Reading	68.6***	-0.0051	-0.000075	29.0	547	0.79
	(12.7)	(0.11)	(0.0015)			
SBRC Writing	68.6***	0.063	0.00091	29.0	547	0.93
	(12.8)	(0.12)	(0.0016)			
SBRC Math (14-15)	72.6***	-0.21	-0.0030	10.1	184	0.35
	(22.8)	(0.19)	(0.0020)			
SBRC Reading (14-15)	72.6***	-0.47**	-0.0065***	10.1	184	0.064
	(22.8)	(0.24)	(0.0022)			
SBRC Writing (14-15)	70.5***	-0.26	-0.0038	9.64	183	0.18
	(22.7)	(0.20)	(0.0023)			
MS Overall GPA	77.3***	-0.095	-0.0012	16.8	242	0.38
	(18.9)	(0.11)	(0.0013)			
MS ELA GPA	77.6***	-0.11	-0.0015	17.4	242	0.55
	(18.6)	(0.15)	(0.0017)			
MS Math GPA	77.9***	-0.14	-0.0018	18.2	237	0.30
	(18.3)	(0.16)	(0.0018)			
Behavioral Outcomes:						
Fraction Days Absent	67.5***	0.24	0.0035	43.2	1029	0.28
	(10.3)	(0.27)	(0.0038)			
Student Interest	70.0***	-0.13	-0.0018	45.8	743	0.25
	(10.3)	(0.080)	(0.0011)			
Respects Others	70.2***	-0.023	-0.00032	48.2	743	0.90
	(10.1)	(0.093)	(0.0012)			
Prepared for Class	68.5***	-0.17*	-0.0025**	32.5	557	0.090
	(12.0)	(0.091)	(0.0011)			
Completes Assignments	68.7***	-0.24***	-0.0035***	31.9	557	0.046
	(12.1)	(0.085)	(0.00092)			
Shows critical thinking	72.3***	0.11	0.0016	10.1	186	0.65
	(22.7)	(0.23)	(0.0026)			
Responsible for learning	72.3***	0.26	0.0036**	10.1	186	0.12
	(22.7)	(0.17)	(0.0017)			
MS Citizenship GPA	80.7***	-0.15	-0.0019*	21.7	242	0.65
	(17.3)	(0.094)	(0.0011)			

Notes: Columns (1) and (2) show the coefficient on $1[t \leq c]$ from respectively estimating equations 1.1 and 1.3 for the variable in the first column. Column (3) reports the 2SLS coefficient from equation 1.2. Each regression uses the same sample restrictions, bandwidth, kernel, control variables and fixed effects as the baseline specification shown in table 1.3. However, in contrast to table 1.3, in this table only individuals who attended the school of the application for the entire year are included and only data from provider 2 is used. Standard errors are clustered at the school by year level and are reported in parentheses below each coefficient. The first-stage F-value is reported in column (4). The number of observations is reported in column (5). Column (6) reports the Wild boostrap p-value of the 2SLS coefficient. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

1.7.3 Local Linear Regressions

Table 1.28: Local Linear Regressions - CST Math

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	81.5***	0.14	0.0017	113.1	620
	(7.67)	(0.16)	(0.0020)		
24 days	83.0***	0.092	0.0011	103.3	561
	(8.16)	(0.17)	(0.0021)		
18 days	85.8***	-0.034	-0.00040	97.2	484
	(8.70)	(0.20)	(0.0024)		
12 days	79.9***	-0.098	-0.0012	44.9	395
	(11.9)	(0.23)	(0.0030)		
12 days	79.9***	-0.098	-0.0012	46.2	395
	(11.8)	(0.23)	(0.0030)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.29: Local Linear Regressions - CST ELA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	81.6***	0.076	0.00093	110.7	619
	(7.75)	(0.16)	(0.0020)		
24 days	83.0***	0.028	0.00033	98.1	560
	(8.38)	(0.17)	(0.0020)		
18 days	85.8***	-0.097	-0.0011	89.8	483
	(9.05)	(0.19)	(0.0023)		
12 days	79.8***	-0.011	-0.00014	46.7	394
	(11.7)	(0.22)	(0.0028)		
17 days	89.2***	-0.056	-0.00063	91.0	466
	(9.35)	(0.21)	(0.0023)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01.$

Table 1.30: Local Linear Regressions -SBRC Math (up to '13-'14)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	70.8***	-0.021	-0.00030	66.0	1111
	(8.71)	(0.097)	(0.0014)		
24 days	73.1***	-0.048	-0.00066	54.5	1024
	(9.90)	(0.098)	(0.0014)		
18 days	77.3***	-0.066	-0.00085	50.3	919
	(10.9)	(0.11)	(0.0015)		
12 days	74.5***	-0.068	-0.00091	27.9	732
	(14.1)	(0.13)	(0.0019)		
8 days	66.4***	-0.17	-0.0025	14.6	590
	(17.4)	(0.14)	(0.0027)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01.$

Table 1.31: Local Linear Regressions –SBRC Reading (up to '13-'14)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	70.7***	-0.074	-0.0010	67.8	1100
	(8.59)	(0.094)	(0.0014)		
24 days	72.9***	-0.083	-0.0011	59.7	1015
	(9.43)	(0.10)	(0.0014)		
18 days	77.0***	-0.097	-0.0013	50.8	911
	(10.8)	(0.12)	(0.0016)		
12 days	74.6***	-0.092	-0.0012	26.3	724
	(14.5)	(0.12)	(0.0018)		
7 days	60.9***	-0.091	-0.0015	11.4	535
	(18.0)	(0.16)	(0.0036)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.32: Local Linear Regressions –SBRC Writing (up to '13-'14)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	70.7***	-0.048	-0.00067	67.1	1100
	(8.64)	(0.095)	(0.0014)		
24 days	72.9***	-0.068	-0.00093	59.2	1015
	(9.47)	(0.099)	(0.0014)		
18 days	77.0***	-0.12	-0.0015	49.2	911
	(11.0)	(0.11)	(0.0015)		
12 days	74.6***	-0.094	-0.0013	26.5	724
	(14.5)	(0.12)	(0.0016)		
8 days	66.0***	-0.17	-0.0025	15.3	584
	(16.9)	(0.15)	(0.0024)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.33: Local Linear Regressions - SBRC Math ('14-'15)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	92.1***	0.13	0.0014	43.8	353
	(13.9)	(0.18)	(0.0019)		
24 days	97.9***	0.054	0.00055	39.3	314
	(15.6)	(0.18)	(0.0019)		
18 days	98.7***	0.032	0.00033	26.0	259
	(19.3)	(0.20)	(0.0021)		
12 days	97.2***	0.095	0.00098	18.0	217
	(22.9)	(0.24)	(0.0026)		
9 days	88.4***	-0.097	-0.0011	16.1	180
	(22.0)	(0.28)	(0.0034)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.34: Local Linear Regressions - SBRC Reading ('14-'15)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	92.1***	0.049	0.00053	45.8	353
	(13.6)	(0.20)	(0.0022)		
24 days	97.9***	0.044	0.00045	38.3	314
	(15.8)	(0.21)	(0.0022)		
18 days	98.7***	0.099	0.0010	26.8	259
	(19.1)	(0.23)	(0.0024)		
12 days	97.2***	0.049	0.00051	19.2	217
	(22.2)	(0.25)	(0.0027)		
10 days	97.4***	0.066	0.00068	22.0	196
	(20.7)	(0.23)	(0.0025)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.35: Local Linear Regressions - SBRC Writing ('14-'15)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	91.7***	0.31	0.0034	41.9	352
	(14.2)	(0.22)	(0.0022)		
24 days	97.6***	0.30	0.0031	36.8	313
	(16.1)	(0.23)	(0.0022)		
18 days	98.3***	0.36	0.0036	26.8	258
	(19.0)	(0.25)	(0.0026)		
12 days	96.7***	0.37	0.0038	19.3	216
	(22.0)	(0.26)	(0.0028)		
9 days	87.8***	0.29	0.0033	16.3	179
	(21.8)	(0.25)	(0.0032)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.36: Local Linear Regressions - MS Overall GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	77.4***	-0.028	-0.00036	39.7	415
	(12.3)	(0.12)	(0.0016)		
24 days	66.6***	-0.082	-0.0012	25.3	380
	(13.2)	(0.10)	(0.0018)		
18 days	66.5***	-0.14	-0.0021	28.1	337
	(12.5)	(0.13)	(0.0023)		
12 days	66.2***	-0.15	-0.0023	16.3	255
	(16.4)	(0.16)	(0.0043)		
14 days	73.4***	-0.14	-0.0020	26.6	304
	(14.2)	(0.14)	(0.0021)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.37: Local Linear Regressions -MS ELA GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	77.4***	0.017	0.00022	40.2	415
	(12.2)	(0.14)	(0.0019)		
24 days	66.6***	-0.026	-0.00039	26.2	380
	(13.0)	(0.12)	(0.0023)		
18 days	66.5***	-0.057	-0.00085	27.4	337
	(12.7)	(0.13)	(0.0023)		
12 days	66.2***	-0.095	-0.0014	16.3	255
	(16.4)	(0.14)	(0.0061)		
13 days	74.3***	-0.10	-0.0014	25.2	286
	(14.8)	(0.12)	(0.0018)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.38: Local Linear Regressions - MS Math GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	76.5***	-0.093	-0.0012	36.7	410
	(12.6)	(0.10)	(0.0015)		
24 days	65.5***	-0.094	-0.0014	24.0	375
	(13.4)	(0.085)	(0.0017)		
18 days	66.4***	-0.048	-0.00073	26.2	333
	(13.0)	(0.13)	(0.0023)		
12 days	66.8***	-0.017	-0.00026	16.5	252
	(16.5)	(0.19)	(0.0056)		
15 days	73.8***	0.0010	0.000014	25.8	313
	(14.5)	(0.13)	(0.0028)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.39: Local Linear Regressions - Fraction Days Absent

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	73.4***	-0.15	-0.0020	136.9	1978
	(6.27)	(0.25)	(0.0034)		
24 days	73.3***	-0.15	-0.0021	108.7	1811
	(7.03)	(0.27)	(0.0037)		
18 days	76.7***	0.028	0.00037	90.9	1603
	(8.04)	(0.29)	(0.0038)		
12 days	74.7***	0.013	0.00018	49.5	1278
	(10.6)	(0.32)	(0.0043)		
10 days	74.4***	0.28	0.0037	40.2	1151
	(11.7)	(0.36)	(0.0051)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.40: Local Linear Regressions - Student Interest

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	75.8***	-0.060	-0.00079	102.9	1473
	(7.48)	(0.077)	(0.0010)		
24 days	78.9***	-0.072	-0.00092	86.7	1347
	(8.48)	(0.079)	(0.0010)		
18 days	82.8***	-0.073	-0.00088	72.5	1184
	(9.72)	(0.091)	(0.0011)		
12 days	80.3***	-0.051	-0.00064	39.9	955
	(12.7)	(0.11)	(0.0014)		
10 days	77.1***	-0.042	-0.00054	33.3	871
	(13.4)	(0.13)	(0.0016)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.41: Local Linear Regressions - Respects Others

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	75.8***	-0.10	-0.0013	96.9	1473
	(7.70)	(0.083)	(0.0011)		
24 days	78.9***	-0.12	-0.0015	89.8	1347
	(8.33)	(0.083)	(0.0011)		
18 days	82.8***	-0.074	-0.00089	74.5	1184
	(9.59)	(0.095)	(0.0012)		
12 days	80.3***	-0.052	-0.00065	40.9	955
	(12.6)	(0.11)	(0.0014)		
13 days	83.1***	-0.068	-0.00081	48.6	1003
	(11.9)	(0.10)	(0.0012)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.42: Local Linear Regressions - Shows critical thinking

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	93.7***	0.20	0.0022	44.6	362
	(14.0)	(0.18)	(0.0018)		
24 days	98.8***	0.16	0.0016	41.5	323
	(15.3)	(0.17)	(0.0017)		
18 days	101.1***	0.27	0.0027	25.8	265
	(19.9)	(0.20)	(0.0019)		
12 days	99.8***	0.23	0.0023	21.6	223
	(21.4)	(0.21)	(0.0021)		
13 days	101.3***	0.26	0.0026	20.6	231
	(22.3)	(0.20)	(0.0021)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.43: Local Linear Regressions - MS Citizenship GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
30 days	77.4***	0.083	0.0011	40.9	415
	(12.1)	(0.12)	(0.0017)		
24 days	66.6***	0.0024	0.000036	26.4	380
	(13.0)	(0.11)	(0.0020)		
18 days	66.5***	-0.087	-0.0013	26.1	337
	(13.0)	(0.13)	(0.0022)		
12 days	66.2***	-0.079	-0.0012	17.0	255
	(16.0)	(0.14)	(0.0033)		
11 days	74.9***	-0.062	-0.00083	17.6	237
	(17.9)	(0.15)	(0.0083)		

Notes: Estimated coefficients are based on local linear regressions with a uniform kernel and no control variables or fixed effects. The optimal bandwidth based on Calonico et al. (2016) is used in the last row of each table. In each table, column (1) reports the first-stage coefficient, column (2) the reduced form coefficient and column (3) the 2SLS coefficient. The F-value corresponding to the first stage coefficient is reported in column (4) and the number of observations is listed in column (5). Bootstrap standard errors, clustered by school and year, and based on 2000 replications are reported in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

1.7.4 Heterogeneous Impacts

Table 1.44: Heterogeneous Treatment Effects - CST Math

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	84.9***	-0.0032	-0.000038	114.0	620
	(7.96)	(0.096)	(0.0010)		
1. Male	88.9***	-0.25*	-0.0028**	57.9	296
	(11.7)	(0.13)	(0.0012)		
Female	82.2***	0.17	0.0021	45.0	324
	(12.2)	(0.15)	(0.0015)		
2. Single-parent HH	134.6***	-0.50**	-0.0037***	94.4	209
	(13.8)	(0.24)	(0.0013)		
Two-parent HH	85.8***	0.037	0.00044	65.7	407
	(10.6)	(0.11)	(0.0011)		
3. Parent has BA	81.5***	-0.031	-0.00038	43.3	291
	(12.4)	(0.19)	(0.0020)		
Parent has no BA	86.2***	0.15	0.0018	30.5	282
	(15.6)	(0.25)	(0.0025)		
4. Lag above median	89.4***	-0.17	-0.0019	32.1	235
	(15.8)	(0.22)	(0.0021)		
Lag below median	105.0***	-0.074	-0.00071	25.1	234
	(20.9)	(0.18)	(0.0014)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.45: Heterogeneous Treatment Effects - CST ELA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	85.3***	0.090	0.0010	122.4	619
	(7.72)	(0.100)	(0.0011)		
1. Male	91.3***	-0.093	-0.0010	66.3	296
	(11.2)	(0.19)	(0.0017)		
Female	81.7***	0.19*	0.0023**	43.6	323
	(12.4)	(0.11)	(0.0011)		
2. Single-parent HH	135.1***	-0.16	-0.0012	113.3	209
	(12.7)	(0.22)	(0.0013)		
Two-parent HH	85.4***	0.11	0.0013	70.7	406
	(10.2)	(0.11)	(0.0012)		
3. Parent has BA	81.4***	0.21	0.0025	45.3	290
	(12.1)	(0.17)	(0.0019)		
Parent has no BA	84.7***	-0.042	-0.00050	29.9	282
	(15.5)	(0.14)	(0.0014)		
4. Lag above median	87.1***	0.16	0.0018	48.7	236
	(12.5)	(0.12)	(0.0012)		
Lag below median	110.8***	-0.13	-0.0012	40.9	234
	(17.3)	(0.15)	(0.0011)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.46: Heterogeneous Treatment Effects - SBRC Math (up to '13-'14)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	78.3***	-0.027	-0.00034	125.1	1111
	(7.00)	(0.061)	(0.00073)		
1. Male	72.8***	-0.0064	-0.000088	56.6	547
	(9.67)	(0.094)	(0.0012)		
Female	83.5***	-0.043	-0.00051	70.5	564
	(9.94)	(0.077)	(0.00083)		
2. Single-parent HH	94.9***	-0.063	-0.00066	78.1	444
	(10.7)	(0.13)	(0.0012)		
Two-parent HH	67.2***	-0.0027	-0.000041	44.3	660
	(10.1)	(0.075)	(0.0010)		
3. Parent has BA	70.5***	-0.010	-0.00015	77.4	488
	(8.01)	(0.089)	(0.0011)		
Parent has no BA	79.4***	0.012	0.00015	42.2	505
	(12.2)	(0.11)	(0.0012)		
4. Lag above median	70.2***	-0.076	-0.0011	60.3	471
	(9.04)	(0.093)	(0.0012)		
Lag below median	72.3***	-0.075	-0.0010	20.5	464
	(16.0)	(0.11)	(0.0013)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.47: Heterogeneous Treatment Effects - SBRC Reading (up to '13-'14)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	78.2***	-0.027	-0.00034	119.7	1100
	(7.14)	(0.073)	(0.00087)		
1. Male	73.1***	-0.080	-0.0011	53.1	539
	(10.0)	(0.12)	(0.0014)		
Female	83.7***	0.040	0.00047	70.1	561
	(10.00)	(0.086)	(0.00094)		
2. Single-parent HH	95.3***	-0.14	-0.0014	80.6	435
	(10.6)	(0.18)	(0.0016)		
Two-parent HH	67.3***	-0.042	-0.00063	43.9	658
	(10.2)	(0.088)	(0.0012)		
3. Parent has BA	70.6***	0.031	0.00044	76.3	480
	(8.09)	(0.11)	(0.0015)		
Parent has no BA	78.6***	-0.0100	-0.00013	39.9	502
	(12.5)	(0.098)	(0.0011)		
4. Lag above median	65.9***	-0.072	-0.0011	34.2	474
	(11.3)	(0.095)	(0.0013)		
Lag below median	75.9***	-0.079	-0.0010	28.6	423
	(14.2)	(0.11)	(0.0013)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.48: Heterogeneous Treatment Effects - SBRC Writing (up to '13-'14)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	78.0***	-0.0063	-0.000080	118.9	1100
	(7.15)	(0.080)	(0.00097)		
1. Male	73.1***	-0.096	-0.0013	53.6	539
	(9.98)	(0.15)	(0.0018)		
Female	83.4***	0.12	0.0015	68.7	561
	(10.1)	(0.083)	(0.00095)		
2. Single-parent HH	95.0***	-0.15	-0.0016	76.4	435
	(10.9)	(0.18)	(0.0017)		
Two-parent HH	67.3***	-0.017	-0.00025	44.6	658
	(10.1)	(0.097)	(0.0013)		
3. Parent has BA	70.1***	-0.15	-0.0022	74.8	480
	(8.11)	(0.11)	(0.0014)		
Parent has no BA	78.6***	0.29***	0.0037***	40.1	502
	(12.4)	(0.098)	(0.0013)		
4. Lag above median	67.1***	-0.053	-0.00078	40.4	451
	(10.6)	(0.13)	(0.0017)		
Lag below median	61.2***	-0.071	-0.0012	21.3	446
	(13.3)	(0.14)	(0.0020)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.49: Heterogeneous Treatment Effects - SBRC Math ('14-'15)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	81.5***	0.052	0.00064	33.2	353
	(14.2)	(0.18)	(0.0020)		
1. Male	66.1**	-0.20	-0.0030	5.72	169
	(27.6)	(0.31)	(0.0042)		
Female	96.5***	0.16	0.0016	22.6	184
	(20.3)	(0.19)	(0.0017)		
2. Single-parent HH	69.9***	-0.021	-0.00030	11.6	175
	(20.5)	(0.27)	(0.0032)		
Two-parent HH	68.6**	0.00069	0.000010	6.12	178
	(27.7)	(0.28)	(0.0034)		
3. Parent has BA	84.4***	0.48**	0.0057***	16.9	173
	(20.6)	(0.19)	(0.0021)		
Parent has no BA	70.6***	-0.33	-0.0047	17.2	162
	(17.0)	(0.31)	(0.0037)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$.

Table 1.50: Heterogeneous Treatment Effects - SBRC Reading ('14-'15)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	81.5***	-0.078	-0.00096	33.2	353
	(14.2)	(0.20)	(0.0022)		
1. Male	66.1**	-0.35	-0.0053	5.72	169
	(27.6)	(0.26)	(0.0043)		
Female	96.5***	0.049	0.00051	22.6	184
	(20.3)	(0.16)	(0.0013)		
2. Single-parent HH	69.9***	-0.038	-0.00055	11.6	175
	(20.5)	(0.33)	(0.0039)		
Two-parent HH	68.6**	-0.32	-0.0047	6.12	178
	(27.7)	(0.24)	(0.0038)		
3. Parent has BA	84.4***	0.036	0.00042	16.9	173
	(20.6)	(0.32)	(0.0031)		
Parent has no BA	70.6***	-0.040	-0.00056	17.2	162
	(17.0)	(0.26)	(0.0031)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$.

