Title
Essays on Human and Social Capital Accumulation

Author
Fischer, Stefanie Jane

Publication Date
2015

UNIVERSITY OF CALIFORNIA
Santa Barbara

Essays on Human and Social Capital Accumulation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Stefanie Jane Fischer

Committee in Charge:

Professor Kelly Bedard, Co-Chair

Professor Peter Kuhn, Co-Chair

Professor Shelly Lundberg

June 2015

The Dissertation of
Stefanie Jane Fischer is approved:

Professor Shelly Lundberg

Professor Peter Kuhn, Committee Co-Chairperson

Professor Kelly Bedard, Committee Co-Chairperson

June 2015

Essays on Human and Social Capital Accumulation

To my mother, I would never have made it through this without your love, support, and unwaivering belief in me.

Acknowledgements

Many people have supported me throughout this six-year long marathon, and for that I am truly grateful. Without the support of my family, friends, fellow graduate students and committee, this feat would not have been possible. My parents have been my constant support – always believing in me even when I didn't believe in myself. They provided endless amounts of encouragement, offered an ear when I needed to vent, and were always ready to celebrate the smallest victories. I thank my mother for helping me see the positive side to everything – even in those times when there didn't seem like there was one. I thank my dad for instilling in me, from a very young age, a strong work ethic. And, I owe both of them credit for helping shape my research agenda – my dad for encouraging me throughout my life to be intellectually curious, my mom for sharing her passion for public education, and both of them for teaching me to care about the wellbeing of others. Thank you to my brother and sister-in-law for their technical support at all hours of the day and night, to my grandmother for teaching me to be a "pushy broad", and to my grandfather for taking me to my first college lectures at a young age sparking my love for learning. Thank you to the Robinson's for being my home-away-from-home.

I am forever grateful to my advisor, Kelly Bedard and to the other members of my committee, Peter Kuhn and Shelly Lundberg. Thank you to Kelly for the countless hours spent mentoring and guiding me in my research and in life. Kelly has pushed me to my highest potential with patients and kindness, and I would not be where I am today without her as a role model. I am humbled to have had the opportunity to work with such an incredible economist and person. Thank you to Peter Kuhn for diligently reading and re-reading drafts, always providing insightful comments. I thank my two coauthors Daniel Argyle and Christiana Stoddard .

Without my fellow classmates I would have given up long ago. Thanks to Ryan Abman, Tom Zimmerfaust, Vedant Koppera, and Sara Sutherland for being the best first year study buddies anyone could ask for. Thank you to Jenna Stearns for the reserach support and friendship, I should have added her to my committee long ago, and to Adam Wright for teaching me how to teach and for keeping things light. Thank you to the Education Working Group for helpful comments. Last, I thank all of my friends near and far, especially the good folks of Santa Barbara, for keeping me balanced and giving me perspective. Thank you to Ann Micka and Corey White for Saturday morning East Mountain loop bike rides, and the girls (Fran, Kelsey and Julie) for getting me out in the mountains and for making me splurge on good dinners every now and then. The past six years, as so many of you know, have been a wild ride – with plenty of triumphs and failures along the way. Thank you for lifting me up and for being part of it.

Curriculum Vitæ

Stefanie Jane Fischer

Education

Ph.D., Economics, University of California Santa Barbara, 2015

M.A. Applied Economics, Montana State University, 2009

B.S. Mathematics, Minor: Economics, Linfield College, 2005

Research Fields

Labor, Applied Econometrics, Applied Microeconomics, Education, Health

Committee Members

Kelly Bedard (Co-Chair), Peter Kuhn (Co-Chair), Shelly Lundberg

Honors and Fellowships

Experimental Research Grant (UCSB Graduate Division), 2013 and 2015

Humanities and Social Sciences Research Grant (UCSB Graduate Division), 2013

Broom Center for Demography Graduate Fellow (UCSB), 2012- 2015

Walter Mead Graduate Fellowship (UCSB), 2009-2010

President's Fellowship (Montana State University), 2007-2008

Faculty Merit Scholarship (Linfield College), 2001-2005

Teaching Experience

Introduction to Econometric Theory (Instructor)

Introduction to Econometric Theory (Teaching Assistant)

Intermediate Microeconomics (Teaching Assistant)

Principles of Economics (Teaching Assistant)

Financial Management (Teaching Assistant)

Public Finance (Teaching Assistant)

Conferences and Seminar Presentations

2015: UC Irvine, Utah State, Colgate University, Sonoma State, Cal Poly (SLO), CNA Corporation

2014: UC Santa Barbara, WEAI's Graduate Student Dissertation Workshop (Denver), Poster Presentation at All California Labor Conference (UC Berkeley)

2013: Western Economic Association International Conference (Seattle), Poster Presentation at All California Labor Conference (UC San Diego)

2012: Poster Presentation All California Labor Conference (San Francisco)

Professional Experience

Invited member of Human Capital Research Working Group (UCSB)

Constructed and administered large-scale field experiment on optimal group size - sampled 1,800 students

Research Assistant for Professor Kelly Bedard, 2013 and 2011

Referee for Economics of Education Review (2), Population Research and Policy Review (1)

Abstract

Essays on Human and Social Capital Accumulation

Stefanie Jane Fischer

This dissertation consists of three essays in applied microeconomics that address topics of human and social capital accumulation. The first essay addresses the topic of women and STEM (science, technology, engineering and math.) Women are substantially less likely than men to graduate college with a STEM degree. This paper investigates whether class composition can help explain why women are disproportionately more likely to fall out of the STEM "pipeline". Identification comes from a standardized enrollment process at a large public university that randomly assigns freshmen to different mandatory introductory chemistry lectures. Using administrative data, I find that women who are enrolled in a class with higher ability peers are less likely to graduate with a STEM degree, while men's persistence in STEM is unaffected by class composition. I also show that the decline for women is most pronounced for those in the bottom third of the ability distribution. I rule out the possibility that this is driven solely by grades because both men and women receive higher grades in classes with higher ability peers. Overall, these results suggest that class composition as an important factor in determining STEM persistence for women and provide a novel explanation for part of the STEM gender gap in post-secondary education.

The second essay, co-authored with Daniel Argyle, examines the effect of the four-day school week in rural Colorado on juvenile crimes. Four-day school weeks are becoming more common nationwide especially in rural areas. Those affected by the policy spend the same number of hours in school each week as students on a typical

five-day week, however the four-day week schedule essentially reallocates unstructured time into larger blocks creating a three-day weekend every weekend, since treated students for the most part have Fridays off. Our difference-in-difference estimates for rural Colorado indicate that switching all students in a county from a five-day week to a four-day week increases juvenile arrests for property crimes, in particular larceny, by about 80%.

In the third essay, co-authored with Christiana Stoddard, we study the academic achievement of American Indians. The academic achievement of American Indians has not been extensively studied. Using NAEP supplements, we find that the average achievement relative to white students resembles other disadvantaged groups. However, there are several differences. Family characteristics explain two times as much of the raw gap as for blacks. School factors also account for a larger portion of the gap than for blacks or Hispanics. The distribution is also strikingly different: low performing American Indian students have a substantially larger gap than high performing students. Finally, racial self-identification is more strongly related to achievement, especially as American Indian students age.

Contents

List of Figures

List of Tables

Chapter 1

The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence

1.1 Introduction

Women exit STEM majors at much higher rates than men, which in part contributes to the gender-wage gap because they miss out on the sizable premiums associated with employment in these fields.[1] To date, little is known regarding why women drop out of the STEM pipeline in college. Recent research finds that the instructor-student gender match matters somewhat (Carrell et al. (2010) and Hoffmann and Oreopoulos (2009)),

[1] Paglin and Rufolo (1990); Murnane et al. (1995); Grogger and Eide (1995); Brown and Corcoran (1997); Weinberger (1999); Weinberger (2001); Murnane et al. (2000); Rose and Betts (2004)

and that women's responsiveness to grades explains some of the phenomenon as well (Rask and Tiefenthaler (2008) and Ost (2010)), but much remains unexplained. Gaining a better understanding of the factors that cause women to leave STEM fields during college is important for developing policy aimed at bolstering this group's STEM persistence. This paper examines a novel pathway: *To what extent does the composition of a woman's introductory university STEM course impact STEM persistence?* By and large, these courses are large and competitive, and to the extent that women are more sensitive to these environments, a classroom composition which magnifies these features may induce marginal women to leave.

I investigate this question using administrative data from a large public research university. A unique feature of this particular setting is the quasi-experimental way in which students are assigned to their first chemistry course, which is a mandatory prerequisite for nearly all STEM majors at most universities, and therefore a significant gateway to success in STEM. Exploiting this quasi-experimental setting, I show that being in a class with higher ability peers reduces the probability that women graduate with a STEM degree and has no effect on men. More specifically, a 15% increase in the number of high ability students in a General Chemistry lecture (one standard deviation) reduces the probability that the average woman graduates with a STEM degree by 3.1 percentage points (6.8%). As one might expect, I further show that the effect is strongest in the bottom third of the math ability distribution. I rule out grades in the same course as the underlying mechanism by showing that there is a positive relationship for both men and women between peer ability and grades in the STEM gateway course analyzed. These results are informative for at least two reasons. One, they are the first to show that classmate influences are an important factor in determining students' academic success in higher education, at least for women. Two, in contrast to most previous work (discussed below), I focus on STEM major completion rather than grades because STEM completion closely relates to occupation choice, which is an important piece to the gender wage gap story (Murnane et al. (2000) and Rose and Betts (2004)).

Very little is known regarding post-secondary classroom composition effects because isolating the causal impact is difficult. Often students, or more indirectly administrators, influence the student make-up of a classroom. Several studies at the elementary school level, which for the most part rely on data from the large scale randomized experiment Project STAR, find a positive relationship between average classmate ability and achievement (Whitmore (2005), Hanushek et al. (2003), Boozer and Cacciola (2001), and Hoxby (2000)). Whether these results extend to a higher education setting, however, is an open question.