Table 1.51: Heterogeneous Treatment Effects - SBRC Writing ('14-'15)

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	81.5***	0.17	0.0021	33.0	352
	(14.2)	(0.21)	(0.0023)		
1. Male	66.1**	-0.063	-0.00096	5.72	169
	(27.6)	(0.21)	(0.0027)		
Female	96.5***	0.35	0.0036	22.6	183
	(20.3)	(0.23)	(0.0023)		
2. Single-parent HH	69.9***	0.43	0.0061	11.6	175
	(20.5)	(0.33)	(0.0037)		
Two-parent HH	67.6**	-0.12	-0.0018	5.89	177
	(27.8)	(0.24)	(0.0031)		
3. Parent has BA	84.4***	0.37	0.0044	16.9	173
	(20.6)	(0.37)	(0.0037)		
Parent has no BA	69.7***	0.042	0.00061	16.9	161
	(17.0)	(0.31)	(0.0036)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.52: Heterogeneous Treatment Effects - MS Overall GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	72.0***	-0.023	-0.00032	27.1	415
	(13.8)	(0.11)	(0.0015)		
1. Male	88.1***	0.051	0.00058	18.8	206
	(20.3)	(0.20)	(0.0019)		
Female	58.6***	-0.093	-0.0016	21.3	209
	(12.7)	(0.10)	(0.0016)		
2. Single-parent HH	111.1***	-0.62**	-0.0056**	25.5	136
	(22.0)	(0.30)	(0.0028)		
Two-parent HH	73.1***	0.15	0.0020	23.0	279
	(15.2)	(0.12)	(0.0015)		
3. Parent has BA	83.7***	-0.0096	-0.00011	20.5	261
	(18.5)	(0.11)	(0.0011)		
Parent has no BA	75.9***	-0.098	-0.0013	29.1	135
	(14.1)	(0.19)	(0.0021)		
4. Lag above median	128.8***	0.062	0.00048	20.0	100
	(28.8)	(0.064)	(0.00034)		
Lag below median	46.3	0.40**	0.0087	1.39	99
	(39.2)	(0.18)	(0.0054)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.53: Heterogeneous Treatment Effects - MS ELA GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	71.9***	-0.021	-0.00029	26.8	415
	(13.9)	(0.13)	(0.0016)		
1. Male	88.4***	-0.051	-0.00058	19.1	206
	(20.2)	(0.26)	(0.0026)		
Female	57.3***	-0.031	-0.00053	20.0	209
	(12.8)	(0.11)	(0.0016)		
2. Single-parent HH	111.0***	-0.51*	-0.0046*	28.2	136
	(20.9)	(0.31)	(0.0027)		
Two-parent HH	73.6***	0.14	0.0019	23.3	279
	(15.2)	(0.13)	(0.0017)		
3. Parent has BA	83.4***	-0.019	-0.00023	20.6	261
	(18.4)	(0.12)	(0.0013)		
Parent has no BA	75.3***	0.028	0.00037	24.1	135
	(15.3)	(0.29)	(0.0031)		
4. Lag above median	138.0***	0.16	0.0011*	44.4	102
	(20.7)	(0.11)	(0.00066)		
Lag below median	58.0	-0.14	-0.0024	2.08	97
	(40.2)	(0.42)	(0.0058)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.54: Heterogeneous Treatment Effects - MS Math GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	72.2***	-0.066	-0.00092	26.7	410
	(14.0)	(0.13)	(0.0016)		
1. Male	85.6***	-0.0079	-0.000092	17.2	203
	(20.6)	(0.29)	(0.0029)		
Female	58.3***	-0.14	-0.0025*	20.6	207
	(12.8)	(0.090)	(0.0014)		
2. Single-parent HH	111.7***	-0.055	-0.00049	21.7	133
	(24.0)	(0.32)	(0.0023)		
Two-parent HH	73.1***	0.025	0.00035	22.4	277
	(15.4)	(0.14)	(0.0017)		
3. Parent has BA	83.9***	-0.020	-0.00024	20.5	260
	(18.5)	(0.13)	(0.0014)		
Parent has no BA	73.9***	-0.39	-0.0053	23.0	132
	(15.4)	(0.26)	(0.0033)		
4. Lag above median	108.5***	0.13	0.0012**	20.6	99
	(23.9)	(0.096)	(0.00054)		
Lag below median	69.9**	-0.089	-0.0013	3.89	99
	(35.4)	(0.27)	(0.0027)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.55: Heterogeneous Treatment Effects - Fraction Days Absent

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	76.2***	-0.10	-0.0013	148.8	1978
	(6.25)	(0.25)	(0.0031)		
1. Male	75.7***	0.24	0.0032	79.8	968
	(8.47)	(0.31)	(0.0039)		
Female	77.3***	-0.72**	-0.0093**	77.5	1010
	(8.78)	(0.34)	(0.0042)		
2. Single-parent HH	95.0***	-0.44	-0.0046	108.8	812
	(9.11)	(0.58)	(0.0056)		
Two-parent HH	65.9***	0.19	0.0028	63.9	1159
	(8.24)	(0.27)	(0.0038)		
3. Parent has BA	72.2***	0.37	0.0051	77.4	968
	(8.21)	(0.34)	(0.0045)		
Parent has no BA	76.2***	-0.68*	-0.0089*	61.1	850
	(9.75)	(0.41)	(0.0049)		
4. Lag above median	64.6***	-0.085	-0.0013	62.2	925
	(8.19)	(0.40)	(0.0057)		
Lag below median	78.4***	0.050	0.00064	66.4	905
	(9.62)	(0.32)	(0.0037)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.56: Heterogeneous Treatment Effects - Student Interest

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	79.1***	0.025	0.00032	150.0	1473
	(6.46)	(0.078)	(0.00095)		
1. Male	72.9***	0.12	0.0016	61.6	723
	(9.28)	(0.12)	(0.0015)		
Female	85.1***	-0.042	-0.00049	93.6	750
	(8.80)	(0.10)	(0.0011)		
2. Single-parent HH	90.4***	0.043	0.00048	110.7	626
	(8.59)	(0.15)	(0.0015)		
Two-parent HH	67.6***	-0.079	-0.0012	50.4	840
	(9.52)	(0.10)	(0.0014)		
3. Parent has BA	73.8***	0.14	0.0019	83.6	665
	(8.07)	(0.12)	(0.0015)		
Parent has no BA	79.3***	0.084	0.0011	54.9	671
	(10.7)	(0.13)	(0.0014)		
4. Lag above median	73.1***	0.067	0.00092	58.9	661
	(9.52)	(0.10)	(0.0013)		
Lag below median	62.1***	-0.28	-0.0046	19.5	449
	(14.1)	(0.20)	(0.0028)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.57: Heterogeneous Treatment Effects - Respects Others

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	79.1***	-0.013	-0.00016	151.0	1473
	(6.44)	(0.088)	(0.0011)		
1. Male	73.2***	-0.015	-0.00021	63.0	723
	(9.23)	(0.17)	(0.0021)		
Female	85.3***	0.0036	0.000042	93.5	750
	(8.82)	(0.11)	(0.0012)		
2. Single-parent HH	90.4***	-0.060	-0.00067	111.3	626
	(8.57)	(0.15)	(0.0015)		
Two-parent HH	67.5***	0.046	0.00068	51.2	840
	(9.43)	(0.10)	(0.0014)		
3. Parent has BA	73.4***	0.031	0.00042	82.8	665
	(8.07)	(0.099)	(0.0012)		
Parent has no BA	79.0***	0.0043	0.000055	56.6	671
	(10.5)	(0.18)	(0.0021)		
4. Lag above median	78.4***	-0.087	-0.0011	106.2	661
	(7.61)	(0.16)	(0.0018)		
Lag below median	57.7***	-0.099	-0.0017	17.8	449
	(13.7)	(0.24)	(0.0037)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.58: Heterogeneous Treatment Effects - Shows critical thinking

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	83.0***	0.18	0.0022	34.1	362
	(14.2)	(0.18)	(0.0020)		
1. Male	66.5**	0.16	0.0023	6.35	176
	(26.4)	(0.27)	(0.0033)		
Female	101.3***	0.22	0.0022	26.8	186
	(19.6)	(0.19)	(0.0017)		
2. Single-parent HH	76.6***	0.29	0.0038**	15.7	182
	(19.3)	(0.20)	(0.0019)		
Two-parent HH	68.5**	0.39	0.0057**	6.12	180
	(27.7)	(0.24)	(0.0024)		
3. Parent has BA	85.1***	0.21	0.0024	17.0	177
	(20.6)	(0.24)	(0.0023)		
Parent has no BA	70.2***	0.026	0.00037	18.8	166
	(16.2)	(0.22)	(0.0026)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ *that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5).* $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.59: Heterogeneous Treatment Effects - MS Citizenship GPA

	(1) FS	(2) RF	(3) 2SLS	(4) FS F-value	(5) N
All students	73.1***	-0.051	-0.00069	27.0	415
	(14.1)	(0.099)	(0.0012)		
1. Male	93.7***	-0.075	-0.00080	23.2	206
	(19.4)	(0.21)	(0.0020)		
Female	56.1***	-0.16	-0.0028	19.9	209
	(12.6)	(0.11)	(0.0018)		
2. Single-parent HH	108.9***	-0.38	-0.0035	29.3	136
	(20.1)	(0.27)	(0.0023)		
Two-parent HH	76.0***	0.023	0.00030	24.0	279
	(15.5)	(0.12)	(0.0015)		
3. Parent has BA	85.2***	-0.054	-0.00064	21.5	261
	(18.4)	(0.10)	(0.0010)		
Parent has no BA	71.6***	-0.092	-0.0013	19.4	135
	(16.3)	(0.22)	(0.0025)		
4. Lag above median	107.5***	0.060	0.00056	21.5	101
	(23.2)	(0.088)	(0.00064)		
Lag below median	88.7***	0.13	0.0015	7.76	98
	(31.8)	(0.23)	(0.0021)		

Notes: Columns (1) and (2) report the coefficient on $1[t \leq c]$ that is obtained by respectively estimating the first stage equation 1 and the reduced form equation 3 for the subgroup of students given in the first column and using the variable in the column header as the dependent variable. Column (3) reports the 2SLS coefficient from equation 2. The sample restrictions and control variables used in the regressions are exactly the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. The F-value corresponding to the first stage coefficient is reported in column (4). The number of observations is reported in column (5). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

1.7.5 Ordinary Least Squares

Table 1.60: OLS Results

	(1) OLS	(2) BS P-value	(3) N
Academic Outcomes:			
CST Math	-0.00022	0.0035	319
	(0.00097)		
CST ELA	0.0017*	0.037	318
	(0.00089)		
SBRC Math	-0.00044	0.73	530
	(0.00066)		
SBRC Reading	0.00014	0.19	525
	(0.00062)		
SBRC Writing	-0.00059	0.63	525
	(0.00071)		
SBRC Math (14-15)	-0.0032***	0.34	185
	(0.00093)		
SBRC Reading (14-15)	-0.0020***	0.11	185
	(0.00066)		
SBRC Writing (14-15)	-0.0028***	0.51	184
	(0.00045)		
MS Overall GPA	0.00060	0.12	218
	(0.00056)		
MS ELA GPA	0.0012	0.13	218
	(0.00071)		
MS Math GPA	-0.00038	0.47	215
	(0.00083)		
Behavioral Outcomes:			
Fraction Days Absent	-0.0059***	0.00001	986
	(0.0017)		
Student Interest	-0.0018***	0.22	719
	(0.00069)		
Respects Others	-0.00096	0.46	719
	(0.00067)		
Prepared for Class	-0.0021***	0.017	530
	(0.00072)		
Completes Assignments	-0.0022***	0.032	530
	(0.00084)		
Shows critical thinking	-0.0027***	0.42	189
	(0.0010)		
Responsible for learning	-0.0011	0.40	189
	(0.00096)		
MS Citizenship GPA	0.00043	0.78	218
	(0.00086)		

Notes: Column (1) shows the coefficient on days of PT attendance that is obtained by estimating equation 2 using an OLS regression and by not controlling for the relative postmark date of the application. The sample only includes students who applied on or at most 29 days before the cutoff date and who attended the school of the PrimeTime application for the entire year. The control variables used in the regressions are the same as in the baseline regressions presented in table 3. Standard errors are clustered at the school by year level and are reported in the parentheses below each coefficient. No weights are used. Column (2) reports the Wild bootstrap p-value of the OLS coefficient based on 1000 repititions. The number of observations is reported in column (3). $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Chapter 2

Personal Income Taxation and College Major Choice: A Case Study of the 1986 Tax Reform Act

Abstract

This paper evaluates whether changes in expected lifetime income by major that result from changes in the individual income tax law have the potential to affect college major choices, using one of the largest federal income tax reforms in recent U.S. history, the 1986 Tax Reform Act (TRA86), as a case study. Using various labor force surveys, I first calculate the change in expected after-tax lifetime income for males due to TRA86 for 47 majors. I find that the average major experienced an increase in after-tax expected lifetime income of \$58.000 or, equivalently, 6.2% due to TRA86. The differential impact across majors is small. Relative to the mean across all majors expected earnings for a particular major changed from -1.2% to 2.0%. I next use estimates of the elasticity of college major choice with respect to major-specific expected earnings from other studies to simulate how the differential TRA86-induced change in expected after-tax lifetime earnings across majors could have affected

the distribution of completed college majors. Due to the limited differential impact on expected earnings across majors, at most 0.25% of males graduating after TRA86 are estimated to have completed a different major due to TRA86.

2.1 Introduction

A large literature estimates the elasticity of taxable income. This literature captures mostly the short-run effects of tax changes on efficiency including changes in hours worked, form of compensation and occupational choice. However, other less-studied responses can take longer to show up in the data and have important implications for the long-run effects of tax changes on efficiency.[1] In particular, changes in the tax law may not only affect the amount of education that a person chooses to obtain, as examined among others by Lucas (1990) and Trostel (1993), but also the type of education pursued. For instance, in college, a reduction in marginal income tax rates should induce students to switch from majors leading to pleasant but low-paying jobs to majors leading to less pleasant but higher-paying jobs. The objective of this paper is to examine the potential for changes in the personal income tax law to affect college major choices, using one of the largest federal income tax reforms in recent U.S. history, the 1986 Tax Reform Act (TRA86), as a case study. In doing so, as far as I know, I am the first to consider the effect of personal income taxation on the choice of educational type.

The expected lifetime earnings of a major have been shown to be an important determinant of college major choice (Berger (1988), Arcidiacono (2004), Arcidiacono et al. (2012), Wiswall and Zafir (2015a)), suggesting a link should exist between the tax law and college major choice. However, the current literature on college major choice uses pre-tax earnings as a proxy for future income streams.[2] There has not yet been a study linking changes in the

[1] Some recent papers look at long-run responses to changes in marginal income tax rates. For instance, Gentry and Hubbard (2004), Powell and Shan (2012) and Kreiner et al. (2015) look at the effect that marginal tax rates have on job mobility and on the long-run allocation of labor across profitable opportunities.

[2] The only paper that uses after-tax earnings by major is Skyt Nielsen and Vissing-Jorgensen (2006). However, they do not explicitely consider how a change in the individual

personal income tax law to changes in the choice of college major.

Studying the effect of personal income taxation on college major choice is challenging for a number of reasons. First, one needs both a very large income tax change and data on major-specific lifetime earnings from the period of the tax change. The largest personal income tax changes in the U.S. in recent history occurred during the 1980s. However, survey data that contain information on an individual's earnings, level of educational attainment and type of educational attainment, which would allow one to construct measures of after-tax expected lifetime earnings by major, are available at the earliest in 1993.

Second, even if one calculates the change in expected after-tax lifetime earnings by major due to a tax reform, it is difficult to evaluate the impact of this change on the types of college majors completed. As shown later, the change in expected lifetime earnings of a major due to a tax reform can be strongly correlated with the pre-tax reform level of expected lifetime earnings of that major. This makes it difficult to distinguish between changes in the composition of completed college majors that are due to the tax reform and those that are due to other factors correlated with baseline expected lifetime incomes by major. In addition, there are no large individual-level datasets in the U.S. that follow more than two or three consecutive cohorts of college students over time. One needs to rely on aggregate data on the number of college degrees completed by major over time when studying the impact of personal income taxation on college major choice. Such data do not allow one to directly control for changes in the composition of college students over time.

In this paper, I evaluate the potential for TRA86 to have affected the college major choice decisions of males. TRA86 was one of the largest federal income tax reforms in recent history. It was a largely regressive reform that simplified the income tax law by reducing the number of marginal tax rates and broadening the tax base.

First, I estimate the change in after-tax expected lifetime earnings by major due to TRA86. To do so, I first use the 1993 National Survey of Col-

income tax law affects college major choice.

lege Graduates (NSCG) to construct a distribution of male college graduates across occupations by age group and college major. Using this distribution allows me to overcome the problem that survey data that contain information on individuals' educational type and earnings are not available before 1993. I then link these distributions to average after-tax income by occupation and age group, calculated under both the pre- and post-TRA86 tax law and with pre-TRA86 March CPS data, to get an estimate of the change in expected after-tax lifetime earnings by major due to TRA86 for a total of 47 majors.

The correlation between pre-TRA86 after-tax expected lifetime earnings by major and the change in these earnings due to TRA86 is 0.93. As discussed above, this high correlation and the lack of large individual-level datasets of college students over time make it difficult to credibly estimate the causal impact of TRA86 on the types of college majors completed. Given this difficulty, I next use estimates of the elasticity of college major choice with respect to major-specific expected earnings from other studies to simulate how the differential TRA86-induced change in expected after-tax lifetime earnings across majors could have affected the distribution of completed college majors. Lastly, I briefly evaluate the welfare effects of this simulated change in the distribution of completed college majors using the framework of Hendren (2016).

I find that the average major experienced an increase in after-tax expected lifetime income of $58.000 or, equivalently, 6.2% due to TRA86. Most of this increase is due to increased after-tax income in the second half of one's working life. The amount gained varies significantly across majors from a low of $39.000 in Home Economics to a high of $97.000 in Medical Sciences and Nursing. Although the absolute differences in increases between majors are large, relative to the mean across all majors expected earnings changed relatively little. The minimum and maximum change relative to the mean is respectively -1.2 and 2.0%.

The mean increase in after-tax expected lifetime income due to TRA86 of 6.2% is sizeable and for instance comparable to the effect that graduating at different points in the business cycle can have on the first five years of labor market earnings (see Beffy et al. (2012)). The differential effect across majors is relatively small though when compared to this business cycle effect. As differential effects are important for college major choice, the simulations show small effects of TRA86 on the distribution of completed college majors. Under conventional estimates of the elasticity of the share of degrees completed in a particular field with respect to that field's expected earnings, at most 0.25% of males graduating after TRA86 are estimated to have completed a different major due to TRA86. In terms of the absolute number of completions, the major Education is most negatively impacted, whereas Political Science, Life Sciences and Medical Sciences including Nursing see the largest positive changes. For a given cohort graduating after TRA86, the corresponding welfare effects are small and positive, and equal at most 0.04% of the expected lifetime income tax payments of that cohort.

The results of this paper are potentially important to both researchers and policy makers. First, the results show that personal income tax reforms can have large effects on the mean level of expected earnings of a given level of education. However, given that all educational types within a given educational level are similarly affected, the differential impact across educational types is not likely to be very large. This suggests that changes in the amount rather than the type of education chosen as a result of a tax reform might be more important. Expected earnings vary more across educational levels than across educational types for a given level of education, and a tax reform therefore likely has a larger differential impact across levels of education.

In addition, the lack of a differential impact across educational types implies that current short-run estimates of the elasticity of taxable income with respect to the net-of-tax rate (see e.g. Saez et al. (2012)), a central parameter for calculating the efficiency losses from taxation, are not heavily understating the long-run elasticity. These current estimates, which rely on the short-run response of taxable income to changes in marginal tax rates, would not be heavily impacted by a small change in the composition of completed college majors in the long-run as a result of a tax change.

This paper proceeds as follows. Section 2 discusses the related literature. Section 3 briefly describes the 1986 Tax Reform Act. Section 4 describes the data. Section 5 discusses the methodology. Section 6 presents the effect of the tax change on expected after-tax lifetime earnings by major. Section 7 discusses the simulated effect on the types of majors completed and section 8 concludes.

2.2 Related Literature

Personal income taxation can cause many distortions to individuals' behavior. The distortion that has been most frequently studied is to hours of work. Other possible distortions include distortions to effort, form of pay, occupational choice and educational choice. In terms of educational choice, both the type and level of education that an individual chooses to obtain can be distorted by personal income taxation. The literature has mostly focused on the distortion to the level of education and ignored the distortion to the type of education.[3] [4]

There is a large literature on the determinants of college major choice. This literature has found the expected earnings of a major to be an important determinant[5], although less so than the consumption value of different majors (e.g. Arcidiacono (2004)) or individuals' major-specific abilities (e.g. Freeman and Hirsch (2008), Stinebrickner and Stinebrickner (2014).).[6] As a

[3]Some papers that study the distortion to the amount of education obtained are Ben-Porath (1967), Boskin (1975), Heckman (1976), Lucas (1990), Trostel (1993) and Bovenberg and Jacobs (2005).

[4]There are two theoretical papers that look at the effect of income taxation on the choice of educational type: Alstadster et al. (2008) and Malchow-Mller et al. (2011). Unfortunately, these papers are both of limited use in forming predictions for the actual observed effects of (changes in) income taxation on the choice of educational type. For instance, in Alstadsaeter et al.'s model there are only two types of education between which an individual can choose, and in Malchow-Moller et al.'s model alll individuals have exactly identical preferences and there are only two different ability levels that individuals can have. These assumptions are hard to reconcile with the literature on college major choice discussed here.

[5]See for instance, Berger (1988), Arcidiacono (2004), Beffy et al. (2012), Arcidiacono et al. (2012), Hastings et al. (2015), Long et al. (2015), Wiswall and Zafir (2015a, 2015b), Altonji et al. (2016b) and Altonji et al. (2016a).

[6]Many other factors such as the major-specific or national unemployment rate at the time of the major choice decision (see Blom (2012), Bradley (2013), Clarke (2015) and Blom et al. (2015)) or the classroom composition of introductory college courses (Fischer (2016)) can

measure of students' expected earnings in a major, some papers have used students' stated expected earnings in that major (e.g. Arcidiacono et al. (2012), Wiswall and Zafir (2015a, 2015b)) whilst others (e.g. Berger (1988), Beffy et al. (2012)) have generated students' expected future earnings for a major under the assumption of rational expectations and with an econometric model estimated with actual future earnings data. Lastly, some papers (e.g. Long et al. (2015)) have used population-based average earnings of college graduates with a particular major at the time that students are in college as a proxy for students' expected future earnings in a major. All these papers have found their measure of expected earnings to be an important determinant for college major choice.

Key to the approach of this paper are the results of Wiswall and Zafir (2015a, 2015b) and Long et al. (2015) that students' choice of college major reponds to changes in the population distribution of earnings by major. At first sight, this is not entirely obvious as many recent papers[7] highlight that the expected monetary returns to a given major can vary dramatically across individuals because they are a function of the multidimensional ability vector of an individual. As a result, one might worry that earnings observed in the population of graduates with a particular major might not be useful to current undergraduates in helping predict their own major-specific lifetime income. The fact that population-based earnings by major do in fact matter for college major choice is key to this paper as I calculate the effect that TRA86 had on average after-tax lifetime earnings by major, which will not be adjusted for self-selection into majors and will be based on the actual earnings observed in the population of people with particular majors.

In this paper I hope to add to the college major choice literature discussed above and the literature on the elasticity of taxable income mentioned in the introduction by being the first paper to explicitely consider the effect of changes in personal income taxation on the choice of educational type. This

also affect the major a student chooses.

[7]See for instance Arcidiacono (2004), Arcidiacono et al. (2012) and Kirkeboen et al. (2016).

involves first considering how (changes in) personal income taxes affect the expected returns to various majors and then considering how these changes in expected returns affect college major choice patterns.

2.3 1986 Tax Reform Act

Prior to the 1986 Tax Reform Act (TRA86), the federal personal income tax schedule had 14 different tax brackets and marginal tax rates that varied from a low of 11 to a high of 50 percent [8]. TRA86 reduced this to essentially two tax brackets and two marginal tax rates. In 1988, when the individual income tax law provisions were fully phased in, a taxpayer filing for instance as a single person faced a marginal tax rate of 15 % on his first $17,850 of taxable income and a marginal tax rate of 28% on the remaining part of his taxable income.[9] Using data from the IRS Public Use Tax Return data file, Poterba and Hausman (1987) estimate that 58.9% of the taxpayers saw a reduction in their marginal tax rates as a result of TRA86. 13.8% saw no change and 27.3% faced a higher marginal tax rate. TRA86 also increased the standard deduction and the personal exemption, meaning that many low-income individuals did not need to pay federal income taxes after the reform.

To compensate for the overall expected revenue shortfall as a result of these measures, TRA86 also included a number of tax base broadening measures and raised corporate income taxes. These tax base broadening measures included the abolishment of a rule that allowed 60 percent of realized capital gains to be excluded in the calculation of adjusted gross income and the implementation of a rule that restricted the use of passive losses to offset other income. Corporate income taxes were increased by lengthening depreciation lives and eliminating the tax investment credit. TRA86 was signed into law by Ronald Reagan on Oct. 22nd 1986 and the individual income tax changes

[8] See Poterba and Hausman (1987), Feldstein (1995) and Auerbach and Slemrod (1997) for a detailed discussion of the 1986 Tax Reform Act

[9] However, a variety of phase-out provisions of for instance the 15 % marginal tax rate bracket meant that an individual filing as a single person could face marginal tax rates up to 33% for some ranges of taxable income in excess of $17,850.

went into full effect for the tax (and calender) year 1988.

In this paper, I will only focus on the effect that the change in the federal personal income tax law had on major-specific expected lifetime earnings. Since I use income data from before TRA86 I will not be able to say anything about the possible effect of the higher corporate income taxes on after-tax expected lifetime earnings by major.[10]

2.4 Data

To calculate the TRA86-induced change in expected after-tax lifetime income by major I use two data sources. I use the 1993 National Survey of College Graduates (NSCG) of the NSF to calculate a distribution of male college graduates across occupations by college major and age group. The NSCG is an extensive survey of around 150.000 college-educated individuals and has information on individuals' occupations and educational histories. Ideally, I would have used a survey such as the NSCG from before the tax change to calculate these occupational distributions, but no such survey exists.

I also use the 1985, 1986 and 1987 (Unicon) versions of the March Supplement to the Current Population Survey. In this supplement questions are specifically aimed at respondents' income and labor market involvement in the previous year. Two aspects of this survey are worth mentioning. First, for each individual, the survey contains information on the principal occupation of the individual in the previous year. This occupation can be linked to the occupations used in the NSCG. Second, since the March CPS is a household survey,

[10]I also considered looking at the impact of other large tax reforms in this paper. I cannot look at the effects of the 1981 Economic Recovery Tax Act as the data on college major completions only starts in 1983/1984. In addition, Saez (2004) shows that only the top 1% of income earners in the population were significantly affected by the 1990 OBRA and 1993 OBRA. For instance, the overall average marginal tax rate (excluding the small change in the payroll tax wage cap) that tax units in the top 1-10% (in terms of gross income) faced remained virtually unchanged from 1990 to 1996. For the top 1% however, it went up from 28.9 to 37.74%. The top 1% of tax units are exactly the people for whom earnings are top-coded in the March CPS, making it hard to calculate the impact of the 1990 OBRA and 1993 OBRA on after-tax expected lifetime earnings by major. Also, a lot fewer people will be affected by the 1990 and 1993 tax reforms than the 1986 tax reform. Hence, I decided to also exclude these tax reforms from this study.

it contains information on the sources of income and labor force involvement of all individuals in the household in the previous year. As the March CPS also contains the tax filing status of each individual, this information can then be used to impute the state and federal tax liabilities *of the tax filing unit* to which the individual belongs for the year on which the March CPS is based.

I impute these tax liabilities in the March CPS using version 9.2 of the NBER's TAXSIM program[11]. This program requires 20 different inputs, which together capture the most important items that one needs to list on form 1040.[12] TAXSIM then uses the federal and state tax law of that year to calculate the state, federal and payroll tax liabilities of the tax unit in that year.[13]

Data on the number of bachelor degrees awarded by institution, field (i.e. major), gender and year that is used to simulate the effect of TRA86 on the composition of completed college majors is obtained from the Higher Education General Information Survey (HEGIS) prior to the academic year 1986/1987 and from the Integrated Postsecondary Education Data System (IPEDS) from 1986/1987 onwards. This data is a census of the number and types of bachelor degrees awarded at all federally accredited institutions of higher education in the U.S. from 1983/1984 to 1996/1997. The six digit code used in these surveys to classify the field of the instructional program (the Classification of Instructional Programs (CIP)) is highly disaggregated and distinguishes between close to a thousand fields of specialization. In this paper, I distinguish between 47 different degree fields.[14]

[11]For more information on TAXSIM, see Coutts and Feenberg (1993).

[12]These include the tax year, state of residence, marital status, number of dependents exemptions, wage and salary income of the taxpayer, wage and salary income of the spouse, dividend income and information on around 10 other sources of income.

[13]In order to use the March CPS data with the TAXSIM, I modified the program of Judith Scott-Clayton that is available at the TAXSIM website of the NBER. See the appendix section 2.9.2 on TAXSIM for more information on how I modified the program of Scott-Clayton to work with the March CPS data of 1985 to 1987.

[14]Most of these 47 degree fields are exact copies of the 2-digit CIP broad field code, but certain fields could naturally be subdivided into a few smaller fields. For instance, the 2-digit broad field "Engineering" was split into five different fields such as "Civil Engineering" and "Mechanical Engineering".

2.5 Methodology

Keeping individuals' before-tax income constant, TRA86 changed the average after-tax lifetime income of a major. To calculate by how much these major-specific incomes changed due to TRA86 alone, I rely on pre-TRA86 income data from a sample of college graduates.[15] In particular, I calculate average after-tax lifetime income by major under the 1985 tax law (before the tax change) and under the 1988 tax law (after the tax change) with income data from 1985 and use the difference as a measure of the mechanical effect of TRA86 on expected after-tax lifetime income by major. By not relying on income data from after the tax change, I can isolate the change in expected lifetime income that is due to the tax change alone. Using post-TRA86 income data as well is problematic because this data among other things also captures the effects that economy-wide time-trends in wage growth across sectors have on major-specific lifetime incomes[16]. One limitation of this approach though is that I might be slightly overstating the relative change in lifetime earnings across majors due to TRA86. By affecting the composition of completed college majors, TRA86 could have general equilibrium effects that, in the long-run, could cause the before-tax lifetime earnings to decrease slightly in majors gaining the most from TRA86.[17]

[15] As shown by Skyt Nielsen and Vissing-Jorgensen (2006), individuals might also care about the variance of expected lifetime earnings by major. This variance might also have changed in response to the TRA. I ignore this possibility here as the limited sample size of the available data does not allow me to accurately calculate the change in the variance of expected lifetime earnings by major due to TRA86.

[16] Over the period 1985-1988, nominal personal income per capita grew by 17.4 percent (Feldstein (1995)). At the same time, the 1980s was a period of rising overall wage inequality and changing (before-tax) relative wages of college graduates in different fields of study (Katz and Autor (1999)).

[17] Another reason for using pre-TRA86 earnings information is that changes in total annual after-tax earnings, as are observed in the March CPS, can also be a result of changes in hours worked (e.g. overtime) in response to TRA86. However, if some people change their working hours in response to the tax-reform act, I would not only need to know the change in utility associated with the change in income due to the change in labor supply, but also the (unobservable) change in utility associated with the decrease in leisure time available to that individual. By the envelope theorem changes in individuals' behavior as a result of a tax change have no first-order effect on their utility. Hence, changes in utility from the change in income should almost cancel out with the change in utility from the decrease in leisure. Since the change in utility from decreased leisure is also unobservable, I therefore ignore the effect of changes in labor supply on major-specific SWE and use pre-TRA86

I use the concept of synthetic work-life earnings (SWE) to approximate expected after-tax lifetime earnings by major. This concept is used in many Census publications (e.g. Julian (2012)) and is an "estimate of the amount of money a person might expect to make over the course of a career" (p.1, Julian (2012)) based on the cross-section of people in the labor force today with the same characteristics as that person. I calculate the SWE of college major h under the tax law at time t for males as follows:

$$SWE_h^t = \sum_{s=1}^{3} \sum_{j=yr_0(s)}^{yr_l(s)} \frac{1}{(1+r)^{j-24}} \sum_{k=1}^{K} \pi_{hks} w_{ks}^t \tag{2.1}$$

$$t = 1985,\, 1988 \qquad h = 1, ..., 47$$

I first calculate the average after-tax income for males in age-group s, with major h and under tax law t: $\sum_{k=1}^{K} \pi_{hks} w_{ks}^t$. The construction of this average after-tax income is explained in more detail below. I distinguish between three age-groups: 25-34, 35-44 and 45-64.[18] For each age-group I then multiply these average incomes by the number of years that the age-group represents in a 40-year working-life: $\sum_{j=yr_1(s)}^{yr_l(s)} \frac{1}{(1+r)^{j-24}}$. $yr_1(s)$ is the first year of age-group s (e.g. 25 for the first age-group) and $yr_l(s)$ is the last year of the age-group (e.g. 34 for the first age-group). r is the discount rate and equals 0 percent.[19]. To get the SWE by major I sum these three age-group and major-specific cumulative earnings measures.

In the March CPS, which I use as the source of the pre-TRA86 income data, I do not observe the college major of college graduates. To construct the measure of expected average after-tax income by age-group and major under the two tax-law regimes $\left(\sum_{k=1}^{K} \pi_{hks} w_{ks}^t\right)$ I therefore first calculate the distri-

earnings information.

[18]Due to sample size issues, I can only distinguish between three age groups

[19]This simplifying assumption makes the results easier to interpret as it means I multiply average after-tax incomes in age-group s and major h by the number of years in age group h when calculating SWE for major h.The results discussed below change little when using a different discount rate. Alternatively, since I do not take into account potential real growth in earnings in the future when calculating SWE by major, a discount rate of 0 can also be interpreted as meaning that $r - g = 0$, where g is the growth rate of average earnings over time.

bution of college-graduates across various occupations (k) by age-group and college major using the 1993 National Survey of College Graduates. In calculating π_{hks}, the probability a person in age-group s and with college major h is in occupation k, I distinguish between 59 different occupations. Being a student or being unemployed are included as separate occupations.[20] The other 57 occupations consist of 32 occupations of which all except 7 are divided into two categories based on individuals' highest attained degree.[21] [22]

Next, I use 1985, 1986 and 1987 March CPS surveys combined with NBER's TAXSIM program to calculate the average after-tax income (from all sources) of college graduates in occupation k and age-group s under the two different tax laws. I start with before-tax earnings of college graduates between the ages of 25 and 64 from the 1985, 1986 and 1987 editions of the March CPS.[23] For each individual, I then calculate the total after-tax income of *the tax filing unit* to which that person belongs under both the 1985 and 1988 tax laws using NBER's TAXSIM program. I define total after-tax income as: income from *all sources (including unearned income)*[24]- federal income tax liabilities - state tax liabilities - 0.5*payroll taxes.[25] w^t_{ks} is then the average

[20]In this paper, synthetic work-life earnings are an estimate of what the average person can expect to earn in a major if he is willing to work full-time for 40 years. As a result, in both the March CPS and 1993 NSCG, I exclude people that are out of the labor force and are not students, and part-time and/or part-year workers working part-time and/or part-year because they wanted to work part-time and/or part-year. Doing so minimizes the possibility that the calculated major-specific expected lifetime income is affected by the fact that preferences for part-time (or part-year) employment might vary across people majoring in different subjects.

[21]For instance, the occupation "Secondary School Teachers" is further subdivided into "Secondary School Teachers with a Bachelor as highest degree" and "Secondary School Teachers with at least a Master's degree". 7 occupations could not be subdivided in such a way due to small sample size or because the occupation requires people to have a graduate degree (e.g. the category "Lawyers and Judges").

[22]See the appendix section 2.9.3 on the National Survey of College Graduates for a detailed discussion of some of the limitations of using the 1993 NSCG survey to approximate π_{hks} in 1985.

[23]In both surveys, I also drop people in the Armed Forces. In the March CPS, I drop anybody filing jointly with somebody over the age of 64 as well.

[24]This includes all reported sources of income in the March CPS: earned income from wages, self-employment or a farm, and unearned income from social security, public assistance, supplemental security income, worker's compensation, interest, dividends and capital gains.

[25]All amounts are adjusted to be in 1985 US dollars. Specifically, to calculate 1985 tax liabilities I adjusted the income amounts reported in the 1985 and 1987 March CPS for

of the total after-tax income of the people in age-group s with occupation k under the tax law at time t.

One issue with SWE_h^t is that the average after-tax incomes for individuals in particular occupations and age-groups, w_{ks}^t, are based on the after-tax incomes of the tax-filing units that those individuals belong to and not just on the income of the individual. That is, w_{ks}^t is not adjusted by the size of the tax filing units. As a result, major-specific SWE will by construction generally be larger the more joint filers are present in the occupations associated with a major. To resolve this issue, I divide the after-tax income calculated above by the income equivalency scale of Short (2001) that is often used to adjust poverty thresholds for family size and composition.[26] [27]

2.6 The Effect on Expected Lifetime Income by Major

Figure 2.1 shows how the average net-of-tax rate (calculated including state and payroll taxes) of a major varies with the 1985 before-tax SWE of the major. The before-tax SWE of a major are calculated in the same way as

inflation using the CPI-U. Similarly, to calculate 1988 federal and state tax liabilities I also adjusted the income amounts reported in the 1985 to 1987 March CPS for inflation with the CPI-U. Afterwards, I converted the estimated 1988 tax liabilities to 1985 dollars.

[26]In the income equivalency scale, single filers with no dependents carry a weight of 1 and two people filing jointly without any dependents have a weight of 1.41 . In general, a child receives less weight than an adult and the first child carries a higher weight for single parents than for two-parent families. This scale is supposed to reflect that people within a household share resources and take advantage of economies of scale.

[27]One last issue with w_{ks}^t is that I also subtract state tax liabilities to arrive at the total after-tax income of an individual in the March CPS. State tax liabilities are calculated based on the state of residence of the individual. w_{ks}^t thus approximates the average after-tax earnings of an individual in age-group s and occupation k at the national level. The state tax laws of states in which an occupation occurs more frequently are thus automatically weighted heavier when calculating average after-tax earnings by occupation and age-group. This approach is valid if students are costlessly mobile. Although this might seem a strong assumption, Groen (2004) finds that attending college in a particular state has only a very modest effect on the probability of working in that state after graduation. This suggests that college graduates are very mobile and that the assumption of costless mobility is not very extreme.

the after-tax ones, but without subtracting the tax liabilities from individuals' annual income. The average net-of-tax rate of a major equals the major's after-tax SWE divided by its before-tax SWE. As is clear from the figure and as can be expected given the progressivity of the tax code, the transformation from before-tax to after-tax SWE by major is nonlinear, both pre- and post-TRA86. The average net-of-tax rate ranges from around 0.65 to 0.70 before TRA86 and from 0.71 to 0.74 after TRA86, and decreases with before-tax SWE.[28] On average, TRA86 caused the average net-of-tax rate across all majors to increase by around 4 percentage points from 0.681 before TRA86 to 0.723 after TRA86.[29]

Figure 2.2 displays the TRA86-induced change in after-tax expected lifetime income by major. For each major, the total length of the bar represents the total change in after-tax SWE for that major. The various colored parts of the bars show what share of this change can be attributed to the various age-groups that contribute to the calculation of SWE. The dashed line at $58.000 is the weighted average across all majors when we use a major's average annual share of all degrees completed between 1985 and 1997 as weight. $58.000 represents an increase in expected after-tax lifetime income of 6.2% relative to the weighted average pre-TRA86 baseline level of $930.000 across all majors. The amount gained varies significantly across majors. The

[28]Majors with the same before-tax level of SWE can face different average net-of-tax rates if the age-earnings profile varies across occupations associated with these majors. For instance, due to the progressivity of the tax code, a major leading to a constant annual before-tax income of $25.000 for 40 years will have a different average net-of-tax rate than a major leading to earnings of $12.500 during the first 20 years of a career and $37.5000 during the last 20 years.

[29]As a side note, the non-linear relationship between before- and after-tax SWE by major illustrates the importance of using after-tax expected lifetime income by major when calculating the sensitivity of major choice to expected earnings. For instance, if students look at after-tax and not before-tax expected earnings when deciding which major to pursue, elasticities of the probability of choosing a particular major with respect to that major's expected earnings will be biased if we use data on before-tax earnings. So far, the literature has solely relied on before-tax earnings. Another point worth noting is that previous studies have often relied on variation in expected earnings by major over time to identify the elasticities of the probability of choosing a major with respect to expected earnings. If students look at after-tax and not before-tax expected earnings when deciding which major to pursue, and tax laws changed during the sample period of some of these studies, the estimated elasticities in these studies could also be biased because they ignore the effect of tax law changes.

Figure 2.1: Average pre- and post-TRA86 net-of-tax rates by before-tax SWE

standard deviation is $9.400, and the increase in earnings varies from a low of $39.000 in Home Economics to a high of $97.000 in Medical Sciences and Nursing.

TRA86 most dramatically reduced top marginal tax rates. This has two main consequences. First, TRA86 affected expected lifetime earnings by major mostly by affecting expected earnings later in life. This is clear from looking at the relative size of the various colored parts of each major-specific bar. In particular, on average 65% of the increase in lifetime earnings by major takes place between the ages of 45 and 64, and only 35% takes place between 25 and 44. Second, the correlation between the pre-TRA86 after-tax SWE by major and the change in SWE induced by TRA86 is very high. This correlation equals 0.93.

Instead of focusing on absolute levels of expected lifetime income by major, some papers (e.g.Long et al. (2015)) have instead focused on expected lifetime incomes by major relative to the mean across all majors. Before TRA86, this measure varied from 0.81 in Protective Services to 1.24 in Medical Sciences including Nursing and had a standard deviation of 0.075. Due to TRA86, this standard deviation increased by 6.1% to 0.079, with majors' rel-

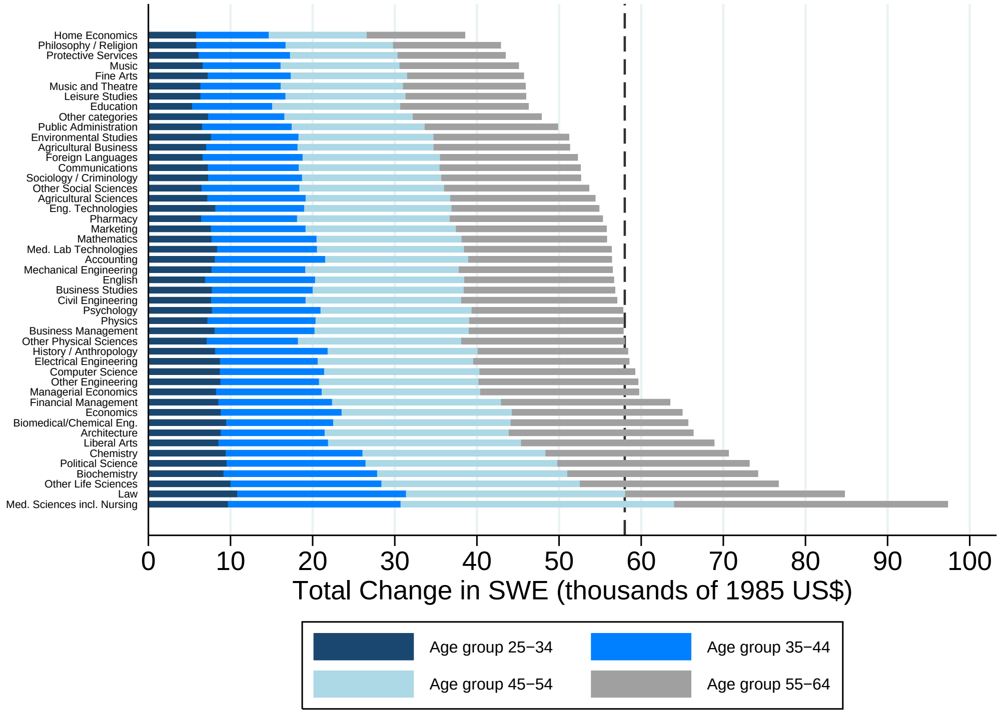


Figure 2.2: Change in after-tax lifetime-income (SWE) by major due TRA86

ative earnings changing by anywhere from -0.01 in Home Economics to 0.025 in Medical Sciences including Nursing.

To place the effect of TRA86 on expected earnings by major into context, we can compare it to variation in expected earnings isolated in other papers on college major choice. For instance, Beffy et al. (2012) look at the before-tax real monthly earnings of French university students during the first five years after leaving the educational system. They find that individuals who entered into the labor market in 1998, when France experienced an economic expansion, on average earned 4.9 percent higher real wages during the first five years in the labor market than individuals entering in 1992, when France experienced a downturn. The average increase in after-tax expected lifetime earnings of 6.2% induced by TRA86 is thus comparable to the average change in earnings induced by graduating at different points in the business cycle.[30]

[30]In fact, the mean effect of TRA86 on expected earnings could even be much larger than the effect of graduating at different points in the business cycle. TRA86 is likely to represent a permanent change in after-tax expected lifetime income by major, whereas the effects of

However, in contrast to TRA86, Beffy et al. find that the moment of graduation has very heterogeneous effects across majors. For instance, science (respectively law, economics and management) majors had 13.5% (respectively, 10.4%) higher earnings in 1998 than in 1992 whilst humanities and social science majors had 4.2% lower earnings in 1998 than in 1992. Beffy et al. exploit this change in relative earnings across majors over time to empirically identify the effect of expected earnings on college major choice. As TRA86 had much less of a differential impact on major-specific expected earnings, it will be more difficult to empirically detect its impact on college major choice.

The variation in relative expected lifetime income across majors due to TRA86 can also be compared to the long-term trends in major-specific returns identified by Altonji et al. (2014). Altonji et al. (2014) use data from the 1993 and 2003 National Survey of College Graduates and from the 2009 to 2011 American Community Survey to estimate (before-tax) wage premiums for 51 different majors[31]. They find that the standard deviation of these wage premiums increased by 24.5% from 1993 to 2003 and, in the longer run, by 13% from 1993 to 2011. Absent tax law changes, the changes in *after-tax* wage premiums by major over this period are likely to have been slightly smaller due to the progressivity of the tax code. I showed that TRA86 caused the standard deviation of the distribution of expected after-tax lifetime income by major relative to the mean across all majors to increase by 6.1%. When compared to the estimates of Altonji et al. (2014), this suggests that tax changes such as TRA86 can significantly contribute to long-run changes in the distribution of after-tax wage premia by major such as those observed between 1993 and 2011.

graduating at a different point in the business cycle might not persist beyond the first few years in the labor market. See also Altonji et al. (2016a) and Oreopoulos et al. (2012) for a further discussion of the impact of graduating at different points in the business cycle on initial earnings.

[31] In particular, they regress log annual earnings for full-time workers with a college degree on a number of individual characteristics and college major dummies. The coefficients on these dummies are the major-specific wage premia and capture both the causal impact of the major and differences across majors in skills and abilities determined prior to college.

2.7 The Effect on College Major Choice

Empirically identifying the effect of TRA86 on the composition of completed college majors using available data runs into a number of difficulties. First, I only have data on the aggregate number of bachelor degrees completed in each major over time. Besides college major choice, TRA86 could also have affected other margins of educational attainment such as whether an individual chooses to go to college, what college he/she attends[32] and whether he/she graduates from college. Such effects will change the composition of students completing college over time, which could also affect the distribution of completed college degrees by major over time. Using aggregate data on the number of bachelor degrees completed in each major, I am not able to control for any such possible changes in the composition of college graduates over time. The aggregate data would for instance allow me to test the prediction following from Berger's (1988) model of college major choice that a major experiencing a relatively larger increase in lifetime income due to TRA86 should see a relatively larger percent increase in the fraction of undergraduate degrees completed in that major in years following TRA86. However, given the inability to control for changes in the composition of college graduates, it is unclear how reliable this estimate of the effect of TRA86 on the types of majors completed would be.[33]

Second, the nature of the variation in expected after-tax lifetime income by major induced by TRA86 makes it difficult to identify a possible effect on the composition of completed college majors. When compared to variation used in other studies, the changes in relative expected lifetime earnings by major due to TRA86 are small, with majors' relative earnings changing by anywhere from -0.01 to 0.025. Moreover, the personal income tax provisions of TRA86 were phased in over a period of three years. As a result, the change in majors' relative expected after-tax lifetime income will have materialized

[32]For instance, an increase in parental income due to TRA86 could cause some students to attend private instead of public colleges.

[33]This empirical strategy is pursued in appendix 2.9.4. Using this approach, I fail to find consistent evidence of a significant impact of TRA86 on the composition of completed college majors.

gradually, making the corresponding effect on the types of college majors completed harder to detect as this effect occurred over multiple years. In addition, the very strong positive correlation (0.93) between major-specific baseline expected after-tax lifetime income and the change in this income induced by TRA86 makes it impossible to distinguish between changes in the composition of completed college majors that are due to TRA86 and those that are due to other factors correlated with baseline expected lifetime income by major. This is an important limitation as the 1980s were a period of rising overall wage inequality and changing (before-tax) relative wages of college graduates in different fields of study (Katz and Autor (1999)). It is possible that these time trends in wage growth caused students to shift towards completing majors with high baseline expected lifetime earnings. This would cause the estimated effect of TRA86 to be biased upwards.

Given the difficulties in empirically identifying the effect of TRA86 on the distribution of completed college majors, I simulate how the actual (pre-TRA86) 1986 composition of completed college degrees of males would have looked if the 1985/1986 cohort had faced the post-TRA86 distribution of after-tax expected lifetime income by major instead of the pre-TRA86 distribution. To do so, I use various estimates of the elasticity of college major choice with respect to major-specific earnings from the literature. Looking at the change in the 1986 distribution assumes that TRA86 does not affect the number of undergraduate degrees completed. I make this assumption because current estimates on the elasticity of college-major choice with respect to expected earnings cannot be used to also simulate how TRA86 changed the composition of college graduates and, correspondingly, the types of majors completed.