To date, the most convincing peer effects study in higher education – which estimates small positive effects on freshman grade point average – relies on the random assignment of students from the United States Air Force Academy to squadrons; which are essentially cohorts (Carrell et al. (2009)).[2] This peer group measure is an improvement over previous studies, which define dorm-mates as the peer group, because squadrons capture a more comprehensive set of students' peer interactions.[3] None of these estimates, however, capture the effects that students within a classroom may have

[2] Lyle (2007) uses a similar military dataset (USMA) and cohort approach and finds no evidence of peer effects. A drawback of both of these studies is that students from military institutions likely are not representative of the general university population, especially women.

[3] The following studies use the random assignment of students to dorms to estimate peer effects and find mixed results. Stinebrickner and Stinebrickner (2006) find small positive peer effects on grades for women. Zimmerman (2003) and Sacerdote (2001) find small positive peer effects on students' grades, grade point average, and the take-up of social networks such as fraternities/sororities. Foster (2006) finds no evidence of peer effects.

on individual outcomes because dorm-mates and squadron members do not necessarily attend the same classes.

There are several ways in which being assigned to a college lecture with relatively higher ability peers could affect students' outcomes. On one hand, this type of environment could be performance enhancing. Students may benefit directly from higher ability classmates through knowledge spillovers during class, office hours, or out-of-class group study sessions. Additionally, the average class ability can affect the classroom standard and students may be motivated to work harder to keep up with their high achieving peers, consequently earning higher grades.

On the other hand, a high achieving classroom environment may be harmful in more subtle ways by negatively impacting self-perception. The higher the ability of the peers in a classroom, the harder it is to be ranked highly. This may make the environment more competitive. While in many situations competition can improve performance, contest theory suggests that large gaps in skills between competitors can have the perverse effect of reducing effort incentives. As such, marginal students in this "small fish in a big pond" environment may feel relatively weaker and either reduce their effort resulting in lower grades in STEM courses or exit the STEM pipeline altogether. Brown (2011) provides empirical evidence for this theoretical prediction by showing that the presence of a superstar in a PGA golf tournament is associated with lower performance by the other competitors. In a related vein, Niederle and Vesterlund (2007) and Garratt et al. (2013) show that women shy away from competition more than men and that the gender performance gap is exacerbated under competition (Gneezy et al. (2003)). These findings suggest that women may be more discouraged by the competitive STEM climate than men.

At least three aspects of the quasi-experimental nature of this setting make it favorable for studying class composition effects. First, the unique no-priority registration process at this institution leaves little room non-random class enrollment (see Section 3.2). Second, the introductory STEM course General Chemistry is a required prerequisite for most STEM majors and students cannot circumvent the course by apply Advanced Placement credits eliminating another avenue for selection. Third, the enrollment of upperclassmen into this introductory course generates substantial exogenous variation in classroom composition (see Section 3.3)

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 discusses identification and outlines the empirical strategy. Section 4 reports the results. Section 5 concludes.

1.2 Data

All data in this study are drawn from the University of California Santa Barbara (UCSB) administrative data system. UCSB is a research intensive public university with a large undergraduate population. These data include all students who are enrolled in the required introductory university STEM course, General Chemistry (CHEM1A), in a fall quarter during the years 1997 through 2007, and follows each of them through graduation. General Chemistry is a particularly favorable setting to study STEM because at UCSB, and at the majority of other universities, this course is the first prerequisite for most STEM majors[4] and tends to be difficult and competitive.[5] This gate-

[4] Appendix Table A1 lists the STEM majors that require CHEM1A.

[5] One indication of the relative difficulty of the CHEM1A sequence at UCSB, is the large fraction of campus tutoring services allocated to it. In the academic year 2012-2013, about 25% of all the students that used the Campus Learning Assistance Services (CLAS) were General Chemistry students.

way course is a yearlong sequence consisting of CHEM1A, CHEM1B and CHEM1C.[6] Moreover, STEM majors at UCSB are required to take the course at the university to advance to any follow-on courses. Students are not able to apply Advanced Placement credits to bypass this prerequisite, a strategy that is often used to circumvent introductory math and statistics.

The data contain two types of students, on-track (freshman) and late-track (sophomore or higher). On-track students are defined as taking CHEM1A in the fall quarter of their first year at the university. Late-track students are upperclassmen taking CHEM1A in a fall quarter other than their freshman year. Of this sample, roughly 85% of the observations are on-track students. The other 15% of students are late-track. There are 12,230 on-track students in the sample. Conditional on being a late-track student, 1,291 are upperclassmen who entered the university as freshmen and 621 are transfers. Anecdotally, there are several reasons to believe that late-track CHEM1A students differ on observable and unobservable characteristics from on-track students. First, a portion of the late-track students are transfers. Transfer students typically come from the local city college and, on average, have lower high school grades and socioeconomic characteristics. Second, upperclassmen enrolled in CHEM1A, even if they are not transfer students, are behind schedule in their major since CHEM1A is the gateway course for almost all other STEM courses/majors. They are either behind because they switched to a STEM major at some point after their first year or because they needed a year of preparatory courses – remedial math and science – before taking CHEM1A.

Table 1.1 presents evidence that on-track students are on average higher achieving than late-track students on observable characteristics. Each column is a separate re-

[6] CHEM1A is offered in the fall, CHEM1B in the winter and CHEM1C in the spring.

gression where the outcome is a different predetermined student characteristics. The variable of interest is a 0-1 indicator for being on-track. Year fixed effects are also included in the specification. As reported in Columns 2 and 4, late-track students appear to have lower high school grade point averages and SAT math scores; two characteristics that are predictors of STEM success. In fact, late-track students on average have two-thirds of a letter grade lower high school grade point average (see Figure 1.1).

One desirable feature of this dataset is that on-track students (freshmen) face a no-priority registration policy. On the other hand, late-track students have the ability to selectively enroll in CHEM1A. To alleviate concerns of possible selection issues on this margin, I analyze the outcomes of on-track students only. The only way in which late-track students enter the analysis is through the composition variable (the share of on-track students per lecture), which is the regressors of interest.[7]

Table 1.2 presents summary statistics for the main sample, which only includes on-track students. From 1997-2007 there are 46 CHEM1A lectures taught by 13 different instructors. The average lecture size is 329 students. On average, on-track students make up 85% of each lecture; the minimum is 71% and the maximum is 96%.[8] The main outcome of interest is STEM completion, defined as graduating with a STEM major from UCSB within five years. Among entering freshmen who take CHEM1A, the average STEM completion rate is 53% for men and 45% for women. Other outcomes used in this analysis are a student's grade in CHEM1A, whether a student takes the direct follow-on course (CHEM1B), and a student's grade in CHEM1B. The average CHEM1A grade for males is a 2.65 GPA (on a 4 point scale) and a 2.49 for women.

[7] While late-track students have the ability to selectively enroll, Table 1.4 shows that the observable characteristics for this group are orthogonal to class composition. See Section 3.3 for more details.

[8] If the sample is restricted to exclude CHEM1A lectures with 100 or fewer students, the percent freshmen per lecture ranges from 75-96%.

Of all freshmen who take CHEM1A, 85% of men and 80% of women continue on to CHEM1B where the average grade earned in CHEM1B for men and women is 2.59 and 2.58 respectively.

UCSB administrative data also include several socioeconomic measures: race, sex, high school grade point average, SAT math and verbal scores, type of high school (public or private), parents' highest education level, English proficiency and age. Limited information is also available regarding instructors and course times. These include an instructor's sex and a unique instructor identification number, as well as the year, day, and time of the lecture. These data are linked to students.

1.3 Empirical Strategy

1.3.1 Econometric Specification

The primary specification is the following linear probability model:

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 F_i * lnO_{tnd} + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_i + \phi_t + \rho_n + \varepsilon_{itnd} \tag{1.1}$$

The variable G_{itnd} denotes STEM major completion for an on-track student i who takes CHEM1A in year t with instructor n at time of day d; tnd uniquely identifies an individual lecture in a specific year. F is a female indicator variable and O_{tnd} is the number of on-track students in a specific lecture. The log transformation allows one to interpret the on-track estimate as a percent change and takes into account that a one student change is proportionately larger from a small base.[9] The coefficient β_2 captures the effect of the number of on-track students per lecture on the outcome for men. The coefficient on the interaction term $F_i * lnO_{tnd}$ is the differential effect of the number of on-track students per lecture for women. Thus, for women the percentage point change in STEM graduation associated with a percent increase in the number of on-track students is $\beta_2 + \beta_3$. C_{tnd} controls for several class level characteristics: the total number of students enrolled in a given lecture and the percent female. The lecture size variable includes the log of the total number of on-track and late-track students enrolled in a given lecture. X_i is a vector of student background characteristics including: race, if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. M indicates that the lecture took place in the morning (starting at 8 a.m. or 9 a.m.). Year fixed effects (ϕ_t) are included to control flexibly for time trends in STEM completion. Since many instructors appear repeatedly, I include instructor fixed effects (ρ_n) to control for time-invariant instructor differences. All standard errors are clustered at the lecture level (instructor/year/time of day).

1.3.2 The Assignment of Students to Lectures

The aim of this study is to understand the differential impact of the number of on-track students in a general chemistry lecture relative to late-track students on STEM completion for men and women. In order to interpret $\hat{\beta}_2$ and $\hat{\beta}_3$ as causal, the vari-

[9] In alternative specifications, I also use percent of on-track students in a lecture as the measure of class composition and obtain quantitatively similar results. This measure is less desirable because it assumes that the marginal effect is constant regardless of the base.

ation in number and type of on-track students per lecture must be uncorrelated with unobservable student and instructor characteristics. For example, student preference for lectures with a certain concentration of on-track students would draw a causal interpretation into question. For my purposes, a primary concern is that $\hat{\beta}_2$ and $\hat{\beta}_3$ will be biased if on-track students with similar characteristics systematically enroll in lectures that have a higher (or lower) share of on-track students. To mitigate possible selection issues, only students in the fall quarter of their first year are included in the sample; Hoffmann and Oreopoulos (2009) take a similar approach.[10] This sample of students combined with UCSB's no-priority freshman enrollment process and the Chemistry Department's strictly enforced "no switching" rule leaves little room for selective enrollment for freshmen in CHEM1A.

During my sample period, 95% of all first year students attend a fee-based two-day summer orientation either in June, July or August where they register for their fall quarter courses.[11] Importantly, there is no priority based registration during or before summer orientation. In each orientation session, a certain percent of the total seats available in a given "first year" course are made available to that particular orientation session. This equalizes the probability of enrolling in a certain lecture/course across all orientation sessions and eliminates the issue of students who attend earlier orientation dates getting all the "good" classes.

[10] Freshmen in their first quarter of school have virtually no information about the instructors or classroom composition since they enroll in their first quarter courses prior to the start of school and they have very little flexibility in their class schedules to allow them to be strategic in course registration.

[11] Each summer 12 freshman orientation dates are offered and students can attend the date of their choice.

At orientation each student is assigned to a group of 15 students.[12] With this group, students attend seminars about university life and register for their first quarter classes under the guidance of a trained orientation leader. Only one student in an orientation group is able to register at a time, as there is only one laptop per group. Within each group a registration queue is formed by random draw (i.e. creating registration position 1- position 15). The student who draws position 1 registers first. Registration opens at the same time for all groups within an orientation session. According to the Director of Orientation Programs and Parent Services at UCSB, the high demand for CHEM1A, the limited seats, the random registration order, and student's general lack of information about instructors makes strategic CHEM1A registration essentially impossible. The one avenue for selection that one might be concerned with is time of day preferences. While there is no evidence that this is happening in a differential fashion, all specifications include controls for time of day. Making selection even harder, the Chemistry Department strictly enforces an no switching policy. A student can only switch lectures during the first week of the quarter and he must find a student in his desired lecture to replace him in his original one: a one for one switch.[13]

Balance tests for the main sample (entering freshmen cohorts from 1997-2007) are reported in Table 1.3. Each column in this table corresponds to a separate regression where a different predetermined student characteristic is regressed on lnO_{itnd} and $F_i * lnO_{itnd}$. Year fixed effects are also included. If selection is present, the coefficient on lnO_{itnd} will attain statistical significance. Furthermore, if gender based selection exists,

[12] Students are placed into orientation groups by declared major, but groups within major are formed randomly.

[13] All information regarding freshman orientation and registration comes from an interview with Kim Equinoa (kim.equinoa@sa.ucsb.edu) who was the director of Orientation Programs and Parent Services at UCSB during the years in which the data for this analysis are from. Information on the Chemistry Department's no switching rule comes from the administrative office within the Chemistry Department.

i.e. all high achieving women enroll in lectures with a low share of on-track students while high achieving men enroll in lectures with a high share of on-track students, the interaction term $F_i * lnO_{itnd}$ will be significant. Although the coefficient on lnO_{itnd} is statistically significant in some cases, the magnitudes are nearly zero. More importantly for my purposes, there is even less evidence of a gender differential in selection across class composition. $F_i * lnO_{itnd}$ is statistically significant in a third of cases but again the magnitudes are minuscule. For instance, Column 2 shows that SAT math and the interaction term are positively related, however, the estimated coefficient indicates that an increase in the log of on-track students in a class of 15% has an additional positive effect for women of 1.7 SAT math points where the SAT math standard deviation is approximately 80 points. Not only are these estimates economically insignificant, the direction works against the main findings of this paper (women's STEM persistence and the average class ability are negatively related) as they are a downward bias. Based on the class composition, there is little evidence that students are selecting into lectures.

The one concern related to student selection that I cannot directly rule out is the possibility that students who attend summer orientation differ from those who do not. Students who do not attend summer orientation register for their fall classes in mid-September prior to the start of the quarter but after all orientation goers have registered. On-track student estimates will be biased if students who do not attend summer orientation are non-representative and systematically register for lectures based on the ratio of on-track to late-track students. If I could observe CHEM1A registration dates I could do a balance test on observables between students who attend summer orientation and those who do not. Since these data are unavailable for my sample period, I have instead obtained registration data for all freshmen enrolled in CHEM1A in fall 2013. Although these students are not in my main sample, the registration behavior should be similar, as the general structure of freshman registration is similar.

For this group of students – all freshmen enrolled in CHEM1A in fall of 2013 – I observe their CHEM1A registration date and time as well as demographics and CHEM1A instructor characteristics. Ninety percent of this sample attended a summer orientation/registration (slightly lower than the main sample). Comparing the observable characteristics of students who attend orientation and those who do not, the only group who is underrepresented in orientation attendance is underrepresented minorities; 38% of the orientation attending group are URMs compared to 49% of the non-orientation attending group. There appears to be no selection into orientation attendance by gender, parent's education level, whether one has a high school GPA in the top half of the distribution for the sample, whether one scores in the top half of the SAT math or SAT verbal distribution for the sample, type of high school one attended, and English language learner status. Most importantly, the data indicate that there is no statistically significant difference in the share of on-track students in a lecture or begin time of the lecture for those who attend summer orientation and those who do not. Appendix Table A2 reports these results.

1.3.3 Source of Variation in Classmate Ability

The approach used to identify classroom composition effects in this study is similar to that of Hoxby (2000). She estimates classroom peer effects among elementary age students by exploiting plausibly random variation in the gender and racial composition within a given grade and school over time. I exploit exogenous variation in the ability composition across CHEM1A lectures within a year and over time. Variation in CHEM1A classroom composition is driven by late-track student enrollment patters. Late-track students register for fall classes the previous spring, before on-track enrollment. Based on estimated freshmen fall enrollment, the university holds a fraction of the CHEM1A seats in each lecture for incoming freshmen; in some cases late-track students fill all of the seats allotted to them in a given lecture and in other cases they do not.

For instance, suppose for simplicity that there are 4 lectures and each has a maximum enrollment of 100. Further suppose that 30 percent of the seats (30 seats in this case) in each lecture in a given year are made available to late-track students. If in one of the lectures all 30 seats are filled, in the second only 25 are filled, and in the third and fourth only 20 and 15 respectively are taken, then variation will arise in the number of late-track students. The second stage of the process is that on-track students are assigned (virtually at random) to the remaining seats. Although in theory late-track students have the ability to select into specific lectures which is why I don't analyze their outcomes, the balance test reported in Table 1.4 show no sign this group is selecting into lectures based on the composition of the class. There is some evidence that better late-track students select morning lectures but, importantly for this paper's empirical approach, time of day is uncorrelated with the share of on-track students.

1.4 Results

Results from the main specification (Equation 1) – which estimates the differential impact of the number of on-track students in a class for men and women – are reported in Table 5. Column 1 of Table A5 reports results for the full sample and shows that increasing the number of on-track students in a class by 15% reduces the probability that a woman graduates with a STEM major by 3.1 percentage points (see Column 1, panel B). Increasing the number of on-track students by 15% in the average class is equivalent to adding 44 more on-track students to a class with 281 on-track students, which is about one standard deviation. To give context to the magnitude of the results, the average STEM graduation rate for women is 45% and 53% for men. Thus, a 15% increase in the number of on-tracks student in a lecture decreases the STEM graduation rate for an average woman from 45% to about 42%, which is a decrease of 6.6%. For men, Column 1 suggests that there is no statistically significant relationship between the ability of the students in his CHEM1A lecture and the rate at which he persists in STEM.[14]

While the aim of this analysis is to understand the total effect of a student's classmates on that student's STEM outcomes, it is worth noting that in theory the total estimated composition effect can embody three distinct effects; exogenous peer effects (also known as contextual effects), endogenous peer effects, and correlated effects. Correlated effects are present when individuals in the same group behave similarly because they have similar individual characteristics (Moffitt et al. (2001)). This is often caused by students self-selecting into a group. The random assignment of students to lectures ensures that the composition estimates are free of correlated effects. Exogenous peer effects arise when a student's classmates' predetermined observable characteristics (high school grade point average, SAT scores etc.) affect her outcomes. Endogenous peer effects are present when a student's classmates' outcomes affect her outcome. For

[14] I re-run Equation 1 using a probit model rather than a linear probability model and obtain similar results. I also use percent on-track students in the class as well as number of on-track students in the class rather than natural log of the number of on-track students and obtain similar results.

instance, although in practice this seems rather unlikely, if a student's classmates' decision to remain in STEM affects her decision to stay, then the composition estimates in this study will capture not only the the exogenous effect but also this endogenous effect.[15] That said, although in the peer effects literature it is often the goal of the empiricists to solve the reflection problem which involves isolating the exogenous effect net of the other two effects, in this this study I am interested in learning how classmates' affect a student's STEM completion. As such, my composition estimates are capturing this total effect and the presence of an endogenous effect does not undermine the empirical findings.

1.4.1 Sensitivity and Heterogeneity Analysis

One factor that threatens the causal interpretation of these results is student assignment to lectures with a particular composition based on characteristics that correlate with student STEM persistence. Although a balance tests reveal that students are statistically similar on observables across classes (Table 1.3), lectures added at the last minute to meet a larger than expected CHEM1A demand are a particular concern.[16] The data do not allow one to specifically identify whether a class is added at the last minute, however, the very small lectures (i.e. those with fewer than 100 students) are likely to be the added course if one exists. Importantly, these very small lectures also appear to be correlated with the percent of on-track students in the class. The average

[15] If one believes that endogenous effects are present and assuming that both the exogenous and endogenous effects are negative, my estimate of the "total effect" will overestimate the exogenous effect. That is, the estimate will be inflated by a social multiplier, which is commonly known as the reflection problem, and be more negative than the true exogenous effect.