Estimates of the elasticity of college major choice with respect to expected earnings vary widely across studies. Since all these studies differ in many ways, there is no one value that is clearly best to use.[34] Most im-

[34]For example, elasticities can differ based on the number of majors a study distinguishes between, the time period and country on which a study is based and whether or not a study includes students of both genders. In addition, elasticities can also differ based on the sample of students included in a study. Some studies look only at those who go on to receive a college degree whilst others look at the choices of entering college students who do not all go on to receive an undergraduate degree.

portantly, elasticities will differ depending on the definition of earnings being used. Some studies use population-based average earnings of college graduates with a particular major at the time that students are in college (e.g. Long et al. (2015)) whilst others use students' individual-specific expected lifetime income across majors, generated with forward-looking data, under the assumption that students have rational expectations (e.g. Beffy et al. (2012)).

Most closely related to my study are the estimates of Long et al. (2015). They calculate average before-tax earnings by major from 1987 to 2011 for 36 majors by relying on a time-invariant major-occupation distribution from the 2009 to 2011 American Community Survey and occupation-specific time-varying wage data from the CPS Outgoing Rotation Groups. They find that a 1% increase in major-specific earnings relative to the mean across all majors in year t is associated with a 0.64% increase in the share of majors completed in that subject in year t+3. The main limitation of this study is though that the estimates are purely correlational.

On the other hand, papers using expected lifetime income calculated separately for each student and major under the assumption of rational expectations and with forward-looking data find different elasticities. For instance, Beffy et al. (2012) estimate a sequential schooling decisions model with student-level French data in which students among other things compare the rationally expected earnings across majors. Using, as mentioned before, variation in expected earnings induced by the business cycle to identify earnings elasticities, they find earning elasticities ranging from 0.09 in sciences to 0.14 in humanities and social sciences.[35] On the other hand, Montmarquette et al. (2002) use data from the 1979 National Longitudinal Survey of Youth to estimate a model in which students have rational expectations and take into account the probability of completing a major and the starting earnings of a major upon graduation. Using mixed multinomial logit and probit models,

[35]The smaller elasticities could not only be the result of the definition of earnings being used. Beffy et al. (2012) also only distinguish between 3 majors and only have data on students during the first five years after leaving the educational system. As a result, they cannot tell whether the elasticities identified in their paper are with respect to a permanent or transitory change in major-specific earnings.

they find expected earning elasticities of around 5. It is unclear how reliable these estimates are though as the earnings variables in this study are estimated without correcting for self-selection into majors.

In figure 2.3 I show the simulated change in the pre-TRA86 1985/1986 distribution of the number of college degrees completed by field when using an elasticity of 0.67[36] [37]. In this figure majors are ordered by the percent change in their relative expected after-tax lifetime income induced by TRA86, with majors experiencing the largest decrease listed at the top and majors experiencing the largest increase listed at the bottom. As a result, non-monotonic variation in the change in the number of degrees completed as we move from the top to the bottom is driven by differences in the initial number of students completing a college degree in a particular major. Figures using an elasticity of 0.1 or 5 will look the same, the only exception being that the simulated changes will respectively be 6.75 times smaller or 7.4 times larger. When using an elasticity of 0.67, the simulations estimate that close to 1200 individuals would have completed a different major under the post-TRA86 tax law. This amounts to 0.25% of all majors completed by males in 1985/1986. When using elasticities of 0.1 or 5, these effects are respectively 0.04% and 1.83%. Thus, even under the most extreme assumptions about the elasticity, the share of undergraduate degrees affected by TRA86 is still small. As can be seen in figure 2.3, in terms of the absolute number of degrees completed, the major Education is most negatively impacted by TRA86, whereas Political Sciences,

[36] Let ϵ_R^A denote the elasticity of the share of students with a particular major with respect to the relative *after-tax* expected lifetime income of that major. One can show that $\epsilon_R^A = \epsilon_R^B \left(\frac{Y_R^A}{Y_R^B} \cdot \frac{1}{\delta Y_R^A / \delta Y_R^B} \right)$, where ϵ_R^B equals the same elasticity with respect to the relative *before-tax* expected lifetime income in a major and Y_R^A and Y_R^B equal respectively the after- and before-tax relative expected lifetime income of a major. Across all majors in my dataset, and both before and after TRA86, $\frac{Y_R^A}{Y_R^B}$ ranges from a minimum of 0.957 to a maximum of 1.034. $\delta Y_R^A / \delta Y_R^B$, which I approximate separately before and after TRA86 by regressing Y_R^A on Y_R^B, equals 0.85 before TRA86 and 0.91 after TRA86. These estimates suggest that by using ϵ_R^B as a proxy for ϵ_R^A, I am underestimating ϵ_R^A by anywhere from 5 to 21%. This relatively minor approximation error will not dramatically affect the nature of the simulation results discussed below.

[37] To calculate this change I take the share of majors completed in major m in 1985/1986 and multiply it by the percent change in relative expected after-tax earnings in this major due to TRA86, the elasticity and the total number of degrees completed in 1985/1986

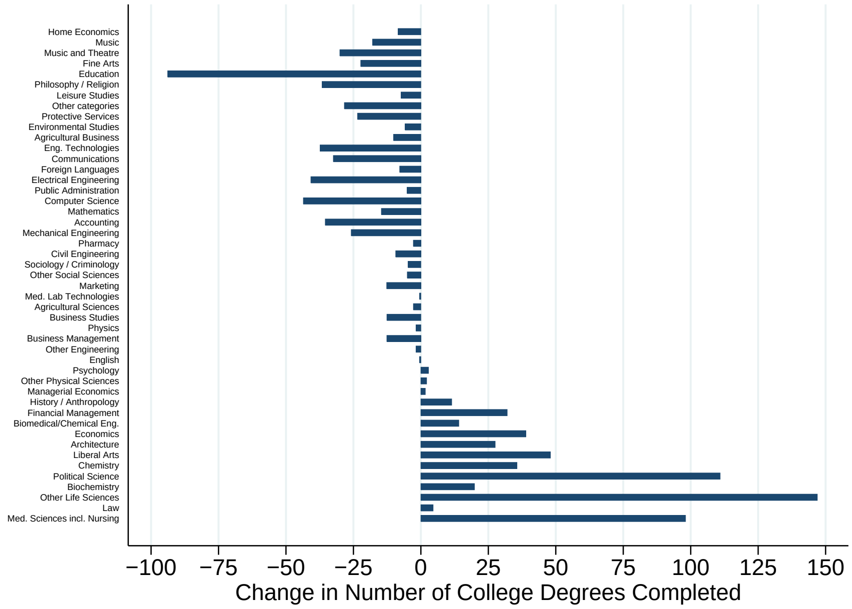


Figure 2.3: Simulated effect of TRA86 on the types of degrees completed by the 1985-1986 graduating cohort of males

Other Life Sciences and Medical Sciences including Nursing see the largest positive changes in the number of completions due to TRA86.

Lastly, using the framework of Hendren (2016), I analyze the welfare impact of the simulated change in the composition of completed college degrees as a result of TRA86. Under some simplifying assumptions discussed in detail in appendix 2.9.1, we can approximate this welfare impact in Hendren's framework by measuring the change in the amount of income tax that a cohort pays over their working life as a result of the TRA86-induced change in the composition of completed college majors in this cohort. To do so, I calculate average total federal and state personal income tax payments by major over a 40-year working life using same method and data that I used to calculate major-specific SWE. I then use the change in the composition of completed college degrees simulated above to calculate by how much government tax revenue would change if this simulated change had taken place.

I find positive welfare effects equal to 62.7 million 1985 US dollars when using an elasticity of 0.67 in the simulation exercise.[38] To put these welfare gains into perspective, 62.7 million 1985 USD equals 0.04% of the expected lifetime income tax payments of the 1985/1986 cohort under the pre-TRA86 tax law.[39] The welfare gains are thus small relative to overall tax payments. Even when using an elasticity of 5 in the simulation exercise, this welfare effect would be only 0.28% of overall expected tax payments for the 1985/1986 cohort of male college graduates.

2.8 Conclusion

The potential for changes in personal income tax laws to affect the choice of educational type, and in particular the choice of college major, has not been studied empirically. This is an important topic however as it has implications for the welfare effects of personal income taxes. In this paper, I combine information from the 1993 NSCG on the distribution of occupations by major and from the March CPS on average after-tax incomes by occupation to first estimate the impact of TRA86 on expected after-tax lifetime earnings by major. I then use estimates of the elasticity of college major choice with respect to major-specific expected earnings from other studies to simulate how the differential TRA86-induced change in expected after-tax lifetime earnings across majors could have affected the distribution of completed college majors.

I find that the average major experienced an increase in after-tax expected lifetime income of $58.000 or, equivalently, 6.2% due to TRA86. The differential impact across majors is small. Relative to the mean across all majors expected earnings for a particular major changed between -1.2 to 2.0%. Due to the limited differential impact across majors, at most 0.25% of males graduating after TRA86 are estimated to have completed a different major due to TRA86. These results show that personal income tax reforms can have large effects on the mean level of expected earnings associated with a given

[38]To evaluate the welfare effects using different elasticities one can simply scale this amount up or down by the elasticity relative to 0.67.

[39]These expected lifetime income tax payments of the 1985/1986 cohort are calculated using the actual observed 1985/1986 distribution of completed college degrees by major.

level of education. However, the differential impact across educational types is not likely to be very large. In addition, the lack of a differential impact across educational types implies that current short-run estimates of the elasticity of taxable income with respect to the net-of-tax rate are not heavily understating the long-run elasticity. These current estimates, which rely on the short-run response of taxable income to changes in marginal tax rates, would not be heavily impacted by a small change in the composition of completed college majors in the long run as a result of a tax change.

In this paper I simulated the effect of TRA86 on the distribution of completed college majors. Further empirical studies are needed to determine whether this simulated effect did actually take place and whether there actually exists a link between the choice of college major and tax reforms in the data. To do so, future studies need to address two empirical challenges, namely changes in the composition of college graduates over time as a result of tax changes and the high correlation between tax-induced changes in expected lifetime income by major and baseline levels of expected lifetime income. To address the former, one needs large individual-level datasets that contain data on the college major choices of many cohorts of college students. To address the latter, one could look to a country other than the U.S. in which both large regressive and progressive tax reforms have been implemented. Looking at various such tax reforms in one study would allow one to more clearly separate the impact of tax reform-induced changes in expected lifetime income from time-trends in college major completions that are correlated with baseline levels of expected lifetime income by major.

Chapter 2, in part, has been submitted for publication of the material. The dissertation author was the sole author of this paper. Gaastra, Sieuwerd.

2.9 Appendix

2.9.1 Welfare Effects

In this appendix section, I will analyze the welfare impact of the change in the composition of completed college degrees as a result of TRA86 using the framework of Hendren (2016).[40] In Hendren's framework, individual welfare is measured by an individual's willingness to pay out of his/her own income for a policy change. The marginal welfare impact to a given individual of pursuing a policy consists of the sum of three terms: (1) the causal impact of the individual's behavioral response to the policy on the government budget, (2) the individual's willingness to pay for the change in publicly provided goods and services as a result of the policy, and (3) the net transfers to the individual as a result of the policy. Since TRA86 was designed to be budget-neutral and was not directly linked to the public provision of goods and services, I will ignore these last two terms and only focus on the impact of the individual's behavioral response to TRA86 on the government budget.

The effect of the behavioral response to a policy on the government's budget matters because of the envelope theorem. The envelope theorem tells us that an individual's behavioral response to a marginal policy change does not have a direct first-order effect on his utility. However, in the presense of such distortions as personal income taxes that create a wedge between private and social costs and benefits, these behavioral responses impose a resource cost on society. This resource cost is captured by the effect of the behavioral response to a marginal policy change on the government budget.

To aggregate the marginal welfare impact of a marginal policy change across all people, I assume that the composition of completed college degrees does not impose non-pecuniary externalities on individuals. If this were the case, I would need to know individuals' marginal rate of substitution between the composition of completed college majors and income in order to be able

[40]The standard framework for analyzing welfare effects, the marginal excess burden framework, is not suitable for my analysis as it requires the decomposition of behavioral responses to policy changes into income and substitution effects. In contrast, Hendren's framework requires only the total causal effects of the policy change.

to value the change in this externality due to TRA86. In addition, I assume that the cost of training a student does not depend on his/her choice of major. If training costs do vary by major, and students would not pay higher tuition to offset the cost of more expensive majors, I would also need to include the change in the training costs due to the change in the composition of completed college majors in my welfare measure.

For the actual welfare analysis, let the parameters of the tax law be a function of some variable θ. When θ equals zero, the tax law is equal to the pre-TRA86 tax law. When θ equals 1, the tax law equals the post-TRA86 tax law. I assume that the tax law parameters are a continuously differentiable function of θ. Under this assumption, the move from $\theta = 0$ to $\theta = 1$ traces out a smooth path of tax policies from before TRA86 to after TRA86. Let N_t denote the number of students graduating from college in year t, which is assumed to be unaffected by TRA86, and $P_{ht}(\theta)$ be the fraction of students graduating in year t who majored in field h as a function of the variable θ. Under the framework and assumptions outlined here, the equivalent variation of TRA86 that is only due to the change in the composition of completed college degrees for cohort t can be written as:

$$\Delta W_t = N_t * \left(\sum_h \left(\int_0^1 \frac{dP_{ht}(\theta)}{d\theta} T_{ht}(\theta) d\theta \right) \right) \qquad (2.2)$$

In this equation, $T_{ht}(\theta)$ denotes the average lifetime (state and federal) income tax payments of marginal entrants into major h[41]. These average lifetime income tax payments are a function of the tax law (θ) and can vary across cohorts (t). $\frac{dP_{ht}(\theta)}{d\theta}$ is the marginal rate at which the tax policy induces students to complete major h instead of another major. ΔW_t thus represents the change in total government income tax revenue that cohort t pays over their working life and that is a result of the change in the composition of completed college

[41]I ignore any taxes paid on compensation that is not taxed as wages and salaries as well as sales tax payments. Moreover, I ignore payroll taxes paid. The reason I do not include payroll taxes here is that the payroll taxes that an individual pays today are directly related to higher benefits in the future. In other words, an increase in government revenue from payroll taxes today will partly be offset by higher costs of benefits in the future.

majors in cohort t as a result of TRA86.

I estimate equation 2.2 for the cohort of 1985/1986 college graduates. This is the last pre-TRA86 cohort, meaning that in contrast to later cohorts the types and number of college degrees completed by this cohort cannot have been affected by TRA86. As a result, we can use the observed value of the share of majors completed in field h in 1985/1986 as an estimate of $P_{h86}(0)$ in equation 2.2. In section 2.7, I simulated the counter-factual distribution of college graduates across majors for this cohort had it been affected by TRA86. This counter-factual distribution can be used in equation 2.2 to approximate $P_{h86}(1)$, the counter-factual share of college graduates in 1985/1986 with major h under TRA86.

To implement equation 2.2 empirically for the 1985/1986 cohort, I need to make a number of further simplifying assumptions. First, I cannot obtain good estimates of the average lifetime income tax payments of marginal entrants into major h for the 1985/1986 cohort, $T_{h86}(\theta)$, as a functon of θ for values of θ other than zero. For instance, estimating $T_{h86}(1)$ properly would need to account for the general equilibrium effects of TRA86. This cannot be done as it will take up to 40 years after TRA86 before the composition of completed college degrees in every cohort of the working-age population reflects the effect of TRA86. Since the composition of completed college degrees will affect relative wages in occupations, this also means that it will take up to 40 years after TRA86 before the full general-equilibrium effects of TRA86 are visible in the relative wages across occupations. To simplify the analysis, I therefore use 1984, 1985 and 1986 March CPS income data to calculate the average total federal and state personal income tax payments by major over a 40-year working life using the same method that I used to calculate major-specific SWE. That is, I approximate $T_{h86}(1)$ by applying post-TRA86 tax laws to pre-TRA86 income data and $T_{h86}(0)$ by applying pre-TRA86 tax laws to pre-TRA86 income data.[42] As was the case with the change in major specific

[42] An alternative approach would be to just use $T_{h86}(0)$ as an estimate of $T_{h86}(1)$ as well. As my approximation of $T_{h86}(1)$ is smaller than $T_{h86}(0)$ for every major, this will result in even larger welfare gains than those reported below. To be consistent with the discussion is section 2.7, when calculating $T_{h86}(0)$ and $T_{h86}(1)$ I use a discount rate of 0% and divide an

after-tax SWE, this approximation of average lifetime income tax payments by major ignores the effects that economic growth after 1985 could have had on major-specific lifetime income tax payments.

Lastly, for the empirical implementation I also assume that $\frac{dP_{h86}(\theta)}{d\theta}$ is constant and that $T_{h86}(\theta)$ is a linear function of θ. I thus implement equation 2.2 empirically as:

$$\Delta W_{86} \approx N_{86} \left(\sum_h (P_{h86}(1) - P_{h86}(0)) * \left(\frac{T_{h86}(1) + T_{h86}(0)}{2} \right) \right) \tag{2.3}$$

I set the fraction of all college degrees completed by the 1985/1986 cohort that are in major h before TRA86, $P_{ht}(0)$, equal to the actual value of P_{h86} observed in the data. I calculate $P_{ht}(1)$ using the results of the simulation exercise described in section 2.7.

Under the assumptions outlined above, the welfare effects are positive and equal 62.7 million 1985 US dollars when using an elasticity of 0.67 in the simulation exercise of section 2.7. To evaluate the welfare effects using different elasticities one can simply scale this amount up or down by the elasticity relative to 0.67. To put these welfare gains into perspective, 62.7 million 1985 USD equal 0.04% of the expected lifetime income tax payments of the 1985/1986 cohort under the pre-TRA86 tax law ($\sum_h P_{h86}(0) * T_h(0)$). The welfare gains are thus small relative to overall tax payments. Even when using an elasticity of 5 in the simulation exercise, the welfare effects would be only 0.28% of overall expected tax payments for the 1985/1986 cohort of male college graduates.

2.9.2 TAXSIM

To get the (Unicon) March CPS data of 1985 to 1987 to work with the TAXSIM program, I modified the program of Judith Scott-Clayton that is available on the TAXSIM website of the NBER. The program of Scott-Clayton adopts the Unicon versions of the 2003 and 2006 March CPS to work

individual's tax payments by 2 if he/she files jointly.

with TAXSIM.

Unfortunately, the 1985 to 1987 versions of the March CPS are less detailed than later editions. Income from veterans' payments (VT) and worker's compensations (WC) cannot be separated from income from unemployment insurance (UI) until the March CPS of 1988. As UI compensation is part of federally taxable income and WC/VT income is not, it is important to separate income from these sources. To do so, I used regression mean imputation based on data from the 1988 March CPS to predict the share of income coming from either UI or WC/VT if an individual reported receiving income from both UI and WC/VT.

Similarly, the 1985 to 1987 editions of the March CPS do not distinguish between income from dividends and income from rent, royalties, estates and trusts. Although these sources of income are taxed in the same way at the federal level, they are not always taxed in the same way at the state level. Since state income tax paid is tax deductible, this means that the division of income between dividends and rent, royalties, estates and trusts is important for determining federal tax liabilities. Here, I used mean imputation instead of mean regression imputation. Using again data from the 1988 March CPS, I calculated the mean division between income from dividends and income from rent, royalties, estates and trusts for 7 different ranges of combined total dividend and rental income. I then used these mean divisions of income to divide up people's combined total rental and dividend income in the 1985 to 1987 March CPS into dividend and rental income.

Another problem with March CPS data is that reported income amounts are subject to top-coding. This can cause me to understate the change in average lifetime earnings by major due to TRA86, as TRA86 had the largest impact on people at the top of the income distribution. To correct for the potential biases introduced by top-coding, I replace the top-coded income amounts in the March CPS by those reported by Larrimore et al. (2008) and Armour et al. (2016). These papers report cell-means for top-coded income amounts from various sources for every year after 1976 using internal March CPS data. The internal March CPS data use higher top-codes for each income source than the public March CPS data. As a result, these cell-means are a better approximation of the true average income amounts above the top-code than just using the top-coded value of the public CPS data. However, the cell-means calculated in this way are still subject to the internal CPS top-code and will therefore still understate the actual mean income amounts of people subject to the top-code. To deal with this issue, Armour et al. (2016) go one step further and impute the true cell-mean by relying on both the internal March CPS data and Pareto estimation methods. This results in higher cell-means thanLarrimore et al. (2008), but these cell-means are not available for all types of income. I use estimates of Armour et al. (2016) when available, and otherwise rely on the estimates of Larrimore et al. (2008).

Lastly, the March CPS lacks information on many of the large itemized deductions that many people claim. For instance, it has no information on people's charitable contributions and home mortgage interest payments. As a result, when calculating tax liabilities, I had to assume that almost everybody claimed the standard deduction. In only a few cases did the sum of possible itemized deductions on which there was information in the March CPS exceed the standard deduction.

2.9.3 National Survey of College Graduates

I use the 1993 National Survey of College Graduates to create a major-specific distribution of people across occupations for three age-groups. Ideally, I would use a survey like the NSCG from before the tax reform, as the major-specific distribution of people across occupations might have changed over time due to structural changes in the economy. Unfortunately no survey like the NSCG exists from before 1993.

It is important to note that the distribution of π_{hks} that I use is the same both before and after the tax reform. This is not an entirely innocuous assumption. For instance, Powell and Shan (2012) and Gentry and Hubbard (2004) find that marginal tax rates and the progressivity of the tax law have a very small, but statistically significant effect, on occupational choice. However, if some people change their jobs in response to the tax-reform act, we would not only need to know the change in utility associated with the change in income due to the job change, but also the change in utility associated with the difference in the non-monetary attributes of the two occupations. We cannot measure the latter part and, as I also do not have access to a time-varying distribution π_{hks}, therefore do not consider the effect of TRA86 on changes in the distribution of π_{hks}. Ignoring the effects of these job changes should be relatively harmless though. By the envelope condition changes in individual's behavior as a result of a tax change have no first-order effect on their utility, meaning that changes in the utility from the change in income should almost cancel out with changes in the utility from the change in the value of the non-monetary attributes.

In addition, using the observed distribution of π_{hks} from the cross-section of people who majored in field h and are in the labor force today to predict the future career path of people who graduate today with the same major has one major drawback. The distribution of π_{hks} by age-group can differ mainly for two reasons. Firstly, in general, when people get older their jobs are generally less connected to their college major, resulting in a distribution of π_{hks} that varies by age-group. Secondly, because of business cycles and structural changes in the economy over time, two people graduating in different years with the same major might face different sets of potential occupations over their working life. To the extent that one's career path is affected by these changing economic conditions, this means that past business cycles and past structural changes in the economy have had a large effect on the distribution of π_{hks} by age-group observed today. These "cohort effects" are different for each cohort of graduating students. There is no guarantee that the cohort effects in the past are on average the same as those that people graduating today will experience. When calculating π_{hks} I unfortunately cannot distinguish between these "cohort effects", that I would not like to include, and the "age effects" that I would like to include.

2.9.4 Empirical Estimation

In this appendix, I attempt to empirically identify the effect of TRA86 on the composition of completed college majors using available data. As described in the main text, I run into a number of difficulties when doing so. As a result, it is not surprising that I fail to find consistent evidence of a significant impact of TRA86 on the composition of completed college majors.

Basic model

Assume, following Berger (1988), that a college student i chooses a field of study that maximizes his/her lifetime utility (V). Utility for person i in major h can be expressed as a function of the characteristics of the major (Y_{ih}), the characteristics of the individual (Z_i) and an unobserved random component (u_{ih}) that is assumed to follow a type I extreme value distribution:

$$V_{ih} = \beta Y_{ih} + Z_i \delta_h + u_{ih} \tag{2.4}$$

Assume that the only characteristics of the major that individuals consider are expected lifetime earnings in that major (Y_{ih}). The Z-vector contains individual specific variables that control for differences in tastes and investment costs across people. To the extent that preferences for other major-specific characteristics (e.g. effort cost of completing the major and nonpecuniary returns from the major) depend on these individual characteristics, Z_i also controls for these major-specific characteristics. Alternatively, one could include these other major-specific characteristics explicitely in the model above. The results below do not depend on this assumption.