[16] Although many years the Chemistry Department accurately estimates the demand for CHEM1A, there are cases where they add an additional lecture the week before the fall term begins.

percent of on-track students for these lectures is 72 compared to 86% for the entire sample. It is also the case that on-track students assigned to these smaller add-on lectures have the greatest potential to be non-representative. For instance, the small percent of on-track students who do not attend a summer orientation/registration session and also enroll in CHEM1A (which is on average 5% percent of an incoming freshman class) are most likely assigned to an add-on lecture during the first week of school. One would expect this non-summer orientation attending group of students to be less advantaged, thereby dampening the estimated on-track student effect found in the main specification.[17] Column 2 of Table 5 reports the estimates for the subsample which excludes lectures with fewer than 100 students; variation in percent on-track student per lecture ranges from 75 to 96%.[18] Results for this subsample indicate that potential add-on lectures are not driving the main findings. In fact, the magnitude of the estimated on-track student effects for women and men are not statistically different from the estimated effects using the whole sample.

Additionally, understanding which group of students is driving the main result is important for developing and implementing interventions. Columns 3-5 of Table 1.6 report results disaggregated by SAT math score. The effect is strongest for women in the bottom third of the SAT math distribution. A 15% increase in the number of on-track students in a class reduces the probability by 5.7 percentage points that a women in this SAT math group completes college with a degree in STEM, and this effect is statistically different from the estimated effects in the other two SAT categories (Columns 4 and 5). Consistent with the main finding, the men in all subsamples appear to be unaffected

[17] Non-orientation attending students are likely less advantaged because summer orientation is an additional cost. According to UCSB office of Orientation Programs and Parent Services, the most common reason students do not attend orientation is due to summer employment.

[18] A balancing test for the subsample is statistically the same as the balancing test for the main sample.

by the classroom composition. While it is intuitive that women in the lower part of the math ability distribution are the group most affected by classroom composition since they are the group most at risk of dropping out of STEM, these results oppose the findings in Carrel et al. (2010). They find that the group influenced by STEM interventions are women at the top of the SAT math distribution. In particular, they document that women in the top 25% of the SAT math distribution with female STEM instructors are more likely to graduate with a STEM major.

One might wonder if the results truly are a gender effect. It is possible that I am capturing an underrepresented minority effect or merely picking up the fact that all students at the bottom end of the SAT math distribution are less likely to graduate with a degree in STEM. Columns 1 and 2 of Table 6 report results from the specifications outlined in Equations 2 and 3 respectively. These models, which are extensions of equation (1), include a triple interaction term allowing one to disentangle differences in the on-track student effect across gender and race (Column 1), as well as gender and position in the SAT math distribution (Column 2).

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 URM_i + \beta_4 F_i * URM_i + \beta_5 lnO_{tnd} * URM_i + \beta_6 F_i * lnO_{tnd} + \beta_7 F_i * URM_i * lnO_i + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_i + \phi_t + \rho_n + \varepsilon_{itnd} \tag{1.2}$$

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 Low_i + \beta_4 F_i * Low_i + \beta_5 lnO_{tnd} * Low_i + \beta_6 F_i * lnO_{tnd} + \beta_7 F_i * Low_i * lnO_i + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_i + \phi_t + \rho_n + \varepsilon_{itnd} \tag{1.3}$$

URM_i denotes whether a student is an underrepresented minority (black, Hispanic, American Indian or Filipino) and Low_i indicates whether a student falls in the bottom third of the SAT math distribution for the sample in a given year. All other variables are as defined in Section 3. Column 1 shows that all women, regardless of race, have STEM persistence rates that are negatively affected by the number of on-track students in her CHEM1A class. As reported in Column 1 Panel B, URM and non-URM women experience a 3.0 percentage point decline in STEM persistence as a result of an increased number of on-track students. Again, there is no detectable class composition effect for men, URM or non-URM.

Results presented in Table 6 Column 2 further support a gender story. These results show that only women (and not men) in the bottom third of the SAT math distribution for the sample have STEM persistence rates that are affected. In fact, women in this group are 4.0 percentage points less likely to graduate in STEM as a result of a 15% increase in the number of on-track students in a class. The results for the women are statistically different from zero and statistically different from men in this same SAT math group.

$$G_{itnd} = \alpha_1 + \beta_1 F_i + \beta_2 lnO_{tnd} + \beta_3 B_i + \beta_4 F_i * B_i + \beta_5 lnO_{tnd} * B_i + \beta_6 F_i * lnO_{tnd} + \tag{1.4}$$

$$\beta_7 F_i * Low_i * lnO_i + \alpha_2 C_{tnd} + \alpha_3 X_i + \alpha_3 M_i + \phi_t + \rho_i + \varepsilon_{itnd}$$

Finally, socioeconomic status may also play a role in one's willingness to leave STEM when placed in a lecture with a higher share of on-track students. I use a similar triple difference specification – as outlined in Equation 4 – and examine the differential effect of on-track concentration on STEM persistence by gender and by parent's level of education. B_i indicates if a student has at least one parent with a bachelor's degree. Results presented in Column 3 of Table 6 show that all women, regardless of whether her parent is a college graduate, have an increased probability of exiting STEM. Consistent with all other specifications, the persistence rate for men in all subgroups seems to be statistically unrelated to the composition of the class. Together, these findings provide strong evidence that *women* in the bottom third of the math ability distribution are the group most affected by the ability of their classmates. There is no evidence to support the conjecture that it is merely reflecting minority status, being in the bottom of the ability distribution, or socioeconomic status.

1.4.2 Possible Mechanisms

These reported findings raise the question: *Why are women in the bottom third of the math ability distribution less likely to graduate with a STEM degree if their first experience with STEM is in a setting with higher ability classmates, and why are men unaffected by this factor?* Because little is known regarding post-secondary classroom composition effects and student outcomes in general, less is known about the mechanisms at work, and in particular why composition matters more for women. One relevant study – a project funded by the National Science Foundation and conducted by the Goodman Research Group – surveyed roughly 25,000 undergraduate women enrolled in engineering programs across 53 institutions between 1999-2001 with the goal of identifying ".. aspects of women's educational experiences that are critical to their retention in engineering." Although engineering students comprise only a small part of the greater group of women in STEM, lessons learned from the engineering study likely have some relevance for other STEM areas since many early prerequisites are common across STEM majors. This study finds that women who leave engineering are most likely to leave in their freshman or sophomore year. The top two reasons for leaving are: (1) they are dissatisfied with their grades and/or the heavy workload, and (2) they dislike the overall climate of the major, including the competitive nature and discouraging faculty and peers. Both of these factors may be directly influenced by the class composition.

Grades

First, I investigate grades in the initial course as a possible mechanism. In addition to the study by the Goodman Research Group, several studies in economics find that women are more responsive to grades than men, and as a result exit STEM majors.[19] If all CHEM1A classes are graded on a similar curve (i.e. 10 percent of the class earns an A grade, 35 percent earns a B grade etc.), then students in lectures with more high ability classmates will receive lower grades relative to their counterpart (those in lectures with fewer on-track students). For instance, suppose that there are two CHEM1A lectures with equal enrollment in a given fall quarter and one has more on-track students than the other. Relying on the fact that lectures with more on-track students are overall higher ability (Table 1.1), then a student receiving a score of 77% in the lecture with more on-track students will be assigned a lower final letter grade than if she was in a class with fewer on-track students.

To explore this possibility, I use a specification similar to Equation 1 with a student's CHEM1A grade as the outcome. CHEM1A grade is a variable taking on values from

[19] The average grade in STEM courses is much lower than humanities, social sciences, arts and interdisciplinary courses. Rask and Tiefenthaler (2008) and Ost (2010) show that women in STEM are more responsive to grades than men.

0 to 4. If grades are the underlying mechanism, students in classes with more on-track peers should receive lower final grades. The results presented in column 1 of Table 1.7, however, show no sign of this. In fact, both men and women experience a bump in CHEM1A grade as a result of an increase in the number of on-track students in a class. Columns 2-7 further show that this result is robust to a variety of subsamples. It appears that all students – male, female, URM, non-URM, the low SAT math scoring group and high SAT math scoring group – experience a marginal increase in their grade. These results are also robust to the subsample which excludes lectures with fewer than 100 students. Quantile regression results (available upon request) further reveal that a positive effect is detectable at all places in the distribution for both men and women which suggests that the overall grade distribution for classes with a higher share of on-track students is shifted to the right. Moreover, if grades are driving the result, controlling for CHEM1A grade in the main specification (Equation 1) should diminish the composition estimates. Table 1.8 reports such estimates and shows that the effect remains despite controlling for initial course grade, and the differential effect persists as well.

It is possible that the grade findings are a result of positive peer effects as it has been shown that higher ability peer groups elicit higher individual grades (Stinebrickner and Stinebrickner (2006); Han and Li (2009); Zimmerman (2003); Sacerdote (2001); Carrell et al. (2009)). Given the structure of the data, however, I cannot rule out that the positive grade effect is merely an artifact of instructors adjusting their grade scales based on the overall ability of the students. For instance, positive effects will emerge if instructors increase everyone's grade in a class because on-whole they are high achieving; a channel that instructor fixed effects will not capture because this behavior is time-varying.

Course Climate

General Chemistry is *competitive* and women, when given the option, are more likely to select out of competitive environments (Niederle and Vesterlund (2007), Garratt et al. (2013)). The competitive environment in the introductory chemistry course comes from at least three source, (1) the course is required for most STEM majors and students are required to keep a "C" average in the introductory sequence to advance to upper division courses, (2) because many STEM majors are graduate school and medical school prerequisites, students are motivated to maintain a high grade point average, and (3) grades are assigned based on a curve. Increasing the share of high ability students therefore increases competition in an already competitive environment.

Additionally, the composition of the class could affect students' *self-perception* about their immediate and future success in the major. Presumably, all students enter the initial course with an expectation about how they will do. Throughout the course they learn about their relative standing and update beliefs about themselves accordingly. Individuals in lectures with relativity higher ability classmates may adjust these believes differently than those who are not. For example, Pop-Eleches and Urquiola (2011) show that students who just make it into better high schools receive a bump in exam scores but also report feeling marginalized and relatively weaker compared to students who are placed in classes with lower ability classmates. To the extent that women's self-perception about their future success is more negatively affected by the ability of those around them, it could explain their much lower retention rate. Along these lines, if women are more *risk averse*, then the marginal women may switch to majors where they perceive having a higher chance of "making it" while marginal men gamble by staying in STEM. Consistent with this idea, Kuziemko et al. (2014) show that men are more likely to gamble to avoid low rank whereas women accept it.[20]

There are many reasons to believe that women might be turned off by the climate in STEM while men are not. As a result, marginal women may become discouraged and either exert lower effort or quit STEM altogether. Some evidence supporting the idea that women become discouraged is in the take-up of the direct follow-on course, CHEM1B. Twenty percent of women exit the general chemistry sequence after the initial course compared to 15% of men and some of this is attributable to the composition of the introductory course. Table 1.9 reports results for a linear probability model similar to Equation 1 where the outcome is equal to one if a student takes CHEM1B and zero otherwise. This table shows that women's and men's CHEM1B take-up is unrelated to CHEM1A classroom composition for the whole sample (Column 1), but when broken out by SAT math subgroups the results indicate that CHEM1B take-up for women in the bottom third of the SAT math distribution – which is the group with STEM major completion most affected by classroom composition – is negatively related to the number of on-track students in a class (Column 5). The estimated effect is statistically different from zero and statistically different from the effect for men in this same group, but due to the noisy estimates in the other subsamples, I cannot reject that the effect for women is the same as the estimated effects for women from the other

[20] Although data for this study comes from the laboratory and manipulates an individual's rank in the wealth distribution, it is reasonable that the detected behavioral response extends to a classroom ability distribution.

SAT math groups (the middle third and the top third). These results are robust to a specification that controls for CHEM1A grade as well.

Next, I estimate a small negative relationship between number of on-track students and CHEM1B grade for women and find no statistically significant effect for men (Table 1.10, Column 1). An increase in the number of on-track students of 15% is related to a reduction in a woman's CHEM1B grade of 0.08 grade points which is 27% of a letter grade. Consider this estimate understated (less negative than one would expect) because some students – particularly those near the bottom of the ability distribution, see Table 1.9 – have already exited the pipeline as a result of the introductory class composition and are no longer in the sample. It is very likely that this follow-on grade finding reflects a student's overall discouraged feeling – i.e. lack of effort – particularly if she feels marginalized in the introductory course. It is also possible that learning or mastering the fundamental skills needed to successfully complete a STEM degree is lower for women in introductory classes with higher quality students. However, this later explanation, is not supported in my data, as I find a positive relationship between the ability composition in a classroom and CHEM1A grade for women.

Finally, I show that women are responding to the composition of their introductory course by switching into majors that are relatively less quantitative.[21] Table 1.11, Column 2 reports that increasing the number of on-track students in a class by 15% leads to 3.3 percentage point increase in the probability that a woman graduates with a humanities, social sciences, art, or interdisciplinary major.[22] This table shows that women are

[21] Appendix Table A3 outlines by sex the percent of students in each major category at entry and at graduation.

[22] I get similar results when I exclude Economics (Econ, Econ-Math, and Econ-Accounting) and Psychology from this group.

still graduating (as shown in column 3) but on average they are graduating in majors that are, lower paying, less quantitative and arguably less competitive.

Although I can not directly point to the channel by which high ability classmates adversely affect women's STEM retention, I can rule out that the effect is operating through grades. There are, however, various other channels through which the climate may discourage women including competition, self-perception and risk aversion. Consistent with this notion, I show that women both directly after the initial course and at other points in their STEM major pursuit give-up on STEM as a response to the share of high ability classmates, in favor of majors that have a more female friendly environment.

1.5 Discussion and Conclusion

It has been well documented that women are less likely than men to persist in STEM majors and careers. This study targets a unique group of students, those taking General Chemistry in their first quarter of college, to better understand how one's first collegiate experience in STEM explains STEM major graduation rates. Relying on data containing roughly 12,000 first year university students from 11 entering cohorts between 1997-2007, I estimate how the ability of one's classmates, as measured by the share of on-track students per class, in a required STEM major course affects a student's STEM major completion.

In summary, women who are assigned to a STEM lecture with higher ability peers early in their university career are less likely to persist in STEM. I show that this result is driven by women in the bottom third of the SAT math distribution. Men in this same

SAT bracket (or any subsample for that matter) do not experience this same negative relationship between classmate ability and STEM persistence.

I have ruled out the possibility that women earn lower grades in classes with higher achieving classmates and as such are less likely to persist. The number of on-track students per class and grades in the initial course are positively related. On the other hand, I cannot rule out the possibility that women's decision to exit is a response to the climate created in classes with higher ability peers. In fact, I find some evidence consistent with a story that marginal women are deterred by the climate, become discouraged, and eventually exit to majors with a more female friendly environment. I show that at least some of these women are leaving the STEM pipeline immediately after the initial General Chemistry course, as a share of them do not persist to the direct follow-on as a result of the ability of their classmates in the initial course.

This study is the first to provide an analysis of the relationship between class composition and STEM degree completion in higher education, and to document the differential response by gender. Although these estimates do not provide direct policy implications, they do reveal (1) an important group for policy to target, and (2) suggest that, for women, a behavioral response may be present. Thus, if the goal of policy is to bolster the participation of women in STEM fields, deepening our understanding of the channels through which classmates affect a woman's STEM behavior is important.

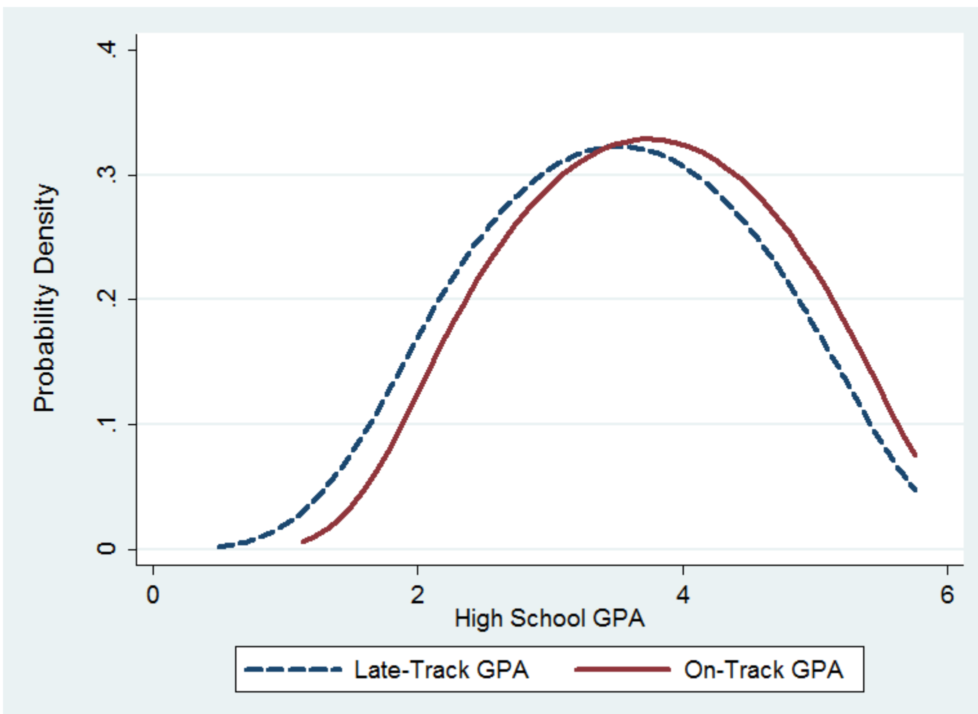

Figure 1.1: Differences in High School GPA Between On-Track and Late-Track

Table 1.1: On-Tracks are Relatively Higher Achieving than Late-Tracks

	On-Track	Late-Track	Diff. (1) - (2)
	(1)	(2)	(3)
Student Background Characteristics			
Women	0.49	0.52	-0.03***
	(0.01)	(0.01)	(0.01)
URM (underrepresented minority)	0.32	0.32	0.01
	(0.47)	(0.47)	(0.01)
High school grade point average	3.75	3.52	0.23***
	(0.32)	(0.44)	(0.01)
SAT math score	612.74	604.34	8.40***
	(80.84)	(80.95)	(2.25)
SAT verbal score	569.58	575.91	-6.34***
	(84.59)	(83.66)	(2.33)
Attended public high school	0.85	0.86	-0.01
	(0.36)	(0.35)	(0.01)
English is only language spoken in home	0.67	0.69	-0.02*
	(0.47)	(0.46)	(0.01)
No parent graduated from college	0.33	0.30	0.03***
	(0.47)	(0.46)	(0.01)
Outcomes			
Graduate with STEM major	0.49	0.44	0.06***
	(0.50)	(0.50)	(0.01)
Graduate	0.81	0.82	-0.014
	(0.40)	(0.38)	(0.01)
CHEM1A grade	2.57	2.26	0.31***
	(0.93)	(1.11)	(0.03)
Took follow-on course (CHEM1B)	0.83	0.63	0.20***
	(0.38)	(0.48)	(0.01)
Grade in follow-on course (CHEM1B)	2.58	2.53	0.05*
	(0.87)	(0.97)	(0.03)
Observations	12,230	1,935	14,165

Notes: On-track students are enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB from 1997 to 2007. Late-track students are enrolled in CHEM1A during this time frame but are taking the course as an upperclassman or transfer student. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Level of significance is indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard deviations are in parentheses.