Since we assumed that u_{ih} follows a type I EV distribution, we have a conditional logit model for which it is true that:

$$ln\left(\frac{P_{ih}}{P_{ik}}\right) = \beta(Y_{ih} - Y_{ik}) + Z_i(\delta_h - \delta_k) \quad \forall k \neq h \tag{2.5}$$

In this equation, P_{ih} denotes the probability that individual i chooses major h. k is another major not equal to h. Absent a behavioral response to the tax change (e.g. ignoring the effects of a possible change in one's future labor supply[43]), TRA86 caused the expected lifetime earnings of major h for person i to change by $\Delta Y_{ih} = Y_{ih}^{post} - Y_{ih}^{pre}$ whilst leaving the non-pecuniary returns of the major, as captured by $\delta_h Z_i$ in equation 2.4, unchanged. In the absence of other changes in the economy this means that for person i and majors h and k:

$$ln\left(\frac{P_{ih}^{post}}{P_{ik}^{post}}\right) - ln\left(\frac{P_{ih}^{pre}}{P_{ik}^{pre}}\right) = \beta(\Delta Y_{ih} - \Delta Y_{ik}) \tag{2.6}$$

Summing over all $k \neq h$ and dividing by (H-1), where H is the total number of majors, we then get that:

$$ln\left(\frac{P_{ih}^{post}}{P_{ih}^{pre}}\right) - \frac{1}{H-1}\sum_{k\neq h} ln\left(\frac{P_{ik}^{post}}{P_{ik}^{pre}}\right) = \beta\left(\Delta Y_{ih} - \frac{1}{H-1}\sum_{k\neq h}\Delta Y_{ik}\right) \tag{2.7}$$

The left-hand side is approximately equal to the percentage change in the probability of choosing major h minus the average percentage change in the probability of choosing any of the other majors due to TRA86. This relative percentage change in the probability of choosing major h is proportional to the change in expected lifetime earnings in major h relative to the average change in expected lifetime earnings in all the other majors, as shown on the right-hand side. The model thus has the intuitive prediction that for a given individual majors that see a relatively larger increase in expected lifetime earnings due to TRA86 should see a relatively larger increase in the probability of being chosen.

I do not have a measure of ΔY_{ih} for each individual enrolled in college around the time of TRA86. Instead, my measure ΔY_h does not vary across individuals and is an estimate of the change in SWE in field h as a result of TRA86. Similarly, I do not have data on the college major choice of individuals. I only have data on the number of completed college degrees by major and year. To estimate equation 2.7 using this data, I replace P_{ih} by P_h, the fraction of all completed college degrees that are in major h, and replace ΔY_{ih}

[43]As discussed in section the main text, the reason I ignore these behavioral responses is that I would then also need to know the change in utility associated with job changes and/or changes in hours worked. These are both unobservable.

by ΔY_h, my estimate of the change in SWE earnings of major h as a result of TRA86.

With these aggregate variables in place of the individual-level variables, equation 2.7 will exactly equal the original version of equation 2.7 when averaged over all I individuals under three conditions. First, I apply a first-order approximation to the log terms on the left-hand side of equation 2.7, approximating $ln\left(\frac{P_{ih}^{post}}{P_{ih}^{pre}}\right)$ by $\frac{\Delta P_{ih}}{P_{ih}^{pre}}$, and average this equation over all I individuals graduating from college. This results in the following equation:

$$\frac{1}{I}\sum_{i}\left(\frac{\Delta P_{ih}}{P_{ih}^{pre}} - \frac{1}{H-1}\sum_{k\neq h}\frac{\Delta P_{ih}}{P_{ih}^{pre}}\right) = \frac{\beta}{I}\sum_{i}\left(\Delta Y_{ih} - \frac{1}{H-1}\sum_{k\neq h}\Delta Y_{ik}\right) \quad (2.8)$$

With the available data, the left-hand side of this equation is replaced by the aggregate term $\left(\frac{\Delta P_h}{P_h^{pre}} - \frac{1}{H-1}\sum_{k\neq h}\frac{\Delta P_h}{P_h^{pre}}\right)$. Noting that $\frac{\Delta P_h}{P_h^{pre}}$ equals $\left(\frac{1}{I}\frac{\sum_i \frac{\Delta P_{ih}}{P_{ih}^{pre}} * P_{ih}^{pre}}{\sum_i P_{ih}^{pre}}\right)$, the left-hand side of equation 2.8 will equal this aggregate term if $\frac{\Delta P_{ih}}{P_{ih}^{pre}}$ is uncorrelated with P_{ih}^{pre}. That is, for the individual level relationship to hold at the aggregate level, I need that the percentage change in an individual's probability of choosing a major due to TRA86 is uncorrelated with that individual's baseline probability of choosing that major.

Second, the right hand side of equation 2.8 can be replaced with the aggregate variables if the average of ΔY_{ih} across all individuals equals my estimate of ΔY_h. As discussed in the literature review, Zafar and Wiswall (2015a, 2015b) suggest that students adjust their own beliefs about their lifetime earnings in various majors in response to changes in the population distribution of lifetime earnings by major. This condition then says that the average change in individuals' beliefs about their expected lifetime earnings by major due to TRA86 should equal the actual mechanical change in average lifetime earnings by major in the population due to TRA86 that I calculated[44].

Third, the composition of graduating college students cannot change

[44] As discussed in the literature review, following Zafar and Wiswall (2015a, 2015b) we might expect that $\frac{1}{I}\sum \Delta Y_{ih} = \gamma \Delta Y_h$ where $0 < \gamma < 1$. This means that the β that I estimate is biased towards zero.

over time. In the above analysis, I showed how TRA86 changed the probability of choosing a particular major for a given cohort of I individuals. In the data, the aggregate variables P_h^{pre} and P_h^{post} are instead based on different cohorts of students. A compositional change would cause the baseline distribution of P_{ih}^{pre} for the cohort of graduates on which I base P_h^{post} to be different from the baseline distribution of P_{ih}^{pre} for the cohort on which I base P_h^{pre}. In order for cross-cohort differences in the fraction of completed college degrees that are in a particular major to be attributable to TRA86, the baseline distribution of P_{ih}^{pre} needs to be the same across subsequent cohorts. In the empirical specification, I will try to control for possible changes in the composition of students over time by looking at the subsample of graduating students that attended highly selective universities.

I estimate equation 2.7 in two ways using aggregate data for the academic years 83/84 to 96/97. First, I treat the academic year 1987/1988 as the last year of the pre-period and the academic year 1996/1997 as the last year of the post-period. Since I have data on college major completions, this means that I am assuming that TRA86 could have affected the college major choice of students who were a sophomore or younger in the year 1986/1987. As students decide on their major early on in college, I do not expect the tax change to cause advanced college students to change their major.[45] The first way of estimating equation 2.7 is as follows:

$$ln\left(\frac{P_h^{96/97}}{P_h^{87/88}}\right) - \frac{1}{H-1}\sum_{k\neq h} ln\left(\frac{P_k^{96/97}}{P_k^{87/88}}\right) = \beta_0 + \beta_1\left(\Delta Y_h - \frac{1}{H-1}\sum_{k\neq h}\Delta Y_k\right)$$
$$+ \beta_2\left(ln\left(\frac{P_h^{87/88}}{P_h^{83/84}}\right) - \frac{1}{H-1}\sum_{k\neq h} ln\left(\frac{P_k^{87/88}}{P_k^{83/84}}\right)\right) + \epsilon_h \quad h = 1, ..., H \quad (2.9)$$

In this equation, P_h^t equals the fraction of all bachelor degrees awarded in academic year t that are in field h. Based on previous literature, I expect $\beta_1 > 0$. Factors that stay constant across time and affect college major choice

[45] I am assuming a 2-year lag between college major choice and the receipt of the degree. This might be an underestimation, as the many students take over 4 years to complete their college degree. This will bias me towards not finding any effect in the earlier years.

are differenced out. There are however many factors other than the tax law that change over time and affect college major choice. In fact, aggregate college major patterns fluctuate widely over time. To control for some of the potentially omitted variables, I include the lag of the dependent variable in the above equation as an independent variable, where the value of the lag is calculated for the period 1983/84 to 1987/88[46]. Although this variable might control for some of the time trends in college major choice, it cannot control for all omitted variables as its coefficient is restricted to be the same for all majors. To better control for omitted variables as well as get a better sense of the effect of TRA86 on college major choice over time, I also estimate the following form of equation 2.9:

$$ln\left(\frac{P_h^t}{P_h^{t-1}}\right) - \frac{1}{H-1}\sum_{k\neq h} ln\left(\frac{P_k^t}{P_k^{t-1}}\right) = \alpha_h + \sum_{s=88/89}^{96/97} \beta_s * \mathbb{1}\{t=s\} * \left(\Delta Y_h - \frac{1}{H-1}\sum_{k\neq h}\Delta Y_k\right) + \epsilon_{ht}$$

$$h = 1, ..., H \quad t = 84/85, ..., 96/97 \quad (2.10)$$

This equation differs from equation 2.9 in three ways. First, we now have panel data and include a major-specific dummy to control for factors, such as skill-biased technical change, that might affect college major choice over the whole time period. Second, the dependent variable is now a first difference and equals the relative percentage change in the fraction of completed undergraduate degrees that are in major h between year t and $t-1$. Lastly, I estimate the coefficient β on the independent variable of interest seperately for every year starting with 88/89. In effect, the coefficient β_t treats year $(t-1)$ as the pre-period and tells us whether the change in SWE by major due to TRA86 caused the composition of completed college degrees to change between years $(t-1)$ and t. Estimating β seperately for each year allows me to pin down

[46]The academic year 1983/1984 is the first year for which I have detailed data on the number of bachelor degree completions by field.

exactly when the tax change affects college major choice.[47] When estimating this equation, I cluster the standard errors at the college major level.

Extensions

Up to this point, I assumed that utility is linear in expected lifetime income. Another assumption that is used among others by Arcidiacono (2004) and Arcidiacono, Hotz and Kang (2012) is that utility is linear in the present discounted value of the sum of expected per-period log income: $\sum_{t=1}^{T} \beta^t \; E \; ln(Y_{ht})$. By using $log(w_{ks}^t)$ instead of w_{ks}^t in equation 2.1, I can easily modify my calculation of major-specific SWE to take into account this alternative form of the utility function. I can then also estimate equations 2.9 and 2.10 to take into account this alternative utility function by replacing the ΔY_h measure calculated using w_{ks}^t by the one calculated using $log(w_{ks}^t)$.

Students who take the same course load in a particular college but graduate in different years might be classified as having completed different majors in the annual completions surveys that colleges submit and that I use as my data source on completions by major and year. For instance, the school-specific name of a particular major might change over time, causing college administrators to change how they classify the major in the completion surveys. Also, a school-specific major might be divided into multiple separate majors at a school as the major grows in size or be absorbed by another major if it becomes too small. For some small fields such as pre-law that not very many schools offer, these changes can result in large percentage changes in the fraction of college degrees completed in that field over time in the data that I use. Importantly, these changes are not related to students' actual course taking behavior in college. To make sure such artificially large percentage changes in completions in a given major do not influence my results, I also estimate equation 2.10 by weighting each observation by the average number of completions in that major over the time-period 1983/1984 to 1996/1997.

[47] However, if the effect of TRA86 on college major choice is small and occurs gradually over a longer time period, estimating β seperately for each year increases the chance of not finding a significant effect.

Baseline Regression Results

Table 2.1 below displays the results of estimating equation 2.9. In columns 1 and 2 SWE are calculated using the level of wages and in columns 3 and 4 using the log of wages. The results below use a discount rate of 5% when calculating SWE, but results using a 0% discount rate are very similar. In columns 2 and 4 SWE are "equivalency adjusted" by adjusting earnings for the size of the tax-filing unit when calculating SWE by major.

Table 2.1 indicates that there is no clear correlation between relative

Table 2.1: Effect of TRA86 on Major Choice - Reduced-form following Berger

	(1)	(2)	(3)	(4)
(Δ rel. SWE, r=5) (level)	-0.11			
	(0.15)			
(EA Δ rel. SWE, r=5) (level)		-0.21		
		(0.24)		
(Δ rel. SWE ,r=5) (logs)			-1.24	
			(1.07)	
(EA Δ rel. SWE ,r=5) (logs)				-1.19
				(1.06)
Dep. var. over 83/84 to 88/89	Yes	Yes	Yes	Yes
Observations	47	47	47	47
R^2	0.056	0.064	0.080	0.077

Standard errors in parentheses

Dep. variable: Rel. perc. change in completions in a given major from 88/89 to 96/97

Rel. change in SWE (level) = $(\Delta Y_h - \frac{\sum_{k\neq h} \Delta Y_k}{H-1})$

Rel. change in SWE (logs) = $\left((\hat{Y}_h^{pre} - \hat{Y}_h^{post}) - \frac{\sum_{k\neq h}(\hat{Y}_k^{pre} - \hat{Y}_k^{post})}{H-1}\right)$ where $\hat{Y}_h^j = \sum_{t=1}^{40} \beta^t Eln(Y_{ht}^j)$

Standard errors are White's standard errors. College major choice data from 83/84 to 96/97.

EA = Equivalency Adjusted, meaning that SWE take into account the size of the tax-filing unit.

Coefficients in columns 1 and 2 are multiplied by 10000.

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

changes in SWE in a major and relative percentage changes in completions in that major over the period 1987/1988 to 1996/1997. If anything, table 2.1 seems to suggest that majors that experienced a relatively larger increase in expected lifetime earnings due to TRA86 experienced a relatively larger decrease in the fraction of undergraduate degrees completed in that major. However, none of the coefficients are statistically significant. As discussed in the previous section, in equation 2.9 I am not able to control for time-varying variables that might be correlated with the change in SWE due to TRA86 and also affect the composition of completed college degrees. To better control for these variables and get a better picture of the effect of TRA86 on college major choice over time, I also estimated equation 2.10.

Effect of TRA86 on college major choice over time

The results of estimating equation 2.10 for various measures of the change in major-specific SWE can be found in tables 2.2 through 2.5. There are a total of 4 tables, one for each way in which I construct the change in major-specific SWE due to TRA86 in equation 2.1. In the first two tables I calculated SWE by using the level of earnings (w_{ks}^t) and in the last two tables I used the log of earnings ($log(w_{ks}^t)$). In addition, I adjusted the earnings for the size of the tax filing unit in tables 2.3 and 2.5. I used a discount rate of 5% in all tables[48]. In all tables, the first column shows the OLS results when not including major-specific fixed effects. The other columns include these fixed effects. In the third column each observation is weighted by the average number of annual completions in that major over the period 83/84 to 96/97. Columns 4 through 7 include additional control variables that will be discussed below.

A few results stand out across columns 1 to 3 in all tables. First, whether or not I include major-specific fixed effects, the coefficient on the relative change in major-specific SWE is often negative and statistically significant in 1989/1990. This significance will disappear when I control for the effect of business cycles on college major choice decisions. Second, the unweighted and weighted FE results are similar in each table. The coefficients from the weighted regression are generally somewhat smaller in absolute value and less precisely estimated. The estimated coefficients are thus not driven by large percentage changes in the fraction of degrees completed in a few small majors. Lastly, the fact that almost all coefficients increase after we include major-specific fixed effects shows that time-trends in completions by major are biasing the OLS coefficients downwards. Following these last two points, I will focus on the unweighted FE results as reported in column 2 of each table.

[48] Results using a 0% discount rate are similar.

In all the tables the coefficients on the change in major-specific SWE are almost exclusively positive for 1990/1991 onwards. However, the coefficients are not precisely estimated. The positive coefficients for years after 90/91 are however consistent with the hypothesis that TRA86 caused college major choice decisions to shift towards majors that saw larger positive relative changes in SWE due to TRA86. Before interpreting individual coefficients, it is important to note that throughout this paper the results for the years 91/92 and 92/93 are to be treated with caution. The data on the dependent variable in 91/92 and 92/93 is very noisy as the way that majors were classified changed in 91/92. I will therefore not discuss individual coefficients that are based on data from these two academic years. Also, before interpreting individual coefficients, I now first adress concerns that the above results might be driven by omitted variables that vary over time and across majors or by changes in the composition of graduating college students over time.

Business cycles are one obvious time-varying factor that have been shown to affect the composition of completed college degrees. Business cycles can affect college enrollment decisions, degree completions and types of degrees chosen (see Blom (2012), Bradley (2013), and Altonji et al. (2016a)). Since the effect of business cycles on completions will differ by major, in columns 4, 6 and 7 of each table I control for business cycles flexibly by including two measures of changes in past national unemployment rates interacted with major-specific dummy variables.[49].

I showed earlier that there is a positive correlation between the change in SWE due to TRA86 and the 1985 baseline level of SWE of a major. This is

[49] In particular, following Blom (2012) and Bradley (2013), suppose students' choice of college major is affected by unemployment rates during their last two years of high school and first two years of college. To limit the number of coefficients I need to estimate, assume that students therefore are affected by the average unemployment rate during their last two years of high school and the average unemployment rate during their first two years of college. Since equation 2.10 is estimated in first-differences, I thus care about the differences in these two average unemployment rates for those graduating in academic year (t-1/t) and academic year (t-2/t-1). Assuming a 4-year time difference between graduating high school and college, I therefore care about differences in the unemployment rate between years $(t-4)$ and $(t-6)$, and between years $(t-2)$ and year $(t-4)$. To flexibly control for the effect of the unemployment rate on college major choice, I interact these differences with major-specific dummies.

problematic if there are other time trends in the growth of expected lifetime earnings by major that are correlated with these baseline levels of SWE by major and also affect college major choice patterns over time. For instance, skill-biased technological change is likely to have caused before-tax expected lifetime earnings to grow more quickly for majors with higher baseline SWE. These changes in expected lifetime earnings could over time have caused more students to major in fields with higher initial SWE. To the extent that these changes occur gradually over time, the major-specific fixed effects should control for them. Nevertheless, in column 5 through 7 of all tables I also include baseline 1985 SWE of a major relative to the other majors interacted with a year dummy for years in which I estimate the effect of the change in SWE on the composition of completed college degrees.[50]. This approach follows from Gruber and Saez (2002) who face a similar problem in their study on the elasticity of taxable income with respect to marginal tax rates as changes in individuals' marginal tax rates are often correlated with baseline levels of income.

Lastly, the estimated coefficients might be biased if the composition of graduating college students changes over time. This could happen if TRA86 not only influenced college major choice probabilities conditional on college enrollment, but also the probability of enrolling in college and/or the probability of completing college. In column 7 of each table I therefore only use data on the number of completions by major and year at the top 25 percent of colleges.[51]. Students attending these colleges are not likely to be on the margin of attending and/or completing college and the composition of students graduating from these colleges should thus be affected less by TRA86 or other changes in the economy.

As can be seen in columns 4 through 6 of every table, controlling for either business cycle effects or relative baseline SWE by major removes any

[50]Baseline 1985 SWE by major relative to other majors are calculated as $\left(SWE_h^{85} - \frac{\sum k \neq h SWE_k^{1985}}{H-1}\right)$

[51]Based on data availability, I define a college as belonging to the top 25 percent of all colleges if the sum of the average SAT math and verbal scores of entering freshmen is among the top 25 percent of schools as reported in 2001 College Board data.

negative effect that the relative change in SWE due to TRA86 seemed to have on completions in a major in 88/89 in the original specifications and increases the size and significance of many of the coefficients. The most dramatic increase in the size and significance of coefficients occurs when controlling for relative baseline SWE by major in columns 5 and 6. This is not surprising given the very high correlation between baseline SWE by major and the change in SWE by major due to TRA86[52]. Some of the coefficients increase by up to one order of magnitude after controlling for relative baseline SWE. Overall, the results that control for business cycle effects, relative baseline SWE by major, or both show a clear positive association between the change in SWE of a major due to TRA86 and the change in the fraction of college degrees completed in that major. When only using data on college major completions at selective colleges in column 7 of each table, these general results do not change. The estimated coefficients are generally smaller in absolute value and not as significant as when using completions at all colleges. Nonetheless, in at least on year, 1989/1990, the coefficient is still positive and statistically significant at the 10% level when calculating SWE in levels. Also, the F-test rejects the null hypothesis that all coefficients on the change in SWE are jointly equal to zero at the 10% level in all tables.

Judging by the years in which the coefficients are significant across the various specifications, the shift towards more completions in majors that saw relative increases in SWE due to TRA86 seems to have occured gradually over time and is not confined to students who entered college around the time that TRA86 was implemented. Based on the results in columns 5 through 7, the shift seems to have ended by the academic year 96/97. No coefficient is statistically significant at any conventional level of significance in this year. Assuming that students take four to five years to finish their degree, these results indicate the shift in the composition of completed college degrees took place among students entering from 1984/1985 to 1990/1991. Since students often don't have to make a final decision on their choice of college major until

[52]The correlations between these two measures are 0.96, 0.94, 0.89 and 0.88 in respectively tables 2.2, 2.3, 2.4 and 2.5

the end of their second year in college, it is not surprising that TRA86 affected the college major choice decisions of students who were in college at the time of TRA86. One reason we see shifts in college major choice patterns occur gradually might be that students are responsive to changes in the population average lifetime earnings by major at the time of their college major choice. Before 1988 the observed population average lifetime earnings by major will not have included the full mechanical effect of TRA86, since the income tax law changes were phased in over the period 1986 to 1988. In addition, the shift towards majors that saw a relatively larger increase in SWE due to TRA86 might take place over multiple years if it takes time for students to update their beliefs on the population average lifetime earnings by major after these change. In general though, based on the reported results it is hard to conclude in which exact years the shift towards majors that saw relative increases in SWE due to TRA86 took place. The years in which coefficients are positive and statistically significant varies widely depending on the included control variables and to a smaller extent depending on the way that SWE are calculated.

In terms of the size of the effect, as an example consider the statistically significant coefficient of 0.26 on the change in SWE in 94/95 in column 6 of table 2.3, where major-specific SWE are calculated in levels and by adjusting earnings for the size of the tax-filing unit. This coefficient indicates that a major that experienced a relative increase in SWE due to TRA86 of \$3000, or one standard deviation of this measure, saw a 7.8% higher relative growth rate in the fraction of undergraduate degrees completed in that major between 93/94 and 94/95. This is a large effect considering that the relative growth rate has a standard deviation of 13.3% and that this one-year effect is only part of the total effect of TRA86 on the composition of completed college degrees that takes place over multiple years. When not controlling for relative baseline SWE by major as in columns 1 through 4, this effect is up to one order of magnitude smaller and not statistically significant.

The estimated effect is similar if I instead use the log of earnings to calculate major-specific SWE in equation 2.1. For instance, consider the same coefficient on the change in SWE in 94/95 in column 6 of table 2.5 where SWE are calculated by using log earnings and by adjusting earnings for the size of the tax-filing unit. In this case, the coefficient indicates that majors that experienced a 1 standard deviation increase in this measure of the relative change in SWE by major due to TRA86 experienced a 4.3% higher relative growth rate in the fraction of undergraduate degrees completed in that major between 93/94 and 94/95. Again, these coefficients are smaller when not controlling for relative baseline earnings by major.

The coefficients mentioned above are hard to compare to the literature on college major choice. In the literature, the major-specific measure of expected lifetime earnings is generally individual specific and corrected for self-selection into a major. Also, due to data limitations, the literature has largely considered choices between five large categories of majors (social sciences, natural sciences, etc.) and not between 47 majors as I do here. Lastly, my measure of the change in SWE due to TRA86 is in 1985 dollars, relative to changes in other majors and constant over the entire time-period that I look at. Nevertheless, when including controls for relative baseline levels of SWE by major, the estimated coefficients seem very large compared to the literature. Wiswall and Zafir (2015a) report that most of the literature finds that the elasticity of the probability of choosing a particular major with respect to expected lifetime earnings earnings in that major is around 0.1. In contrast, the coefficient of table 2.3 that I discussed above seems to suggest an elasticity of around 9.[53] Moreover, some of the specifications show that TRA86 changed college major choice patterns in multiple cohorts. The total effect of this one-time change in SWE by major might therefore be even larger.

Placebo Tests

I also run a placebo test. Specifically, I estimate coefficients on the

[53]The average baseline level of SWE earnings by major used in this table is approximately \$345.000. An increase in SWE in one major relative to all the other majors of \$3000 would on average represent an increase of around 0.87% in SWE for that major. According to my estimates, this 0.87% increase causes the major to experience a 7.8% higher relative growth rate in the fraction of undergraduate degrees completed in that major between 93/94 and 94/95.

relative change in SWE for the pre-TRA86 years using the same specifications as discussed above. The TRA86 is not supposed to affect college major choices of the cohorts graduating from college before 88/89. Hence, the estimated coefficients should not be statistically different from zero. Tables 2.6 through 2.9 display the results of these placebo tests. Each table uses a different measure of the relative change in SWE by major due to TRA86. Many of the estimated coefficients are negative and statistically significant, irregardless of the included control variables and the way that SWE are constructed. Moreover, the F-test of joint significance of all the coefficients is statistically significant at the 5% level in most specifications. Both of these results do not hold when I restrict the sample to degree completions at selective colleges in column 7 of each table. In this case, none of the estimated coefficients are statistically significant at any conventional level and the F-tests of joint significance are also never statistically significant. These results suggest that there are omitted variables that cause the estimated coefficients to be biased downwards during this time period. If these omitted variables are also present in later years, this could cause the reported coefficients for those years as discussed above to be biased downwards as well.

Table 2.2: Effect of TRA86 over time - Levels

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(Δ rel. SWE, r=5)*(t=88/89) (level)	0.037	0.053	-0.010	0.083*	0.32*	0.18	0.016
	(0.062)	(0.065)	(0.012)	(0.047)	(0.17)	(0.12)	(0.065)
(Δ rel. SWE, r=5)*(t=89/90) (level)	-0.081*	-0.064*	-0.023	-0.045	0.067	-0.018	0.13*
	(0.042)	(0.037)	(0.019)	(0.053)	(0.11)	(0.14)	(0.072)
(Δ rel. SWE, r=5)*(t=90/91) (level)	-0.0046	0.012	-0.013	0.030	0.19***	0.12**	0.0062
	(0.024)	(0.023)	(0.032)	(0.019)	(0.061)	(0.051)	(0.098)
(Δ rel. SWE, r=5)*(t=91/92) (level)	0.043	0.059	0.041	0.073	0.14	0.082	0.16
	(0.067)	(0.066)	(0.041)	(0.068)	(0.18)	(0.19)	(0.17)
(Δ rel. SWE, r=5)*(t=92/93) (level)	-0.065	-0.048	-0.023	-0.051	0.18	0.19	-0.045
	(0.041)	(0.045)	(0.034)	(0.051)	(0.15)	(0.18)	(0.16)
(Δ rel. SWE, r=5)*(t=93/94) (level)	-0.0060	0.010	0.014	-0.0038	0.14	0.21**	0.059
	(0.027)	(0.028)	(0.035)	(0.030)	(0.092)	(0.100)	(0.12)
(Δ rel. SWE, r=5)*(t=94/95) (level)	0.0067	0.023	0.018	0.012	0.11	0.16*	0.033
	(0.025)	(0.027)	(0.035)	(0.028)	(0.082)	(0.086)	(0.10)
(Δ rel. SWE, r=5)*(t=95/96) (level)	0.0067	0.023	0.016	0.021	0.11	0.11	0.12
	(0.020)	(0.021)	(0.026)	(0.022)	(0.076)	(0.083)	(0.089)
(Δ rel. SWE, r=5)*(t=96/97) (level)	0.0037	0.020	0.00089	0.032*	0.080	0.025	-0.086
	(0.014)	(0.018)	(0.021)	(0.017)	(0.061)	(0.065)	(0.081)
$(Unem_{t-2} - Unem_{t-4})$ × Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6})$ × Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. SWE × Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.019	0.020	0.010	0.159	0.033	0.167	0.161
F-Statistic	2.41	2.48	1.95	2.65	1.71	1.37	2.24
P value of F stat.	0.025	0.021	0.067	0.015	0.11	0.23	0.036

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (level) = $(\Delta Y_h - \frac{\sum_{k \neq h} \Delta Y_k}{H-1})$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97. Coefficients are multiplied by 10000.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.3: Effect of TRA86 over time - Levels / Equivalency Adjusted

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(EA Δ rel. SWE, r=5)*(t=88/89) (level)	0.076	0.100	-0.023	0.15*	0.49*	0.30*	0.086
	(0.11)	(0.12)	(0.025)	(0.083)	(0.25)	(0.18)	(0.091)
(EA Δ rel. SWE, r=5)*(t=89/90) (level)	-0.14*	-0.11*	-0.045	-0.086	0.081	-0.029	0.19*
	(0.073)	(0.066)	(0.035)	(0.095)	(0.16)	(0.21)	(0.097)
(EA Δ rel. SWE, r=5)*(t=90/91) (level)	-0.0067	0.018	-0.033	0.045	0.30***	0.19***	0.031
	(0.042)	(0.041)	(0.055)	(0.033)	(0.086)	(0.069)	(0.13)
(EA Δ rel. SWE, r=5)*(t=91/92) (level)	0.034	0.058	0.049	0.080	0.56***	0.47**	0.51**
	(0.10)	(0.10)	(0.071)	(0.10)	(0.19)	(0.20)	(0.21)
(EA Δ rel. SWE, r=5)*(t=92/93) (level)	-0.086*	-0.062	-0.049	-0.065	0.028	0.033	-0.22
	(0.047)	(0.055)	(0.058)	(0.064)	(0.16)	(0.18)	(0.20)
(EA Δ rel. SWE, r=5)*(t=93/94) (level)	-0.019	0.0055	-0.00029	-0.015	0.22*	0.29**	0.099
	(0.047)	(0.048)	(0.064)	(0.051)	(0.13)	(0.13)	(0.17)
(EA Δ rel. SWE, r=5)*(t=94/95) (level)	0.0022	0.026	0.0096	0.010	0.20*	0.26**	0.091
	(0.044)	(0.048)	(0.063)	(0.048)	(0.11)	(0.12)	(0.15)
(EA Δ rel. SWE, r=5)*(t=95/96) (level)	0.0014	0.026	0.0072	0.023	0.19*	0.21*	0.22
	(0.034)	(0.036)	(0.048)	(0.039)	(0.11)	(0.12)	(0.13)
(EA Δ rel. SWE, r=5)*(t=96/97) (level)	0.0047	0.029	-0.0097	0.047	0.15	0.074	-0.095
	(0.025)	(0.033)	(0.038)	(0.031)	(0.089)	(0.096)	(0.13)
$(Unem_{t-2} - Unem_{t-4}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. EA SWE $\times$ Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.016	0.016	0.008	0.155	0.034	0.165	0.168
F-Statistic	2.69	2.44	1.71	2.63	3.45	2.71	3.19
P value of F stat.	0.013	0.023	0.11	0.015	0.0025	0.013	0.0045

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (level) = $(\Delta Y_h - \frac{\sum_{k \neq h} \Delta Y_k}{H-1})$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97. Coefficients are multiplied by 10000.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.4: Effect of TRA86 over time - Logs

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(Δ rel. SWE,r=5)*(t=88/89) (logs)	0.25	0.39	-0.12	0.66*	1.63	1.25*	0.18
	(0.48)	(0.50)	(0.11)	(0.35)	(0.99)	(0.70)	(0.33)
(Δ rel. SWE,r=5)*(t=89/90) (logs)	-0.65**	-0.51*	-0.24	-0.35	-0.034	-0.29	0.46
	(0.31)	(0.28)	(0.16)	(0.41)	(0.63)	(0.85)	(0.36)
(Δ rel. SWE,r=5)*(t=90/91) (logs)	-0.070	0.068	-0.17	0.23	0.83***	0.61**	0.29
	(0.19)	(0.18)	(0.24)	(0.15)	(0.31)	(0.26)	(0.58)
(Δ rel. SWE,r=5)*(t=91/92) (logs)	0.12	0.25	0.21	0.38	0.32	0.19	0.63
	(0.49)	(0.49)	(0.31)	(0.50)	(0.88)	(0.91)	(0.93)
(Δ rel. SWE,r=5)*(t=92/93) (logs)	-0.42	-0.28	-0.20	-0.29	0.74	0.86	-0.041
	(0.25)	(0.29)	(0.27)	(0.33)	(0.68)	(0.80)	(0.83)
(Δ rel. SWE,r=5)*(t=93/94) (logs)	-0.11	0.030	-0.016	-0.081	0.46	0.73	-0.049
	(0.21)	(0.22)	(0.28)	(0.24)	(0.42)	(0.48)	(0.53)
(Δ rel. SWE,r=5)*(t=94/95) (logs)	-0.0069	0.13	0.030	0.040	0.39	0.56	-0.069
	(0.19)	(0.22)	(0.28)	(0.22)	(0.39)	(0.41)	(0.51)
(Δ rel. SWE,r=5)*(t=95/96) (logs)	-0.0017	0.14	0.032	0.12	0.36	0.36	0.44
	(0.16)	(0.17)	(0.21)	(0.18)	(0.38)	(0.41)	(0.45)
(Δ rel. SWE,r=5)*(t=96/97) (logs)	-0.0025	0.14	-0.022	0.24*	0.29	0.13	-0.48
	(0.12)	(0.15)	(0.16)	(0.14)	(0.31)	(0.30)	(0.40)
$(Unem_{t-2} - Unem_{t-4})$ × Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6})$ × Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. SWE (logs) × Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.015	0.015	0.008	0.153	0.027	0.162	0.157
F-Statistic	2.61	2.22	1.48	2.69	1.49	1.81	1.88
P value of F stat.	0.016	0.038	0.18	0.013	0.18	0.091	0.079

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (logs) = $\left((\Delta \hat{Y}_h) - \frac{\sum_{k \neq h}(\Delta \hat{Y}_k)}{H-1}\right)$ where $\hat{Y}_h^j = \sum_{t=1}^{40} \beta^t Eln(Y_{ht}^j)$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.5: Effect of TRA86 over time - Logs / Equivalency Adjusted

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(EA Δ rel. SWE,r=5)*(t=88/89) (logs)	0.27	0.41	-0.11	0.66*	1.56*	1.19*	0.38
	(0.49)	(0.51)	(0.11)	(0.36)	(0.88)	(0.62)	(0.30)
(EA Δ rel. SWE,r=5)*(t=89/90) (logs)	-0.63**	-0.49*	-0.22	-0.34	0.015	-0.21	0.49
	(0.31)	(0.28)	(0.16)	(0.41)	(0.55)	(0.74)	(0.30)
(EA Δ rel. SWE,r=5)*(t=90/91) (logs)	-0.060	0.076	-0.16	0.22	0.83***	0.61***	0.25
	(0.19)	(0.18)	(0.24)	(0.15)	(0.28)	(0.21)	(0.43)
(EA Δ rel. SWE,r=5)*(t=91/92) (logs)	0.032	0.17	0.19	0.29	1.69**	1.52*	1.66**
	(0.46)	(0.46)	(0.31)	(0.46)	(0.75)	(0.77)	(0.78)
(EA Δ rel. SWE,r=5)*(t=92/93) (logs)	-0.33*	-0.19	-0.18	-0.20	-0.081	-0.050	-0.62
	(0.19)	(0.23)	(0.26)	(0.27)	(0.61)	(0.69)	(0.76)
(EA Δ rel. SWE,r=5)*(t=93/94) (logs)	-0.11	0.030	-0.029	-0.070	0.53	0.70	0.11
	(0.21)	(0.22)	(0.28)	(0.23)	(0.38)	(0.43)	(0.52)
(EA Δ rel. SWE,r=5)*(t=94/95) (logs)	-0.0099	0.13	0.013	0.041	0.52	0.66*	0.14
	(0.19)	(0.22)	(0.28)	(0.22)	(0.35)	(0.36)	(0.47)
(EA Δ rel. SWE,r=5)*(t=95/96) (logs)	-0.0051	0.13	0.021	0.11	0.52	0.55	0.65
	(0.15)	(0.17)	(0.21)	(0.18)	(0.36)	(0.39)	(0.42)
(EA Δ rel. SWE,r=5)*(t=96/97) (logs)	-0.000097	0.14	-0.026	0.23	0.40	0.25	-0.30
	(0.12)	(0.15)	(0.16)	(0.14)	(0.28)	(0.28)	(0.41)
$(Unem_{t-2} - Unem_{t-4})$ $\times$ Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6})$ $\times$ Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. EA SWE (logs) $\times$ Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.013	0.013	0.006	0.151	0.031	0.161	0.167
F-Statistic	3.34	2.13	1.34	2.81	2.76	3.27	3.13
P value of F stat.	0.0032	0.045	0.25	0.010	0.012	0.0037	0.0051

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (logs) = $\left((\Delta\hat{Y}_h) - \frac{\sum_{k\neq h}(\Delta\hat{Y}_k)}{H-1}\right)$ where $\hat{Y}_h^j = \sum_{t=1}^{40} \beta^t Eln(Y_{ht}^j)$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.6: Placebo Tests - Levels

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(Δ rel. SWE, r=5)*(t=84/85) (level)	0.012	0.019	0.029	0.034	-0.28***	-0.22	-0.040
	(0.020)	(0.026)	(0.024)	(0.042)	(0.092)	(0.14)	(0.14)
(Δ rel. SWE, r=5)*(t=85/86) (level)	-0.012	-0.0053	0.0085	-0.0035	-0.15**	-0.038	-0.053
	(0.012)	(0.019)	(0.021)	(0.033)	(0.073)	(0.11)	(0.14)
(Δ rel. SWE, r=5)*(t=86/87) (level)	-0.026*	-0.019	-0.0014	-0.020	-0.11*	-0.017	-0.094
	(0.014)	(0.021)	(0.021)	(0.025)	(0.063)	(0.081)	(0.11)
(Δ rel. SWE, r=5)*(t=87/88) (level)	-0.041*	-0.034	-0.045*	-0.041**	-0.042	-0.11**	-0.033
	(0.024)	(0.023)	(0.023)	(0.016)	(0.072)	(0.054)	(0.069)
$(Unem_{t-2} - Unem_{t-4}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. SWE $\times$ Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.004	0.003	0.008	0.140	0.016	0.145	0.153
F-Statistic	2.34	1.99	2.91	2.89	3.08	1.74	0.29
P value of F stat.	0.069	0.11	0.032	0.033	0.025	0.16	0.88

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (level) = $(\Delta Y_h - \frac{\sum_{k \neq h} \Delta Y_k}{H-1})$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97. Coefficients are multiplied by 10000.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.7: Placebo Tests - Levels / Equivalency Adjusted

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(EA Δ rel. SWE, r=5)*(t=84/85) (level)	0.018	0.033	0.062	0.062	-0.41***	-0.22	-0.021
	(0.037)	(0.048)	(0.043)	(0.064)	(0.13)	(0.18)	(0.20)
(EA Δ rel. SWE, r=5)*(t=85/86) (level)	-0.015	-0.00073	0.030	0.011	-0.28***	-0.12	-0.11
	(0.021)	(0.035)	(0.038)	(0.058)	(0.10)	(0.16)	(0.20)
(EA Δ rel. SWE, r=5)*(t=86/87) (level)	-0.039	-0.025	0.011	-0.019	-0.20**	-0.082	-0.19
	(0.024)	(0.038)	(0.038)	(0.044)	(0.090)	(0.12)	(0.17)
(EA Δ rel. SWE, r=5)*(t=87/88) (level)	-0.060	-0.046	-0.062	-0.062**	-0.097	-0.23***	-0.12
	(0.043)	(0.041)	(0.043)	(0.027)	(0.10)	(0.068)	(0.088)
$(Unem_{t-2} - Unem_{t-4}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. EA SWE $\times$ Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.003	0.002	0.008	0.139	0.016	0.144	0.154
F-Statistic	1.83	1.50	2.33	3.81	3.19	2.92	0.94
P value of F stat.	0.14	0.22	0.070	0.0093	0.021	0.031	0.45

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (level) = $(\Delta Y_h - \frac{\sum_{k \neq h} \Delta Y_k}{H-1})$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97. Coefficients are multiplied by 10000.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.8: Placebo Tests - Logs

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(Δ rel. SWE,r=5)*(t=84/85) (logs)	0.100	0.20	0.31	0.30	-1.01**	-0.61	0.084
	(0.16)	(0.21)	(0.19)	(0.30)	(0.46)	(0.65)	(0.67)
(Δ rel. SWE,r=5)*(t=85/86) (logs)	-0.097	0.0022	0.15	0.018	-0.52	0.075	0.17
	(0.10)	(0.16)	(0.17)	(0.26)	(0.34)	(0.55)	(0.69)
(Δ rel. SWE,r=5)*(t=86/87) (logs)	-0.21*	-0.11	0.052	-0.11	-0.42	0.052	-0.48
	(0.11)	(0.17)	(0.16)	(0.19)	(0.30)	(0.40)	(0.63)
(Δ rel. SWE,r=5)*(t=87/88) (logs)	-0.35*	-0.25	-0.29	-0.30**	-0.27	-0.64**	-0.22
	(0.19)	(0.18)	(0.19)	(0.13)	(0.37)	(0.26)	(0.35)
$(Unem_{t-2} - Unem_{t-4})$ × Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6})$ × Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. SWE (logs) × Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.004	0.003	0.009	0.139	0.012	0.143	0.153
F-Statistic	2.53	1.94	2.96	3.34	2.19	1.84	0.29
P value of F stat.	0.053	0.12	0.029	0.017	0.085	0.14	0.88

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (logs) = $\left((\Delta\hat{Y}_h) - \frac{\sum_{k\neq h}(\Delta\hat{Y}_k)}{H-1}\right)$ where $\hat{Y}_h^j = \sum_{t=1}^{40} \beta^t Eln(Y_{ht}^j)$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Table 2.9: Placebo Tests - Logs / Equivalency Adjusted

	(1) OLS	(2) FE	(3) FE Weighted	(4) FE	(5) FE	(6) FE	(7) FE T25
(EA Δ rel. SWE,r=5)*(t=84/85) (logs)	0.084	0.18	0.29	0.26	-1.01**	-0.37	0.12
	(0.16)	(0.21)	(0.18)	(0.28)	(0.42)	(0.60)	(0.64)
(EA Δ rel. SWE,r=5)*(t=85/86) (logs)	-0.094	-0.0011	0.14	0.031	-0.73**	-0.22	-0.086
	(0.10)	(0.16)	(0.16)	(0.26)	(0.31)	(0.51)	(0.63)
(EA Δ rel. SWE,r=5)*(t=86/87) (logs)	-0.20*	-0.10	0.058	-0.089	-0.55*	-0.19	-0.66
	(0.11)	(0.17)	(0.16)	(0.19)	(0.27)	(0.38)	(0.57)
(EA Δ rel. SWE,r=5)*(t=87/88) (logs)	-0.34*	-0.24	-0.27	-0.29**	-0.37	-0.82***	-0.44
	(0.19)	(0.18)	(0.19)	(0.13)	(0.32)	(0.21)	(0.28)
$(Unem_{t-2} - Unem_{t-4}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
$(Unem_{t-4} - Unem_{t-6}) \times$ Major dummy	No	No	No	Yes	No	Yes	Yes
1985 rel. EA SWE (logs) $\times$ Year Dummies	No	No	No	No	Yes	Yes	Yes
Observations	611	611	6528938	611	611	611	611
R^2	0.004	0.002	0.008	0.139	0.014	0.142	0.154
F-Statistic	2.27	1.63	2.83	4.26	2.30	3.89	1.15
P value of F stat.	0.076	0.18	0.035	0.0052	0.073	0.0083	0.34

Standard errors in parentheses

Dep. variable: Rel. perc. change in annual completions in a given major

Rel. change in SWE (logs) = $\left((\Delta\hat{Y}_h) - \frac{\sum_{k\neq h}(\Delta\hat{Y}_k)}{H-1}\right)$ where $\hat{Y}_h^j = \sum_{t=1}^{40} \beta^t Eln(Y_{ht}^j)$

Standard errors are clustered by major. College major choice data from 83/84 to 96/97.

When weighted, each obs. is weighted by the average no. of BA degrees awarded in that major in the period 83/84 to 96/97.

T25 means that only the top 25 perc. of colleges in terms of av. SAT scores of freshmen in 2000 are included.

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

Chapter 3

The peer effects of classmates who speak a language other than English at home

Abstract

Using publicly available data from California for the year 2012-2013, this paper studies the peer effects of students who speak a language other than English at home (ESL) on native English speaking students (ELO) in grades 2 to 6. In contrast to the previous literature, I do not just focus on the peer effects of English Language Learners but rather on the peer effects of the larger group of students who do not speak English at home as the share of students who are English Language Learners in any given year, school and grade is more likely to be endogenous. I identify peer effects by exploiting variation in the distribution of ESL students across grades within a given school and year. I find that higher concentrations of ESL students have no effect on average math test scores and a negative effect on average English test scores of ELO students. Surprisingly, I find that these negative peer effects are driven almost entirely by the negative peer effects of the small group of ESL students who do not speak Spanish at home.

3.1 Introduction

In 2015, 22% of children between the ages of 5 and 17 in the U.S. spoke a language other than English at home[1], up from 18% in 2000. This percentage varied widely across states from a low of 2% in West-Virginia to a high of 45% in California. Given the corresponding large share of students with English as a second language (ESL) in K-12 in the U.S., there is a concern that these students could have negative peer effects on so-called English language only (ELO) students. These negative peer effects could be a result of cultural differences between the two groups of students or the lower level of English language proficiency of ESL students, as a little over 50% of all ESL students in K-12 in California for instance classify as English Language Learners (ELL) and receive additional instruction to improve their English language proficiency.

Previous literature has looked at the peer effects of immigrant students in Europe[2], ESL students in Canada and the U.K.[3] and ELL students in the U.S.[4]. In contrast to studies from other countries, the U.S. literature consistently finds negative peer effects on test scores in both mathematics and English Language Arts (ELA) for students in kindergarten through grade 8. The studies from other countries generally find no or very small negative peer effects for students in high school only. One possible reason for this discrepancy is that U.S. studies focus on the share of students who are ELL, which could be endogenous, whereas other studies focus on more exogenous shares of students that are based on student characteristics that are time-invariant.

In this paper, I estimate the peer effects of ESL students on ELO students in a U.S. context using publicly available data from California for the year 2012-2013 for students in grades 2 to 6. Whether or not a student classifies as ESL is time-invariant and determined when a student enters the school district. As a result, by focusing on ESL students I avoid the problems as-

[1] This is based on data from the Census Bureau.

[2] e.g. Gould et al. (2009), Ohinata and van Ours (2013), Jensen and Wurtz Rasmussen (2011) and Brunello and Rocco (2013)

[3] e.g. Friesen and Krauth (2011) and Geay et al. (2013)

[4] e.g. Cho (2012), Diette and Oyelere (2012, 2014, 2016), Ahn and Jepsen (2015)

sociated with the share of ELL students in a grade being endogenous. In addition, since the concentration of ESL students is higher in California than in any other state of the U.S. and than in any other English-speaking country, I also study how the peer effects of ESL students vary based on the average percentage of students in a school who are ESL. This type of analysis is not possible in other studies as most schools in these other studies have very few ESL students. Lastly, the high shares of ESL in California also allow me to study how the peer effects of ESL students vary by the language spoken at home by the ESL students. This is important as heterogeneous impacts provide policy makers with guidance when thinking about ways to address the peer effects of ESL students.

To identify peer effects, I estimate how the mean level of achievement of ELO students in a particular school and grade on standardized tests in math and English varies with the share of students in that school and grade who are ESL. To control for selection into schools, all regressions include school fixed effects. I find that higher concentrations of ESL students have no effect on average math test scores and have a small negative effect on average ELA test scores of ELO students. Further analysis suggests that these peer effects do not systematically differ based on the average share of students in a school who are ESL. Also, I find that the observed negative peer effects are driven almost entirely by the negative peer effects of the small group of ESL students who do not speak Spanish at home. I also look at the peer effects of Fluent English Proficient (FEP) students and ELL students, who together make up the group of ESL students. FEP students are ESL students who have demonstrated a sufficient command of English as not to require supplemental English language instruction. These estimates are to be treated more carefully as the share of students who are ELL and FEP could be endogenous. These analyses show the negative peer effects on ELA test scores to be solely due to FEP students and also suggest that higher concentrations of ELL students have a negative effect on math test scores of ELO students.