Table 1.2: Summary Statistics (sample includes on-tracks only)

	Women (1)	Men (2)
Classroom Characteristics		
% on-track in a lecture	0.85	0.85
	(0.05)	(0.05)
CHEM1A lecture size	329.37	328.53
	(46.73)	(47.37)
Student Background Characteristics		
URM (underrepresented minority)	0.34	0.31
	(0.47)	(0.46)
High school grade point average	3.80	3.70
	(0.31)	(0.33)
SAT math score	587.59	636.41
	(78.16)	(75.49)
SAT verbal score	564.88	573.97
	(82.63)	(85.60)
Attended public high school	0.86	0.84
	(0.35)	(0.37)
English is only language spoken in home	0.69	0.65
	(0.46)	(0.48)
No parent graduated from college	0.36	0.29
	(0.48)	(0.46)
Outcomes		
Graduate with STEM major	0.45	0.53
	(0.50)	(0.50)
Graduate	0.82	0.80
	(0.39)	(0.40)
CHEM1A grade	2.49	2.65
	(0.95)	(0.91)
Took follow-on course (CHEM1B)	0.80	0.85
	(0.40)	(0.36)
Grade in follow-on course (CHEM1B)	2.58	2.59
	(0.87)	(0.87)
Observations	5,942	6,288

Notes: The sample includes only on-track students, those enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 1.3: Balance Tests – Are On-Track Students Selectively Enrolling?

	URM	SAT Math	SAT Verbal	H.S. GPA	Parent is College Grad	Public H.S.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Point Estimates						
Ln(no. of on-track)	-0.076***	-0.043	5.032	0.043**	0.064**	0.001
	(0.028)	(4.696)	(5.040)	(0.020)	(0.028)	(0.021)
Ln(no. of on-track) X Fem.	0.025	12.66**	7.110	-0.011	0.030	-0.053*
	(0.039)	(6.325)	(6.906)	(0.027)	(0.039)	(0.027)
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track						
The differential effect	0.30	1.80**	1.00	0.00	0.40	-1.00*
Women	-0.70*	1.80***	1.70**	0.00	1.00***	-1.00***
Men	-1.00***	-0.01	0.70	0.01**	1.00***	0.00
Observations	12,230	12,142	12,142	12,160	12,230	12,230

Notes: Each column is a separate regression and also includes year fixed effects. On-track students are those enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Level of significance is indicated as follows: *** p<0.01, ** p<0.05, * p<0.1. Standard deviations are in parentheses.

Table 1.4: Balance Tests – Are Late-Track Students Selectively Enrolling?

	URM	SAT Math	SAT Verbal	H.S. GPA	Parent is College Grad	Public H.S.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Point Estimates						
Ln(no. of on-track)	-0.003	-4.549	-1.5072	0.041	0.006	-0.045
	(0.051)	(11.6906)	(9.730)	(0.043)	(0.055)	(0.039)
Ln(no. of on-track) X Fem.	-0.056	-5.082	-6.653	0.062	0.020	0.068
	(0.078)	(14.890)	(14.680)	(0.065)	(0.081)	(0.056)
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track						
The differential effect	-1.00	-0.70	-0.90	0.01	0.00	1.00
Women	-1.00	-1.40	-1.10	0.01*	0.00	0.00
Men	0.00	-0.60	-0.20	0.01	0.00	-1.00
Observations	1,935	1,443	1,443	1,896	1,935	1,935

Notes: Each column is a separate regression and also includes year fixed effects. On-track students are those enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Level of significance is indicated as follows: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard deviations are in parentheses.

Table 1.5: The Effect of the Number of On-Track Students on STEM Major Completion

	Full Sample	Lectures > 100	Bottom 1/3 of SAT Math	Middle 1/3 of SAT Math	Top 1/3 of SAT Math
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Ln(no. of on-track)	-0.100	-0.037	-0.232	-0.71	-0.057
	(0.105)	(0.103)	(0.171)	(0.299)	(0.138)
Ln(no. of on-track) X Fem.	-0.112***	-0.137***	-0.166**	0.010	-0.055
	(0.030)	(0.045)	(0.067)	(0.061)	(0.077)
Instructor, Year, Time of day FE	X	X	X	X	X
Student Characteristics	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track					
Women	-3.10**	-2.40*	-5.70**	-0.90	-1.60
Men	-1.50	-0.50	-3.20	-1.00	-0.80
Observations	12,230	12,122	4,206	3,438	4,586

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** p<0.01, ** p<0.05, * p<0.1. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 1.6: Heterogeneity Analysis – STEM Major Completion for Various Subgroups

	URM Effect?	Low Ability Effect?	Low SES Effect?
	(1)	(2)	(3)
Panel A: Point Estimates			
Ln(no. of on-track)	-0.096	-0.116	-0.143
	(0.114)	(0.106)	(0.119)
Ln(no. of on-track) X Fem.	-0.125**	0.802*	-0.087
	(0.060)	(0.472)	(0.070)
Ln(no. of on-track) X URM	-0.050		
	(0.070)		
Ln(no. of on-track) X Fem. X URM	0.033		
	(0.101)		
Ln(no. of on-track)X bottom 1/3		0.013	
		(0.064)	
Ln(no. of on-track) X Fem. X bottom 1/3		-0.144*	
		(0.084)	
Ln(no. of on-track) X Parent col. grad			0.054
			(0.063)
Ln(no. of on-track) X Fem. X Parent col. grad			-0.042
			(0.087)
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track			
Non-URM – women	-3.10**		
Non-URM – men	-1.30		
URM – women	-3.30**		
URM – men	-2.00		
Bottom 1/3 – women		-4.00**	
Bottom 1/3 – men		-1.40	
Top 2/3 – women		-2.00	
Top 2/3 – men		-1.60	
Col. grad parent – women			-3.00**
Col. grad parent – men			-1.20
No col. grad parent – women			-3.20**
No col. grad parent – men			-2.10
Observations	12,230	12,230	12,230

Note: Each column is a separate specification. The Column 1 regression also includes a dummy for URM and an interaction term between URM and woman. URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. The Column 2 regression also includes a dummy for being in the bottom 1/3 of the SAT math distribution and the interaction between being in the bottom and a woman. The Column 3 regression also includes a dummy for having at least one parent with a college degree and the interaction between being that dummy and woman. Additionally, all three regressions include controls for percent female in a class, class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** p<0.01, ** p<0.05, * p<0.1. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students).

Table 1.7: The Effect of the Number of On-Track Students on CHEM1A Grade

	Full Sample	Lectures > 100	Non-URMs	URMs	Bottom 1/3	Middle 1/3	Top 1/3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Point Estimates							
Ln(no. of on-track)	0.831*	0.856*	0.799*	1.005*	1.376**	0.893**	0.766*
	(0.441)	(0.447)	(0.431)	(0.546)	(0.610)	(0.416)	(0.453)
Ln(no. of on-track) X Fem.	0.029	0.005	0.036	0.010	-0.079	-0.092	0.121
	(0.068)	(0.125)	(0.087)	(0.111)	(0.092)	(0.119)	(0.148)
Instructor, Year, Time of day FE	X	X	X	X	X	X	X
Student Characteristics	X	X	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track (0.3 grade points = 1 letter grade)							
Women	0.12*	0.12*	0.11*	0.14*	0.18**	0.11**	0.12**
Men	0.12*	0.12*	0.11*	0.14*	0.19**	0.13**	0.11*
Observations	12,230	12,122	8,281	3,949	4,206	3,438	4,586

Note: Each column is a separate specification. "Bottom 1/3" refers to the bottom 1/3 of the SAT math distribution. "Middle 1/3" and "Top 1/3" refer to the middle 1/3 and top 1/3 of the SAT math distribution for the class respectively. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 1.8: STEM Major Completion – Controls for CHEM1A Grade

	Full Sample	Lectures > 100
	(1)	(2)
Panel A: Point Estimates		
Ln(no. of on-track)	-0.256*	-0.185
	(0.140)	(0.139)
Ln(no. of on-track) X Fem.	-0.117***	-0.138***
	(0.034)	(0.053)
CHEM1A Grade	0.169***	0.169***
	(0.005)	(0.005)
Instructor, Year, Time of day FE	X	X
Student Characteristics	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track		
Women	-5.20***	-4.50**
Men	-3.60*	-2.60
Observations	12,230	12,122

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** p<0.01, ** p<0.05, * p<0.1. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 1.9: The Effect of the Number of On-Track Students on CHEM1B Enrollment

	Full Sample	Lectures > 100	Non-URMs	URMs	Bottom 1/3	Middle 1/3	Top 1/3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Point Estimates							
Ln(no. of on-track)	0.000	0.015	-0.039	0.110	-0.221	0.141	0.125
	(0.086)	(0.093)	(0.065)	(0.191)	(0.167)	(0.160)	(0.110)
Ln(no. of on-track) X Fem.	-0.045**	-0.058*	-0.049	-0.025	-0.098*	0.040	-0.035
	(0.019)	(0.031)	(0.037)	(0.051)	(0.050)	(0.044)	(0.047)
Instructor, Year, Time of day FE	X	X	X	X	X	X	X
Student Characteristics	X	X	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track (0.3 grade points = 1 letter grade)							
Women	-0.60	-0.60	-1.10	1.20	-4.50**	2.50	1.30
Men	0.01	0.20	-0.40	1.50	-3.10	2.00	1.60
Observations	12,230	12,122	8,281	3,949	4,206	3,438	4,586

Note: Each column is a separate specification. "Bottom 1/3" refers to the bottom 1/3 of the SAT math distribution. "Middle 1/3" and "Top 1/3" refer to the middle 1/3 and top 1/3 of the SAT math distribution for the class respectively. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 1.10: The Effect of the Number of On-Track Students on CHEM1B Grade

	Full Sample	Lectures > 100	Non-URMs	URMs	Bottom 1/3	Middle 1/3	Top 1/3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Point Estimates							
Ln(no. of on-track)	-0.431	-0.374	-0.288	-0.660**	-0.208	-0.625	-0.555*
	(0.260)	(0.281)	(0.320)	(0.287)	(0.373)	(0.585)	(0.276)
Ln(no. of on-track) X Fem.	-0.130*	-0.205**	-0.163*	-0.156	-0.096	-0.244	0.136
	(0.067)	(0.099)	(0.085)	(0.143)	(0.131)	(0.169)	(0.132)
Instructor, Year, Time of day FE	X	X	X	X	X	X	X
Student Characteristics	X	X	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track (0.3 grade points = 1 letter grade)							
Women	-0.08**	-0.08**	-0.06	-0.11**	-0.04	-0.12	-0.06
Men	-0.06	-0.05	-0.04	-0.09**	-0.03	-0.09	-0.08**
Observations	10,086	10,003	6,942	3,144	3,187	2,857	4,042

Note: Each column is a separate specification. "Bottom 1/3" refers to the bottom 1/3 of the SAT math distribution. "Middle 1/3" and "Top 1/3" refer to the middle 1/3 and top 1/3 of the SAT math distribution for the class respectively. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** p<0.01, ** p<0.05, * p<0.1. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Table 1.11: Where are the Women Going?

	Graduate in Non-STEM	Graduate
	(1)	(2)
Panel A: Point Estimates		
Ln(no. of on-track)	1.22	-0.035
	(0.099)	(0.113)
Ln(no. of on-track) X Fem.	0.107***	-0.074*
	(0.030)	(0.038)
Instructor, Year, Time of day FE	X	X
Student Characteristics	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track		
Women	3.20**	-0.50
Men	1.70	-1.50
Observations	12,230	12,230

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** p<0.01, ** p<0.05, * p<0.1. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Chapter 2

Juvenile Crime and the Four Day School Week

2.1 Introduction

The Office of Juvenile Justice and Delinquency Prevention reports that the majority of juvenile crimes are committed during non-school hours; they peak between 3 p.m. and 6 p.m. (Snyder and Sickmund, 2006). This fact has led to the common perception that lengthening the time students are in school, or increasing funds to expand youth programs, will reduce youth crime. The idea is that students who might engage in criminal activity are deprived of these opportunities when they are supervised. However, the extent to which school or youth program participation changes youth crime, if at all, remains an open question. Establishing a causal link between school attendance and youth crime is a challenge because it requires variation in school schedules. Existing studies rely on sporadically occurring disruptions to schedules which reduce the number of days that students are in school in a given week, e.g. teacher inservice days (Jacob and Lefgren, 2003), and randomly occurring events such as strikes (Lu-allen, 2006), or furloughs (Akee et al., 2014). All of these studies focus on a temporary response to an unexpected school schedule change.[1]

We offer an alternative approach, one that exploits the adoption of the four-day school week policy across rural counties and years within the state of Colorado, one of the states where the four-day week is most common. The main contribution of this study is to identify changes to youth criminal behavior associated with a more intentional and prolonged schedule change. On one hand, parents, schools, and communities can more effectively prepare for this more permanent schedule change. On the other hand, the effects associated with this type of policy may be larger than those found in previous work because this type of schedule change is regular and long-lasting. As such, gaining a better understanding of the impact of more permanent schedule changes on criminal behavior has important policy implications since many school districts throughout the U.S. have started to experiment with alternative schedules (i.e.year-round school or the four-day weeks, etc.) in an attempt to cut costs and/or boost student performance.

Using data on reported crimes by day-of-the-week and aggregating to the county-year level, we show that on average property crimes in rural Colorado counties increase as a result of the policy. In particular, larceny crimes increase substantially when four-day school weeks are adopted. Alternatively, we find no statistically significant evidence that juvenile drug or violent crime rates change. Our results are consistent with a

[1] While inservice days (Jacob and Lefgren, 2003) are planned ahead of time by school officials, they appear sporadically throughout the year and often vary widely from location to locations.

2011 report from the U.S. Department of Justice that shows larceny is the most common juvenile crime.[2]

The paper proceeds as follows. In section 2.2 we provide a conceptual framework as well as review the current literature on the effects of school attendance on crime and the four-day school week policy. section 2.3 and section 2.4 describe the data and empirical framework respectively. The baseline results are presented in section 2.5 and section 2.6 shows evidence that the findings persist across a variety of robustness checks. section 2.7 concludes.

2.2 Juvenile Crime and Four-Day School Week Policies

2.2.1 How might the four-day school week impact youth crime?

It is unclear whether the four-day week schedule promotes or hinders youth criminal activity. Different types of crime may be affected very differently. There are several channels through which students' school attendance may affect youth criminal behavior. *Incapacitation* refers to the notion that juveniles who are kept busy and who are under adult supervision will remain out of trouble during these hours. Stated somewhat differently, students who have less idle time are less likely to engage in illegal activities. School, after-school programs, and other youth programs are forms of juvenile incapacitation. The incapacitation component of school is therefore expected to reduce juvenile property and violent crime rates during hours in which school is in session. It

[2]Property crime is defined as the unlawfully taking of property from the possession of another without the use of force, threat or fraud and comprises several types of theft including larceny, burglary, arson and motor vehicle theft. Larceny is the most common type of property crime committed by juveniles and includes shoplifting, pick-pocketing, bicycle theft, theft from a vehicle including vehicle parts, or theft from a building or structure where no break in was involved.

is also possible that incapacitation has positive indirect effects. For example, students may be exposed to a new activity or made aware of healthy choices while incapacitated and choose to spend time on this outside of school hours rather than getting into trouble.

An alternative channel – one that receives less attention in policy circles and the popular press – is that school or other youth programs increase the *concentration* of youth which may increase crime. First, the concentration of students at school provides increased interactions and a space for conflict to arise in the form of fights or other violent exchanges. Although this mechanism suggests that violent crimes could be elevated during school hours, these crimes may also increase outside of school as a result of concentration if the involved parties plan to settle their differences at a later time. Second, school may provide a low cost way to coordinate crimes such as drug transactions or to plan property crimes that may be executed outside of school. A concentration story predicts that school attendance increases all types of crime during or after hours.

In the context of this paper, an incapacitation story suggests that juvenile crime will likely increase in counties that have adopted a four-day school week schedule. The direct effect of incapacitation suggests that that crime will increase on the days of the week that students are not attending school (Monday or Friday) due to lack of adult supervision. Not only are students not at school, but many parents are likely at work and therefore many teenaged students are likely unsupervised. An indirect effect associated with incapacitation will lead to an increase in crime during non-school hours. A concentration story, on the other hand, implies that crime will decline in treated areas specifically on the weekday that students are off from school, but potentially also during non-school hours, because there are fewer opportunities to interact and coordinate with other students.

2.2.2 Related Literature

There is a substantial body of empirical research examining incapacitation. The general finding is that incapacitation decreases criminal activity and other risky behavior such as teen pregnancy. However, the primary focus of this literature has been on variation in the number of years spent in school based on compulsory schooling laws (Anderson, 2012; Anderson et al., 2013; Berthelon and Kruger, 2011; Black et al., 2008; Lochner and Moretti, 2004). This means that the effects are often temporally distinct from the actual program occupying the individuals, so these studies provide little insight into the effect of a program to keep students out of trouble in the immediate future. Additionally, these policies don't allow for the estimation of concentration effects.[3]

Research that addresses the distinction between incapacitation and concentration effects typically relies on exogenous variation in the amount of time that a student spends in school. Jacob and Lefgren (2003) rely on teacher in-service days to estimate a causal relationship between school attendance and crime in urban settings. They find that juvenile property crime declines by 14% on days when school is in session but violent crime for this same group increases by 28% percent on school days. A reduction in property crime associated with increased school attendance is consistent with inca-

[3] A related body of literature relies on experimental interventions in after-school programs to determine the impact of school attendance on youth criminal activity. Insight from these studies is limited due to selective participation; programs of this nature are typically not mandatory and those most at risk may avoid them (see Cross et al. (2009); Rodríguez-Planas (2012)).

pacitation since the reduction was detected on days (and hours) when students were in school. The spike in violent crime on these days provides evidence for a concentration story.

Luallen (2006) exploits school attendance variation caused by teacher strikes – which resulted in canceled school days – to estimate an incapacitation effect. He finds that juvenile property crime and violent crime increase on days with strikes, but that the results are solely driven by urban areas. Akee et al. (2014) estimate the school-crime relationship based on public school teacher furlough days in Hawaii and find that time off from school is associated with significantly fewer juvenile crimes which supports a concentration story. In contrast to these studies, four-day school weeks are primarily located in rural areas and adopted over a long time frame.

2.2.3 Four-Day School Week Policies

As of 2008, seventeen states had switched a portion of their school schedules from a five-day week to a four-day week.[4] The primary motivation for states to implement this policy is to reduce transportation costs, which are especially salient for rural schools. The four-day school week became particularly popular during the energy crisis in the 1970s, at which time many states began changing laws regarding days spent in school. During this period the Colorado legislature changed their law from a mandatory number of school days to a mandatory number of hours, enabling districts in the state to adopt a four-day school week. The number of hours a student spends at school per year

[4]Starting with South Dakota in the 1930s, the following states have schools on the four-day week: Arizona, California, Colorado, Idaho, Kansas, Kentucky, Louisiana, Michigan, Minnesota, Montana, New Mexico, Oregon, South Dakota, Texas, Utah, Wisconsin and Wyoming. However, many of these programs are very limited. See http://www.ncsl.org/research/education/school-calendar-four-day-school-week-overview.aspx for background on specific state legislation regarding four-day schools.

remains constant and is typically set by state statute. To compensate for one fewer day of instruction, those on the four-day week schedule attend school for more hours per day and/or more days in the year. Over the following decades there was a slow shift towards the four-day week schedule in rural districts. As of 2009, 20% of students in Colorado attend four-day week schools. Of the schools that have switched in Colorado, roughly 80% are on a Monday through Thursday schedule with Fridays off with the remainder on a Tuesday-Friday schedule with Mondays off.