These results have a number of implications. First, the high levels of ESL students in California are unlikely to have a large impact on the academic performance of ELO students in math and ELA. To the extent that existing policies for ESL students in California can be copied by other states, this is important for appeasing worries of large negative peer effects in other states.

Most importantly, I find no negative peer effects of ESL students in some key areas. In particular, there are no negative peer effects of ELL students on ELA test scores of ELO students and very small or no negative peer effects in math and ELA of Spanish-speaking ESL students who make up the bulk of ESL students. Instead, I find negative peer effects of FEP students on ELA test scores and of ELL students on math test scores. In addition, I find that the estimated negative peer effects on ELA test scores are mostly driven by non-Spanish speaking ESL students. These patterns are consistent with a situation in which teachers have more experience dealing with Spanish-speaking ESL students and policies are focused on mitigating potential negative peer effects of ESL students in areas where we might expect them. It seems that large gains could still be made by focusing on ESL students where we might not expect negative peer effects such as ELL students in math classes and FEP students in ELA classes, and by focusing on non-Spanish speaking ESL students.

The remainder of this paper is structured as follows. Section 2 provides an overview of the literature. Section 3 discusses the data and presents some summary statistics. Section 4 discusses the empirical methodology. Section 5 discusses the various results and section 6 concludes.

3.2 Literature Review

This paper is connected to two strands of literature. First, there is a mostly Europe-based literature that looks at the peer effects of immigrant[5]

[5]The papers in the literature sometimes distinguish between first- and second generation immigrant children, although many just use first generation immigrant children. First generation immigrant children are generally defined as those born outside the current country of residence with at least one parent also born abroad. Second generation immigrant children are those who have at least one parent who is born abroad.

children on the educational performance of native children[6]. The papers in this literature suggest that immigrant concentrations in high school have a small but statistically significant negative impact on the academic performance of native students in both reading and mathematics. Immigrant concentrations in elementary school however have not been found to affect contemporary test scores (Ohinata and van Ours (2013)), but do have a small negative effect on long-term academic outcomes such as high school graduation rates (Gould et al. (2009)).

Second, there is a literature in English speaking countries that looks at the effects of concentrations of non-native English speakers on the test scores of native English speaking peers[7]. This literature, which mostly uses data from grades 4 to 8, has looked either at the peer effects of all students who speak a language other than English at home or only considered the peer effects of English Language Learners. In the former case, Geay et al. (2013) find that in the UK the concentration of students who speak a language other than English at home has no significant effect on the test scores of students who speak English at home. Friesen and Krauth (2011) are further able to differentiate between the students who do not speak English at home by the language spoken at home[8]. Using Canadian data, they find significant negative effects, mostly on math and not on reading scores, of higher concentrations of Punjabi speaking students and marginally significant positive effects of higher concentrations of Chinese speaking students.

The U.S. based literature has only looked at the peer effects of English Language Learners. Papers by Diette and Oyelere (2012, 2014, 2016) and Ahn and Jepsen (2015) generally find that higher concentrations of ELLs are associated with lower test scores of native students. Cho, who looks at the reading and math test scores of kindergarten and first-grade students, finds that the negative effects are concentrated among girls and students coming

[6]See for instance Gould et al. (2009), Ohinata and van Ours (2013), Jensen and Wurtz Rasmussen (2011) and Brunello and Rocco (2013).

[7]See for instance Friesen and Krauth (2011), Cho (2012), Geay et al. (2013), Diette and Oyelere (2012, 2014, 2016), Ahn and Jepsen (2015).

[8]Due to confidentiality reasons, they can only distinguish between students who speak Chinese, Punjabi or another language at home.

from low-income households. Diette and Oyelere, using data on 4^{th} to 8^{th} graders, however find that the negative effects are concentrated among boys and students at the top of the ability distribution. Ahn and Jepsen do not look at heterogeneous impacts but find average negative impacts on test scores in both mathematics and reading for students in middle school.

The U.S. based literature thus focuses on the peer effects of English Language Learners, whereas studies from other countries focus on the peer effects of students who do not speak the country's official language at home (ESL) or on the peer effects of immigrant students more generally. In contrast to studies from other countries, the U.S. literature consistently finds negative peer effects on test scores in both mathematics and ELA for students in kindergarten through grade 8. Studies from other countries generally find no peer effects or very small negative peer effects for students in high school only. One reason for the discrepancy between U.S. studies and studies from other countries could be that they look at peer effects of different groups of students. English Language Learners tend to be a particularly disadvantaged subset of students who do not speak the country's official language at home or of immigrant students more generally. In this light, the more pronounced negative peer effects from ELL students are not surprising.

However, another reason, as explained in more detail in section 3.4, for the more pronounced negative peer effects of ELLs is that the share of ELLs in a particular grade, school and year could be endogenous. For instance, in California, students who are classified as ELLs when first entering a public school can be reclassified as Fluent English Proficient (FEP) at the beginning of each school year based primarily on their performance on a number of standardized tests in English at the end of the previous school year (see Hill et al. (2014)). Students' performance on these tests could be influenced by the performance of native English speaking students in their cohort. If the composition of a cohort in a school does not change dramatically over time, this could mean that we observe low shares of ELL students in a given cohort due to positive peer effects of native English speaking students and also observe high performance of native English speaking students in that cohort. This could cause us to falsely conclude that low shares of ELL students lead to better performance of native English speaking students even though the actual relationship is the other way around.

In this paper, I add to the above literature by looking at the peer effects of ESL students on non-ESL students in a U.S. context. In California, the state that I will be looking at, whether or not a student classifies as ESL is determined by his/her parents' answers to the home language survey when enrolling in the school district and cannot change in response to particularly able non-ESL peers. As a result, by focusing on ESL students I avoid the problems associated with the share of ELL students in a grade being endogenous. In addition, since the concentration of ESL students is higher in California than in any other state of the U.S. and than in any other English speaking country, I can also study how the peer effects of ESL students vary based on the average percentage of students in a school who are ESL. This type of analysis is not possible in other studies as most schools in these other studies have very few ESL students. Lastly, following Friesen and Krauth (2011), I also look at how the peer effects of ESL students differ based on the language that they speak at home. This is important as heterogeneous impacts provide policy makers with guidance when thinking about ways to address the peer effects of ESL students.

3.3 Data

I use various publicly available data sets from the California Department of Education[9]. This data covers the universe of public schools in California. In particular, for the year 2012-2013 that I look at, I have data on the number of ELL and FEP students by school, grade and home language. This data will be used to construct the main independent variables of the paper. In addition, I also have data on the demographic characteristics of the students, allowing me to calculate the share of students in a particular school and grade who are male or belong to a particular ethnic group. These shares will be used

[9]This data can be found at http://dq.cde.ca.gov/dataquest/.

as control variables. Lastly, as outcome variables, I use mean test scores in English and math on the California Standards Test (CST) for English Language Only students in grades 2 to 6. 2012-2013 was the last year in which all students in grades 2 to 11 took the CSTs towards the end of the school year. I focus on grades 2 to 6 because the CSTs are only grade-specific up to grade 6. For the analysis, I normalize the mean test scores using statewide means and standard deviations by subject and grade. The coefficients will therefore reflect changes in z-scores.[10]

Summary statistics for the data are shown in table 3.1. On average, 42% of all students in grades 2 to 6 are ESL students. Of these ESL students, close to two-third are ELLs and close to 80% speak Spanish at home. The other languages most commonly spoken at home by ESL students include a variety of East Asian languages such as Tagalog, Vietnamese, Chinese and Korean. As we can see in figure 3.1, the share of students who are ESL in a particular school and grade varies widely across schools and grades. This will allow us later to see how the peer effects of ESL students vary based on the school-wide average share of ESL students. Lastly, the average cohort size is a little over 86 students. In total, I have data from 22785 school by grade combinations from 5425 different schools.

3.4 Methodology

The basic level model I estimate is as follows:

$$y_{gst} = \alpha + \beta_1 ESL_{gst} + \beta_2 X_{gst} + \gamma_{st} + \gamma_g + \eta_{gst} \tag{3.1}$$

In this equation, y_{gst} is the average z-score in ELA or math on the California Standards Test of ELO students in grade g of school s in year t. Our main independent variable of interest is ESL_{gst}, which is the share of students in grade g, school s and year t who speak a language other than English at home.

[10]Due to confidentiality concerns, test scores are only available for school by grade combinations in which 10 or more English Language Only students took the CST test.

X_{gst} contains a number of grade-school-year specific control variables such as the percentage of students that are male and the share of students that belong to particular ethnic groups[11] We also include grade-specific fixed effects. Standard errors are clustered at the school level in each model.

The main problem facing the empirical strategy of this paper is that students do not randomly select into schools. For instance, the types of ELO students who choose to attend a school with a high share of ESL students might be different from the types of ELO students who choose to attend a school with a low share of ESL students. This can result in a correlation between the average outcomes of ELO students, y_{gst}, and the share of the students that are ESL, ESL_{gst}, even if this share has no effect on the average outcomes of ELO students.

Following previous literature, we include school by year fixed effects γ_{st} to control for selection into schools. β_1 will then be consistently estimated if the assumption of strict exogeneity of η_{gst} is satisfied. In particular, for this assumption to be satisfied we need that conditional on the school of choice in a given year, there is no relationship between the share of students in a given grade who are ESL and the unobserved characteristics of ELO students in the same grade. This is satisfied if in any particular year, each grade cohort represents a random draw from a fixed school-specific distribution of students in that year. In this case, following Hoxby (2000), we can exploit the small but plausibly random variation in peer group composition across grades within a school and year to identify the causal relationship between the share of students who are ESL and the outcomes of ELO students.

However, it is important to note that this assumption can be violated. For instance, suppose that for some exogenous reason the number of ESL students enrolling in first grade of a particular school is increasing over time. This could change the types of ELO students enrolling in first grade of that school each year as parents enrolling their ELO children in later years might have a higher preference for diversity. If students are not likely to switch elementary schools after first grade and if parents' preference for diversity is correlated

[11] I include the share of students who are Black, White, Hispanic, Asian or other ethnicity.

with factors affecting students' outcomes, this would violate the identification assumption as for a given school and year, the unobserved characteristics of ELO students in a particular grade that affect test scores would be correlated with the share of ESL students in the same grade.

Figure 3.2 shows the distribution of the share of students who are ESL across grades relative to the school mean. Since I include school by year fixed effects, this is the variation that I will be exploiting in this paper. This figure, which captures over 99% of all observations in my sample, highlights that I am exploiting small variations in peer group composition across grades within a school and year. In absolute value, the average deviation from the school-wide average in a particular grade is 3.7 percentage points. In the average cohort with 86 students of whom 36 are ESL, this amounts to a little over 3 ESL students. Such small deviations from school-wide averages in any particular year are unlikely to change the types of ELO students enrolled in a particular cohort and are therefore likely to be exogenous. As smaller variations across grades within a school are more likely to be exogenous, as a robustness check, I also will run analyses in which I drop the top 2% or top 20% of schools in terms of the maximum variation in the share of ESL students across grades.

All related U.S. based papers look at the effect of the share of students who are English Language Learners in a particular school, grade and year on the outcomes of ELO students in the same school, grade and year. As mentioned before, ESL students include both ELL students and FEP students who speak a language other than English at home but are fluent in English. The problem with looking at the effect of the share of students that are ELL is that this share could be endogenous. For instance, suppose the share of ESL students is the same in all grades in a given school and year. If in some grades there are some particularly able ELO students who have positive peer effects on the ESL students, we might observe a smaller share of ELL students and a higher share of FEP students in that grade as the ESL students who were ELL initially are more likely to have been reclassified to FEP. In this case, the share of ELL students in a particular grade and year would, even conditional on school by year fixed effects, be endogenous. Some studies have used shares of ELL students in a particular cohort in earlier years as an instrument for the current share to control for such endogenous reclassification. However, given the positive correlation in these shares over time such a lagged share might still not be completely exogenous. For instance, suppose the ELO students in a cohort of an elementary school are particularly able and cohort composition does not change dramatically over time. In this case, we might observe both low shares of ELL students in that cohort in grade 3 due to positive peer effects of the ELO students and high performance of ELO students in grade 6. Using then the share of ELL students in grade 3 would not resolve the endogeneity problem. To avoid the problem of endogenous reclassification entirely I look at the share of students that are ESL, that is, either ELL or FEP. Whether or not a student classifies as ESL is determined by his/her parents' answers to the home language survey when enrolling in the school district and cannot change in response to particularly able ELO peers.

Since I estimate a levels regression and ESL_{gst} is not likely to change dramatically from grade to grade for a given cohort, the coefficient on ESL_{gst} can be seen as capturing both the contemporaneous effect of having a particular peer group of ESL students as well as the effect of having a very similar peer group in past years. As Friesen and Krauth (2011) note, the coefficient β_1 is therefore likely to overstate the contemporaneous effect of having a peer group in the current grade and understate the effect of having that same peer group in every grade up to the current grade. Also, the share of ESL students in a particular grade could affect the resources available to ELO students in the same grade, as class sizes might for instance be affected or particular teachers could be asked to teach in grades with a high number of ESL students. Since I do not have control variables measuring the availability of various resources at the grade level, the coefficient β_1 can be interpreted as capturing both the direct peer effects of ESL students as well as the effect of changes in resources as the share of students that are ESL changes.

Lastly, to place the results in the appropriate context, it is important to know that in California elementary schools, FEP students are in the same classrooms and take the same courses as ELO students. ELL students with low levels of English proficiency often follow English immersion programs with other similar ELL students in separate classrooms. These programs, which can be either only in English or in both the students' native language and English, are tailored to improving their English language skills as well as to teaching the other core subjects using methods that are effective for students with a low level of proficiency in English. More advanced ELL students are sometimes still taught separately from the FEP and ELO students, but can also be placed into mainstream classes. These students will follow the same curriculum and courses as the FEP and ELO students, with some time each day set aside for additional instruction focusing on their English language skills. In general though, the way ELL students are taught can vary widely across school districts within California. In particular, schools have flexibility in whether or not ELL students are placed in the same classroom as ELO students.

3.5 Results

The baseline results for math test scores are shown in table 3.2. As shown in column 1 there is a clear negative association between the share of students in a grade that are ESL and the math performance of ELO students. However, once we control for selection into schools by adding school fixed effects in column 2 this negative association becomes small and statistically insignificant. It remains so when we add school-grade level control variables in column 3 such as the share of students in a grade that are male and belong to various ethnic groups. The results are also unaffected by excluding the top 2% or top 20% of all schools in terms of the maximum variation in the share of ESL students across grades within the school. Large variations in the share of students who are ESL across grades could have caused grade-specific selection patterns into a school in a given year and are therefore less likely to be exogenous. In column 4, I show the coefficient separately for the share of students who are ELL and for the share of students who are FEP. The coefficient on the share of students who are ELL is statistically significant and negative.

The difference between the coefficients on the share of students who are ESL and on share of students who are ELL is in line with the previous U.S. and international literature and, as discussed in section 3.4, could partially be a result of the share of students who are ELL being endogenous. This does not seem to be the full explanation though, as we would expect the coefficient on FEP to be positive and statistically significant if endogeneity played a large role. Another explanation is that in some schools, ELL students share classrooms with ELO students and follow the same courses. It is possible that ELL students hinder the math development of ELO students by slowing down the class.

The baseline results for ELA test scores are shown in table 3.3. As shown in column 1 there is a clear negative association between the share of students in a grade who are ESL and the ELA performance of ELO students. Adding school and grade fixed effects, various control variables and looking at only a subset of schools in columns 5 and 6 decreases the size of the coefficient, but it remains statistically significant and negative. Under the baseline effect in column 3, a one standard deviation increase of 0.24 in the share of students who are ESL in a particular grade and school is expected to decrease the average English test scores of ELO students by around 0.03 standard deviations. For a student at the 50^{th} percentile, assuming test scores follow a normal distribution, this is equivalent to moving the student to the 48.8^{th} percentile. To place this effect in the right context, this is the same effect that increasing class size by around 3 students has on mathematics and reading test score gains in fourth grade as estimated by Rivkin et al. (2005).

The effect I find here is comparable in size to the effect of the share of ELL students in a grade on English test scores as found by Diette and Oyelere (2012, 2014, 2016) and Ahn and Jepsen (2015) for students in grades 4 to 8 in North Carolina. Since these studies are able to control for students' test scores in the previous year, the effect they find can be interpreted as the contemporaneous effect of having a certain share of classmates who are ELL. Since I cannot control for test scores in the previous year and the share of students who are ESL in a cohort is not likely to change dramatically from grade to grade, the effect I estimate captures both the contemporaneous effect of having a particular peer group of ESL students as well as the effect of having a very similar peer group in past years. In this light, we expect the contemporaneous effect alone to be smaller than that found in the previous literature that focused on ELL students.

One surprising feature of the result highlighted in column 4 is that the overall coefficient is entirely driven by FEP students. The English test scores of ELO students are not affected by ELL students but are negatively impacted by higher shares of FEP students. This is not very surprising since, as explained earlier, ELL students are often placed in separate classrooms, whereas FEP students follow the mainstream curriculum in classrooms with ELO students. One possible explanation for these results is then that FEP students might not be sufficiently prepared for mainstream English classes and slow down the progress of these classes for ELO students. This result is surprising however in light of the negative peer effects of ELL students on the math performance of ELO students shown earlier. However, it is possible that ELL students share the classroom with ELO students for math, but not for ELA classes. Also, as mentioned earlier, ELL students receive additional English Language Development help to make sure that they make adequate progress in English. This additional help, which does not exist for math, could allow English teachers to not change the content of the English classes when teaching to both ELO and ELL students in the same classroom.

Lastly, in tables 3.4 and 3.5 I study how the peer effects of ESL students vary based on the average percentage of students in a school who are ESL. This type of analysis is not possible in other studies as most schools in these other studies have very few ESL students. In addition, following Friesen and Krauth (2011), in these tables I also look at how the peer effects of ESL students differ based on the language that they speak at home. This is important as heterogeneous impacts provide policy makers with guidance when thinking about ways to address the peer effects of ESL students.

When looking at math test scores, I find no convincing evidence that the peer effects of ESL students vary based on the average share of students in a school who are ESL. The peer effects of ESL students on math test scores are only negative and statistically significant for students who do not speak Spanish or an East-Asian language at home. This could be due to the fact that schools have less experience assisting students who speak these other languages at home as they do not encounter such students frequently. When looking at English test scores, the peer effects seem to be most negative at schools with the highest shares of students who are ESL. However, we cannot reject the hypothesis that the coefficient in the top third of schools in terms of the share of students who are ESL is the same as the coefficients in the bottom or middle third of schools. As with math scores, the coefficient is least negative for Spanish-speaking ESL students. In fact, the peer effects of the ESL students who speak a language other than Spanish at home are negative and around 3 to 4 times larger than the peer effects of ESL students who speak Spanish at home, a difference that is statistically significant.

3.6 Discussion and Conclusion

In this paper I estimated the peer effects of students who do not speak English at home on students who speak English at home using publicly available data for students in grades 2 to 6 in California in the year 2012-2013. In contrast to the previous literature, I did not just focus on the peer effects of English Language Learners but rather on the peer effects of the larger group of students who do not speak English at home as the share of students who are English Language Learners in any given school, year and grade is more likely to be endogenous. I identified peer effects by exploiting variation in the distribution of ESL students across grades within a given school in the school year 2012-2013.

I found that higher concentrations of ESL students have no effect on average math test scores but have a small negative effect on ELA test scores of ELO students. Further analysis showed that these peer effects do not systematically differ based on the average share of students in a school who are

ESL. In addition, I found that the observed negative peer effects are driven almost entirely by the negative peer effects of the small group of ESL students who do not speak Spanish at home. Lastly, I also looked at the peer effects of FEP and ELL students separately. These estimates are to be treated more carefully though, as the share of students who are ELL and FEP is more likely to be endogenous. These analyses showed the negative effects on ELA test scores to be solely due to FEP students and also showed that higher concentrations of ELL students could have a negative effect on math test scores of ELO students.

These results have a number of implications. First, high levels of ESL students in California are unlikely to have a large impact on the academic performance of ELO students in math and English. To the extent that existing policies for ESL students in California can be copied by other states, this is important for appeasing worries of large negative peer effects in other states. Second, the definition of the group whose peer effects we study matters for the results. In particular, looking at ESL students, as opposed to ELL students that the previous US literature has focused on, I found negative peer effects on ELA but not on math. On the other hand, looking at ELL students, I found negative effects on math but not on ELA test scores. The fact that the coefficient on FEP is not statistically significant when looking at math test scores further suggests that these negative peer effects of ELL students on the math test scores of ELO students cannot completely be attributed to the share of ELL students being endogenous.

Most importantly, I found no negative peer effects of ESL students in some key areas. In particular, there are no negative peer effects of ELL students on ELA test scores of ELO students and very small or no negative peer effects in math and ELA of Spanish-speaking ESL students who make up the bulk of ESL students. Instead, I found negative peer effects of FEP students on ELA test scores and of ELL students on math test scores. In addition, I found that the estimated negative peer effects on ELA test scores are mostly driven by non-Spanish speaking ESL students. These patterns are consistent with a situation in which teachers have more experience dealing with Spanish-speaking ESL students and policies are focused on mitigating potential negative peer effects of ESL students in areas where we might expect them. It seems that large gains could still be made by focusing on ESL students where we might not expect negative peer effects such as ELL students in math classes and FEP students in ELA classes, and by focusing on non-Spanish speaking ESL students.

This paper provided evidence on the existence of peer effects of ESL students on ELO students in California. The above estimates can be used by policy makers to identify areas in which there is potential for policy improvements to mitigate negative peer effects of ESL students. Further research is needed however on the peer effects of ESL students on ELO students. In particular, one has to be cautious about the results of this paper as I relied on school by year fixed effects to control for selection into schools and as such did not have true exogenous variation in peer group composition. In addition, I also lacked information on the way ESL students are treated in individual schools and as such can say little about potential mechanisms driving the above results. Future research could try to identify such mechanisms by for instance exploiting policy changes such as the repeal of Proposition 227 in California in 2016 that dramatically affect how ELL students are taught.

Chapter 3 is currently being prepared for submission for publication of the material. The dissertation author was the sole author of this paper. Gaastra, Sieuwerd.

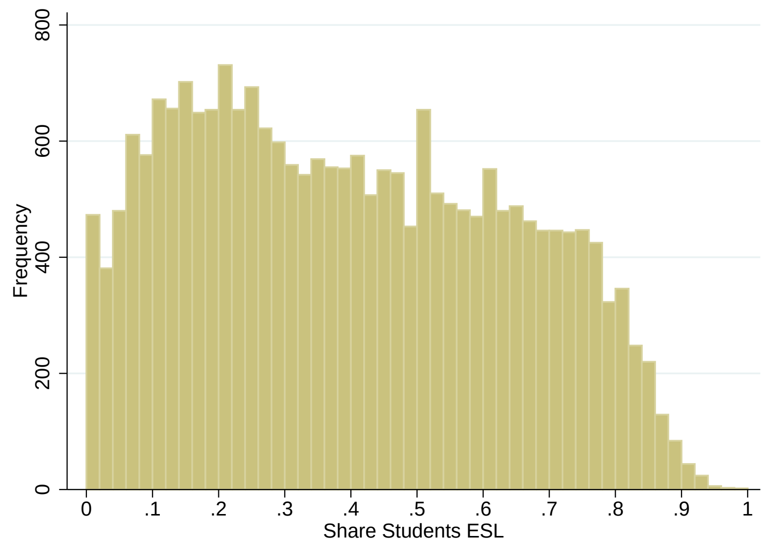


Figure 3.1: Distribution of the share of students who are ESL

Notes: Figure 3.1 shows the distribution of the share of students who are ESL across all school by grade combinations in my sample. In particular, this sample includes all public schools in California in 2012-2013 and only looks at grades 2 to 6. School by grade combinations with less than 10 ELO students are excluded as the outcomes of ELO students are withheld in these cases. The sample also excludes schools with only one school by grade combination as there is no variation across grades within these schools to exploit in these cases.

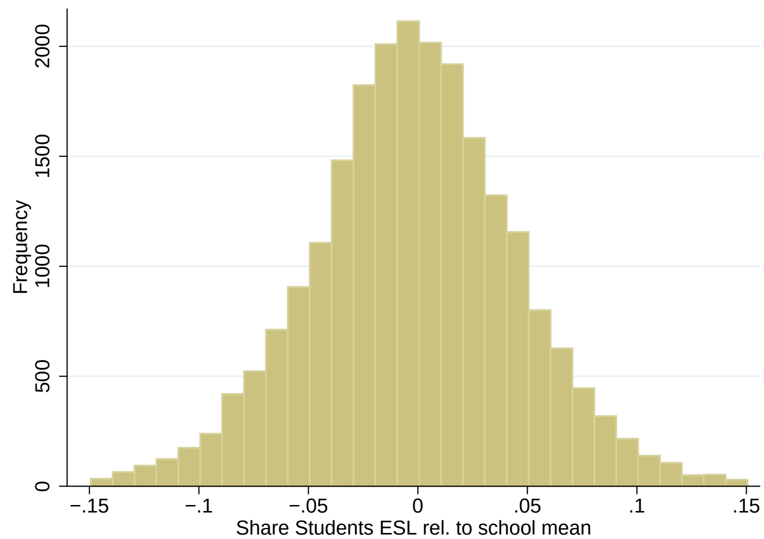


Figure 3.2: Distribution of the share of students who are ESL across grades within a school

Notes: Figure 3.2 shows the distribution of the share of students who are ESL relative to the school mean for school and year combinations in my sample (see description of figure 3.1). This figure captures over 99% of all observations in my sample. In absolute value, the average deviation from the school wide average in a particular grade is 3.7 percentage points.

Table 3.1: Descriptive Statistics

	Mean	Std. Dev.
Share Students ESL	0.42	(0.24)
Share Students ELL	0.26	(0.20)
Share Students ESL - Spanish	0.33	(0.26)
Share Students ESL - East Asian Lang.	0.036	(0.082)
Share Students ESL - Other Lang.	0.056	(0.075)
Share Students Male	0.51	(0.056)
Share Students Black	0.062	(0.094)
Share Students Hispanic	0.52	(0.30)
Share Students Asian	0.12	(0.16)
Share Students Other Ethnicity	0.038	(0.048)
Cohort size	86.4	(35.9)
Observations	22785	

Notes: Table 3.1 shows the summary statistics for my sample (see description of figure 3.1). Each observation is a school by grade combination and the summary statistics are calculated by weighting each observation's value by its cohort size. The category "East Asian Languages" includes Vietnamese, Cantonese, Mandarin, Korean and Filipino. The category "Other Languages" includes all languages other than Spanish and these East Asian languages. The percent of students who are "Other Ethnicity" refers to the percent of students who are not White, Hispanic, Black or Asian.

Table 3.2: Effect of ESL share on Math performance

	(1) No FEs	(2) FEs	(3) FEs + Controls	(4) ELL/FEP	(5) No top 2%	(6) No top 20%
Share Students ESL	-0.67***	-0.030	-0.035		-0.016	-0.096
	(0.028)	(0.037)	(0.039)		(0.044)	(0.079)
Share Students ELL				-0.094**		
				(0.042)		
Share Students FEP				0.022		
				(0.042)		
Grade FEs	No	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
School FEs	No	Yes	Yes	Yes	Yes	Yes
Observations	1082624	1082624	1082624	1082624	1060957	866051

Notes: Table 3.2 shows the result of estimating equation 3.1 with normalized average math CST test scores of ELO students as the dependent variable. The sample restrictions are the same as described below figure 3.1. Each observation is a school by grade combination and is weighted by the number of ELO students in that school and grade. Standard errors are clustered at the school level. The control variables include the shares of students in the school and grade who are male, White, Black, Hispanic and Asian. Columns (5) and (6) respectively exclude the top 2% and top 20% of all schools in terms of the maximum variation in the share of ESL students across grades within the school. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 3.3: Effect of ESL share on English performance

	(1) No FEs	(2) FEs	(3) FEs + Controls	(4) ELL/FEP	(5) No top 2%	(6) No top 20%
Share Students ESL	-0.81***	-0.12***	-0.12***		-0.11***	-0.21***
	(0.025)	(0.030)	(0.032)		(0.036)	(0.064)
Share Students ELL				-0.017		
				(0.035)		
Share Students FEP				-0.22***		
				(0.036)		
Grade FEs	No	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
School FEs	No	Yes	Yes	Yes	Yes	Yes
Observations	1078155	1078155	1078155	1078155	1056528	862406

Notes: Table 3.2 shows the result of estimating equation 3.1 with normalized average ELA CST test scores of ELO students as the dependent variable. The sample restrictions are the same as described below figure 3.1. Each observation is a school by grade combination and is weighted by the number of ELO students in that school and grade. Standard errors are clustered at the school level. The control variables include the shares of students in the school and grade who are male, White, Black, Hispanic and Asian. Columns (5) and (6) respectively exclude the top 2% and top 20% of all schools in terms of the maximum variation in the share of ESL students across grades within the school. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 3.4: Heterogeneous Effects of ESL share on Math performance

	(1) Bottom Third	(2) Middle Third	(3) Top Third	(4) Languages
Share Students ESL	-0.021	-0.021	-0.050	
	(0.11)	(0.067)	(0.054)	
Share Students ESL - Spanish				0.052
				(0.047)
Share Students ESL - East Asian Lang.				-0.11
				(0.11)
Share Students ESL - Other Lang.				-0.26***
				(0.078)
Grade FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
School FEs	Yes	Yes	Yes	Yes
Observations	357453	357109	368062	1082624

Notes: Table 3.2 shows the result of estimating equation 3.1 with normalized average math CST test scores of ELO students as the dependent variable. The sample restrictions are the same as described below figure 3.1. Each observation is a school by grade combination and is weighted by the number of ELO students in that school and grade. Standard errors are clustered at the school level. The control variables include the shares of students in the school and grade who are male, White, Black, Hispanic and Asian. In columns (1)-(3), the sample is restricted to schools in respectively the top, middle and bottom third of the distribution in terms of the average percentage of students in a school who are ESL. In column (4), the category "East Asian Languages" includes Vietnamese, Cantonese, Mandarin, Korean and Filipino. The category "Other Languages" includes all languages other than Spanish and these East Asian languages. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 3.5: Heterogeneous Effects of ESL share on English performance

	(1) Bottom Third	(2) Middle Third	(3) Top Third	(4) Languages
Share Students ESL	-0.046	-0.066	-0.10^{**}	
	(0.081)	(0.055)	(0.046)	
Share Students ESL - Spanish				-0.068^{*}
				(0.039)
Share Students ESL - East Asian Lang.				-0.28^{***}
				(0.092)
Share Students ESL - Other Lang.				-0.23^{***}
				(0.065)
Grade FEs	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
School FEs	Yes	Yes	Yes	Yes
Observations	356049	355671	366435	1078155

Notes: Table 3.2 shows the result of estimating equation 3.1 with normalized average ELA CST test scores of ELO students as the dependent variable. The sample restrictions are the same as described below figure 3.1. Each observation is a school by grade combination and is weighted by the number of ELO students in that school and grade. Standard errors are clustered at the school level. The control variables include the shares of students in the school and grade who are male, White, Black, Hispanic and Asian. In columns (1)-(3), the sample is restricted to schools in respectively the top, middle and bottom third of the distribution in terms of the average percentage of students in a school who are ESL. In column (4), the category "East Asian Languages" includes Vietnamese, Cantonese, Mandarin, Korean and Filipino. The category "Other Languages" includes all languages other than Spanish and these East Asian languages. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Bibliography

Afterschool Alliance, "America After 3PM: Afterschool Programs in Demand," 2014. Washington, D.C.

Ahn, T. and C. Jepsen, "The effect of sharing a mother tongue with peers: evidence from North Carolina middle schools," *IZA Journal of Migration*, 2015, *4* (5), 1–21.

Aizer, A., "Home alone: supervision after school and child behavior," *Journal of Public Economics*, 2004, *88*, 1835–1848.

Alstadster, A., A.S. Kulm, and B. Larsen, "Money or joy: The choice of educational type," *European Journal of Political Economy*, 2008, *24*, 107–122.

Altonji, J.G., L.B. Kahn, and J.D. Speer, "Trends in Earnings Differentials Across College Majors and the Changing Task Composition of Jobs," *American Economic Review, Papers and Proceedings*, 2014, *104* (5), 387–393.

—, —, and —, "Cashier or Consultant? Entry Labor Market Conditions, Field of Study and Career Success," *Journal of Labor Economics*, 2016, *34* (S1), S361–S401.

—, P. Arcidiacono, and A. Maurel, "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," in E.A. Hanushek, S. Machin, and L. Woessmann, eds., *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016, pp. 305–396.

Angrist, J. and J.S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2008.

Angrist, J.D. and I. Fernandez-Val, "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," in D. Acemoglu, M. Arrellano, and E. Dekel, eds., *Advances in Economics and Economet-*

rics, Cambridge University Press, 2013.

Arcidiacono, P., "Ability sorting and the returns to college major," *Journal of Econometrics*, 2004, *121*, 343–375.

_ , V.J. Hotz, and S. Kang, "Modeling college major choice using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 2012, *166*, 3–16.

Armour, P., R.V. Burkhauser, and J. Larrimore, "Using the Pareto Distribution to Improve Estimates of Topcoded Earnings," *Economic Inquiry*, 2016, *54* (2), 1263–1273.

Auerbach, A.J. and J. Slemrod, "The Economic Effects of the Tax Reform Act of 1986," *Journal of Economic Literature*, 1997, *35* (2), 589–632.

Baker, M., J. Gruber, and K. Milligan, "Universal Child Care, Maternal Labor Supply, and Family WellBeing," *Journal of Political Economy*, 2008, *116* (4), 709–745.

Barnett, W. S., "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes," *The Future of Children*, 1995, *5* (3), 25–50.

Beffy, M., D. Fougere, and A. Maurel, "Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter?," *The Review of Economics and Statistics*, 2012, *94* (1), 334–347.

Ben-Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, *75*, 352–365.

Benjamini, Y. and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, 1995, *57* (1), 289–300.

Berger, M.C., "Predicted Future Earnings and Choice of College Major," *Industrial and Labor Relations Review*, 1988, *41*, 418–429.

Bertrand, M., E. Duflo, and S. Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Blom, E., "Labor market determinants of college major," 2012. mimeo.

— , **B.C. Cadena, and B.J. Keys**, "Investment Over the Business Cycle: Insights from College Major Choice," 2015. IZA Discussion Paper 9167.

Boskin, M.J., "Notes on the Tax Treatment of Human Capital," 1975. NBER Working Paper 0116.

Bovenberg, A.L. and B. Jacobs, "Redistribution and education subsidies are Siamese twins," *Journal of Public Economics*, 2005, *89*, 2005–2035.

Bradley, E.S., "The Effects of the Business Cycle on Freshman Major Choice," 2013. mimeo.

Brunello, G. and L. Rocco, "The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores," *Economics of Education Review*, 2013, *32*, 234–246.

California Department of Education, "After School Education and Safety - Funding Profile," Available online at http://www.cde.ca.gov/fg/fo/profile.asp?id=3638 2014.

— , "Current Expense of Education & per-Pupil Spending," Available online at http://www.cde.ca.gov/ds/fd/ec/currentexpense.asp 2016.

Calonico, S., M.D. Cattaneo, M.H. Farrell, and R. Titiunik, "Regression Discontinuity Designs Using Covariates," 2016. Working Paper.

Carrell, S.E. and M.L. Hoekstra, "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2010, *2* (1), 211–228.

Cho, R.M., "Are there peer effects associated with having English Language Learner classmates? Evidence from the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K)," *Economics of Education Review*, 2012, *31*, 629–643.

Clarke, B., "Shocked out of your major: Do labor market shocks prompt major switching?," 2015. mimeo.

Coutts, E. and D. Feenberg, "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 1993, *12* (1), 189–194.

Dahl, G.B. and L. Lochner, "The Impact of Family Income on Child Achievement: Evidence from Changes in the Earned Income Tax Credit," *American Economic Review*, 2012, *102* (5), 1927–1956.

Dahl, G.B, K.V. Loken, and M. Mogstad, "Peer Effects in Program Participation," *American Economic Review*, 2014, *104* (7), 2049–2074.

Damm, A.P. and C. Dustmann, "Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?," *American Economic Review*, 2014, *104* (6), 1806–1832.

Davidson, R. and J. MacKinnon, "Wild bootstrap tests for IV regression," *Journal of Business and Economic Statistics*, 2010, *28*, 128–144.

Diette, T.M. and R.U. Oyelere, "Do Significant Immigrant Inflows Create Negative Education Impacts? Lessons from the North Carolina Public School System," 2012. IZA Discussion Paper Series 6561.

_ and _ , "Gender and Race Heterogeneity: The Impact of Students with Limited English on Native Students' Performance," *American Economic Review: Papers & Proceedings*, 2014, *104*, 412–417.

_ and _ , "Gender and Racial Differences in Peer Effects of Limited English Students: A Story of Language or Ethnicity?," 2016. IZA Discussion Paper Series 9661.

Dynarski, M., M. Moore, J. Mullens, P. Gleason, S. James-Burdumy, L. Rosenberg, C. Pistorino, T. Silva, J. Deke, W. Mansfield, S. Heaviside, and D. Levy, "When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: First-Year Findings," Technical Report, U.S. Department of Education 2003.

Feldstein, M., "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 1995, *103* (3), 551–572.

Figlio, D.N., "Boys Named Sue: Disruptive Children and Their Peers," *Education Finance and Policy*, 2007, *2* (4), 376–394.

Fischer, S., "The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence," 2016. mimeo.

Fletcher, J.M. and S.L. Ross, "Estimating the Effects of Friendship Networks on Health Behaviors of Adolescents," 2012. NBER Working Paper

18253.

Freeman, J. and B. Hirsch, "College majors and the knowledge content of jobs," *Economics of Education Review*, 2008, *27*, 517–535.

Friesen, J. and B.V. Krauth, "Ethnic enclaves in the classroom," *Labour Economics*, 2011, *18* (5), 656–663.

Garces, E., D. Thomas, and J. Currie, "Longer-Term Effects of Head Start," *The American Economic Review*, 2002, *92* (4), 999–1012.

Gaviria, A. and S. Raphael, "School-based peer effects and juvenile behavior," *Review of Economics and Statistics*, 2001, *83* (2), 257–268.

Geay, C., S. McNally, and S. Telhaj, "Non-native speakers of English in the classroom: What are the effect on pupil performance?," *The Economic Journal*, 2013, *123*, F281–F307.

Gentry, W.M. and R.G. Hubbard, "The Effects of Progressive Income Taxation on Job Turnover," *Journal of Public Economics*, 2004, *88*, 2301–2322.

Gottfredson, D., A. Brown Cross, D. Wilson, M. Rorie, and N. Connell, "Effects of Participation in After-School Programs for Middle School Students: A Randomized Trial," *Journal of Research on Educational Effectiveness*, 2010, *3* (3), 282–313.

_ , _ , _ , N. Connell, and M. Rorie, "A Randomized Trial of the Effects of an Enhanced After-School Program for Middle-School Students," 2010. Final report submitted to the U.S. Department of Education Institute for Educational Sciences.

Gould, E.D., V. Lavy, and M.D. Paserman, "Does Immigration Affect the Long-term Educational Outcomes of Natives? Quasi-experimental Evidence," *The Economic Journal*, 2009, *119*, 1243–1269.

Groen, J.A., "The Effect of College Location on Migration of College-Educated Labor," *Journal of Econometrics*, 2004, *121* (1-2), 125–142.

Gruber, J. and E. Saez, "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 2002, *84*, 1–32.

Hastings, J., C. Neilson, and S. Zimmerman, "The Effects of Earnings

Disclosure on College Enrollment Decisions," 2015. NBER Working Paper 21300.

Havnes, T. and M. Mogstad, "No Child Left Behind: Subsidized Child Care and Childrens Long-Run Outcomes," *American Economic Journal: Economic Policy*, may 2011, *3*, 97–129.

Heckman, J.J., "A Life-Cycle Model of Earnings, Learning, and Consumption," *Journal of Political Economy*, 1976, *84*, S11–S44.

Hendren, N., "The Policy Elasticity," *Tax Policy and the Economy*, 2016, *30*, 51–89.

Hill, L.E., J.R Betts, B. Chavez, A.C. Zau, and K. Volz Bachofer, "Pathways to Fluency: Examining the Link between Language Reclassification Policies and Student Success," Technical Report, Public Policy Institute of California 2014.

Hoxby, C., "Peer Effects in the Classroom: Learning from Gender and Race Variation," 2000. NBER Working Paper Series 7867.

Huang, D. and J. Wang, "Independent Statewide Evaluation of ASES and 21st CCLC After School Programs May 1, 2008-December 31, 2011," Technical Report, National Center for Research on Evaluation, Standards, and Student Testing (CRESST) 2012.

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

James-Burdumy, S., M. Dynarski, and J. Deke, "When Elementary Schools Stay Open Late: Results From the National Evaluation of the 21st Century Community Learning Centers Program," *Educational Evaluation and Policy Analysis*, 2007, *29* (4), 296–318.

_ , _ , and _ , "After-School Programs Effects on Behavior: Results from the 21st Century Community Learning Centers Program National Evaluation," *Economic Inquiry*, 2008, *46* (1), 13–18.

_ , _ , M. Moore, J. Deke, W. Mansfield, C. Pistorino, and E. Warner, "When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program: Final Report," Technical Report, U.S. Department of Education , Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance

(NCEE) 2005.

Jensen, P. and A. Wurtz Rasmussen, "The effect of immigrant concentration in schools on native and immigrant children's reading and math skills," *Economics of Education Review*, 2011, *30*, 1503–1515.

Julian, T., "Work-Life Earnings by Field of Degree and Occupation for People With a Bachelors Degree: 2011," *American Community Survey Briefs*, 2012, *11* (04), 1–4.

Kane, T.J., "The impact of after-school programs: Interpreting the results of four recent evaluations," 2004. Working Paper..New York, NY: W.T. Grant Foundation.

Karoly, L.A., P.W. Greenwood, S.S. Everingham, J. Hoube, M.R. Kilburn, C.P. Rydell, M. Sanders, and J. Chiesa, *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions*, RAND, 1998.

Katz, L.F. and D.H. Autor, "Changes in the wage structure and earnings inequality.," in O.C. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3, Elsevier, 1999, pp. 1463–1555.

Kirkeboen, L., E. Leuven, and M. Mogstad, "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, 2016, *131* (3), 1057–1111.

Kreiner, C.T., J.R. Munch, and H.J. Whitta-Jacobsen, "Taxation and the Long Run Allocation of Labor: Theory and Danish Evidence," *Journal of Public Economics*, 2015, *127*, 74–86.

Kugler, M.R., "After-School Programs Are Making a Difference," *NASSP Bulletin*, 2001, *85* (626), 3–11.

Larrimore, J., R.V. Burkhauser, S. Feng, and L. Zayatz, "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976-2007)," *Journal of Economic and Social Measurement*, 2008, *33* (2-3), 89–128.

Lauer, P.L., M. Akiba, S.B. Wilkerson, H.S. Apthorp, D. Snow, and M.L. Martin-Glenn, "Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students," *Review of Educational Research*, 2006, *76* (2), 275–313.

Lavy, V. and E. Sand, "The Effect of Social Networks on Students Academic and Non-Cognitive Behavioral Outcomes: Evidence from Conditional Random Assignment of Friends in Schoo," 2014. Working Paper.

_ **, M.D. Paserman, and A. Schlosser**, "Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom," *The Economic Journal*, 2012, *122*, 208–237.

_ **, O. Silva, and F. Weinhardt**, "The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools," *Journal of Labor Economics*, 2012, *30* (2), 367–414.

Le, T.N., I. Arifuku, L. Vuong, G. Tran, D.F. Lustig, and F. Zimring, "Community Mobilization and Community-Based Participatory Research to Prevent Youth Violence Among Asian and Immigrant Populations," *American Journal of Community Psychology*, 2011, *48*, 78–88.

Lee, D.S. and T. Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48*, 281–355.

Long, M.C., D. Goldhaber, and N. Huntington-Klein, "Do completed college majors respond to changes in wages?," *Economics of Education Review*, 2015, *49*, 1–14.

Lucas, R. E. Jr., "Supply-Side Economics: An Analytical Review," *Oxford Economic Papers*, 1990, pp. 293–316.

Malchow-Mller, N., S. B. Nielsen, and J. R. Skaksen, "Taxes, Tuition Fees and Education for Pleasure," *Journal of Public Economic Theory*, 2011, *13* (2), 189–215.

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

Montmarquette, C., K. Cannings, and S. Mahseredjian, "How do young people choose college majors?," *Economics of Education Review*, 2002, *21* (6), 543–556.

Nielsen, A. Skyt and A. Vissing-Jorgensen, "The Impact of Labor Income Risk on Educational Choices: Estimates and Implied Risk Aversion," 2006. mimeo.

Office of Juvenile Justice and Delinquency Prevention, "OJJDP Statistical Briefing Book," Available online at http://www.ojjdp.gov/ojstatbb/offenders/qa03301.asp?qaDate=2010 2014.

Ohinata, A. and J.C. van Ours, "How immigrant children affect the academic achievement of native Dutch children," *The Economic Journal*, 2013, *123*, F308–F331.

Oreopoulos, P., T. von Wachter, and A. Heisz, "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 2012, *4* (1), 1–29.

Patacchini, E., E. Rainone, and Y. Zenou, "Dynamic aspects of teenage friendships and educational attainment," 2011. CEPR Discussion Paper No. DP8223.

Poterba, J. and J. Hausman, "Household Behavior and the Tax Reform Act of 1986," *Journal of Economic Perspectives*, 1987, *1* (1), 101–119.

Powell, D. and H. Shan, "Income Taxes, Compensating Differentials, and Occupational Choice: How Taxes Distort the Wage-Amenity Decision," *American Economic Journal: Economic Policy*, 2012, *4*, 224–247.

Rivkin, S.G., E.A. Hanushek, and J.F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 2005, *73* (2), 417–458.

Saez, E., "Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications," 2004. NBER Working Paper 10273.

_ , J. Slemrod, and S.H. Giertz, "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 2012, *50* (1), 3–50.

Short, K., "Experimental Poverty Measures: 1999," Technical Report, U.S. Census Bureau 2001. Current Population Reports: Consumer Income.

Sickmund, M., H.N. Snyder, and E. Poe-Yagamata, "Juvenile Offenders and Victims: 1997 Update on Violence," Technical Report, Office of Juvenile Justice and Delinquency Prevention 1997.

Stinebrickner, R. and T.R. Stinebrickner, "A major in science? Initial beliefs and final outcomes for college major and dropout?," *Review of*

Economic Studies, 2014, *81* (1), 426–472.

Trostel, P.A., "The Effect of Taxation on Human Capital," *Journal of Political Economy*, 1993, *101*, 327–350.

What Works Clearinghouse, "Procedures and Standards Handbook Version 3.0," Available online at ¡http://ies.ed.gov/ncee/wwc¿ 2014.

Wiswall, M. and B. Zafir, "Determinants of College Major Choice: Identification using an Information Experiment," *Review of Economic Studies*, 2015, *82* (2), 791–824.

— **and** — , "How Do College Students Respond to Public Information about Earnings?," *Journal of Human Capital*, 2015, *9* (2), 117–169.