Given that cost considerations are central to the decision to switch, research on four-day school weeks has primarily focused on financial savings. Grau and Shaughnessy (1987), using data from ten school districts in New Mexico, document that districts operating on a four-day week experience a 10%-25% savings on fuel, electricity and transportation costs. Griffith (2011) examines six school districts that are either on the four-day week or in transition to that schedule and finds that the policy yields a maximum of about 5.5% savings.[5] Despite their growing prevalence, little work has been done to understand the impact of four-day school weeks on students. To the best of our knowledge, the only study at this point which evaluates the impact of four-day school weeks is Anderson and Walker (2014). Their analysis focuses on the state of Colorado and they find a modest, but statistically significant, positive relationship between the policy and elementary school students' math and reading test scores. Their findings suggest that switching to a four-day week does not compromise student achievement, and may even improve it.

[5]Four day school weeks have been of interest in popular media as well and journalists have gone to some effort to examine specific cases of the policy change. A TIME Magazine article (Kingsbury, 2008) reports that some rural school districts experienced large savings on transportation, utility, and insurance costs as a result of the policy and a Wall Street Journal article (Herring, 2010) sheds light on the savings that the policy has brought to a rural district in Georgia.

2.3 Data

We combine several data sources for our analysis which includes 47 rural counties in Colorado for the years 1993-2009. Because four day schools are primarily undertaken in rural areas and there exists almost no variation in school schedules across urban areas, we restrict our analysis to rural counties only.[6] For our purposes, a county is defined as rural if it is not part of a metropolitan statistical area (MSA).[7] Data indicating which schools are on a four-day school week and the timing of when they switch from a five-day schedule to a four-day schedule come from the Colorado Department of Education.[8] We link these schools with counties since crime data is aggregated to the county level.[9] Nearly one-third of all rural counties have at least one school that is on a four-day school week and approximately 20% of students attend a school on this schedule.

We use the Common Core of Data (CCD) from the National Center for Education Statistics, which contains the universe of Colorado schools, to obtain total student population by county and several measures of student body composition including the percent white, the percent on free or reduced lunch and the student/teacher ratio. The CCD and the list of four-day schools are combined to obtain the treatment variable:

[6] Less than 5% of the total student body in urban areas are affected by the policy compared to 20% in rural areas. Including urban counties in the analysis will introduce noise, which is generated by smaller changes in student population, into the treatment variable. We want our estimates to reflect changes due to policy adoption. Table 7 reports the baseline results for all of Colorado. While there are some minor differences, the results are largely the same.

[7] The MSAs omitted are Denver, Boulder, Greeley, Colorado Springs, Fort Collins, Pueblo, and Grand Junction. Other possible definitions of rural, such as population counts or only omitting the Denver MSA, have been considered and have little influence on the estimated results.

[8] We thank Mark Anderson of Montana State University for helping us obtain this data.

[9] While there is crime data available at more granular jurisdictional level than counties, these jurisdictions do not overlap in any consistent way with school boundaries.

the percentage of students in grades 6 through 12 on a four-day school week.[10] We use these grades because juveniles age 11-17 account for over 99% of juvenile crimes reported in Colorado.

We use two crime datasets. The first is county level arrests from the Colorado Department of Public Safety. This dataset contains reported arrests by crime type and by county in each year for the entire sample, 1993 through 2009. Summary statistics are reported in Table 2.1 (column 1). This dataset has the advantage of covering all arrests in Colorado. However, since it is aggregated annually the timing does not match up perfectly with the treatment variable which is reported by academic years. The second data source is the National Incident Based Reporting System (NIBRS) program. NIBRS provides detailed information on reported crimes at the incident level, including the date of the crime. This allows us to match the reported crimes with the treatment variable by academic year and avoid a timing problem. Reported crimes are a superset of arrests because they include people who are cited for a crime without actually being arrested, which is a common occurrence. The number of offenses is aggregated up to the county level to match the level of treatment and the other dataset.

In addition to providing detailed data on individual crimes, an advantage of the NIBRS data is that it flags the exact date of the offense as well as detailed demographic characteristics of the offenders; this allows us to precisely identify juveniles in the sample as well as examine changes in crime by day-of-the-week. Unfortunately, Colorado has only been fully participating in the NIBRS program since 1997, so the data are not available for the entire sample period. As such, all regressions using NIBRS data include the years 1997 through 2009. Additionally, while the data covers most of

[10] The results are similar if we use all students on a four-day school week regardless of age and percentage of schools on a 4 day week.

Colorado, there are some agencies (approximately 10% of agencies covering approximately 10% of the state's population) who do not participate in the program, reducing the number of rural counties in the NIBRS sample to 44.[11] Column 2 of Table 2.1 contains summary statistics for offenses from NIBRS. The outcomes are normalized by population and defined as crime (offenses or arrests) per 1,000 population in the county.[12]

The crime outcomes studied in this analysis include property crimes, violent crime and drug violations. Property crime is a general category consisting of larceny, breaking and entering, grand theft auto, and arson. Also included are violent crimes which consist of homicide, sexual assault, robbery, and assault. Drug crimes are a separate category. Some incidents reported in the NIBRS dataset involve more than one offense (i.e. breaking and entering while in possession of illegal drugs). In these cases we count the incident in both categories. While this results in some double counting of incidents, the alternatives are less palatable. Dropping all of these incidents results in a substantial loss of data. Some systems use a hierarchy such that an incident is categorized as its most "severe" offense type; however, these distinctions are often arbitrary and can result in under-counting of some kinds of crimes, especially drug offenses.

[11] While we are able to adjust the sample by dropping agencies that do not fully participate in NIBRS, any remaining underreporting should cause the coefficient to be biased downwards, as the number of reported crimes would be lower than actual number of crimes in the county.

[12] Due to the 10% of non-participating agencies in the NIBRS dataset we are not able to normalize the offense data by youth population, instead we normalize by the population provided by NIBRS which adjusts for the non-compliers. In order to make estimates comparable across datasets, we also normalize the arrest data by total population in a count-year. Note, we have also normalized arrest data by youth population and obtain similar results.

2.4 Empirical Framework

We estimate the following difference-in-difference model

$$Y_{ct} = \beta_0 + \beta_1 T_{ct} + X'_{ct}\alpha + \gamma_c + \delta_t + \varepsilon_{ct} \tag{2.1}$$

where variable Y_{ct} is the number of crimes (arrests or offenses depending on the dataset) in county c and in year t.[13] T_{ct} is the percent of students in a county in a given year that are on a four-day school week schedule, X_{ct} is a vector of county-year level covariates (unemployment, percent of students in county eligible for free lunch, race, and student/teacher ratio), γ_c is a county level fixed effect, δ_t is a year fixed effect, and ε_{ct} is the usual error term. The treatment variable T_{ct} is constructed by dividing the total number of students in grades 6-12 in a county who are on the four-day week in a given year by the total number of students in grades 6-12 in a county-year. When using the Colorado DPS arrests data we drop the year that a school first adopts this policy since the crime data is reported by calendar year (as opposed to academic year) and the initial year is only partially treated. We exclude summer months because no students are treated during this time, and all regressions are weighted by the average county population.

We estimate the above model using two methods, OLS (Ordinary Least Squares) where the data generating process is assumed to be linear, and Fixed Effect Poisson QMLE (Quasi Maximum Likelihood Estimation); a non-linear specification. Although we report estimates from both models, due to the non-linear nature of the data, the

[13]The outcome is a count in the Poisson models and a rate (crimes per 1,000 residents) in the OLS specifications.

preferred estimates come from the FE Poisson.[14] Additionally, the FE Poisson corrects for overdispersion and excess zeros; which are both present in the data. In particular, clustered standard errors at the county level correct for the issue of overdispersion, and the fixed effects deal with the excess zeros (Wooldridge (1999)).

Anecdotally, there is little reason to believe that the adoption of the policy is in response to crime patterns in a given county or correlated with unobservable time varying county characteristics that also correlate with juvenile crime; two issues which would undermine the causal interpretation of the policy's effect. In Colorado, the schools that have adopted a four-day school week most often cite financial savings as the reason (Anderson and Walker, 2014; Donis-Keller and Silvernail, 2009; Grau and Shaughnessy, 1987). The Colorado Department of Education states that four-day schedules are almost entirely adopted by schools in rural districts that serve a dispersed group of students because they can save on transportation costs. Other reasons schools have decided to switch include parent support, improved attendance, and increased academic performance. We return to this issue in Section 6 and show that selection into the policy is not a major concern.

2.5 Results

Table 2 reports regression coefficient estimates obtained from the arrest data in the odd numbered columns and results obtained from the reported offense data (NIBRS) in the even numbered columns. Four crime outcomes (for both arrests and offenses) are included in each table: larceny, property crime, drug violations and violent crime.

[14]Due to the large number of zeros in the crime data (our outcome), we can't simply take a log transformation and run OLS; an approach often used when dealing with a non-linear outcome.

Because it is the convention in this literature to use a count model, Table 2a reports the point estimates for Equation 2.1 from a FE Poisson specification. The results in columns 1 and 2 show a significant increase in juvenile larceny arrests and offenses. Switching all students in a county to a four-day week leads to larceny crimes nearly doubling. Companion Table 2b reports OLS results, which are consistent with the Poisson results, and shows that switching all students in a county to a four-day school week leads to a 1.3 increase in juvenile arrests for larceny per 1,000 population or a 1.7 increase in reported juvenile offenses for larceny per 1,000 population. This effect is large, it is about a 50% increase from the mean or a little more than 1 standard deviation for larceny arrests and approximately 87% of a standard deviation for reported larceny offenses, although effects of this magnitude are not unusual in the literature.[15]

Columns 3 and 4 (of Table 2) report the estimated effect of the policy on all property crime. Since larceny is part of property crime, it is not surprising that the effects are similar to the larceny estimates but slightly smaller in magnitude. It is not implausible that students who have an additional day off during the week are more likely to engage in minor offenses such as shoplifting and other petty theft, while property crimes such as arson and breaking and entering are not affected. The aggregation of all of these crimes may account for the fact that property crime show a somewhat smaller positive coefficient. For both arrests and offenses the estimated effects of the percent of students on a four-day school week on drug violations and violent crime (columns 5-8) are statistically and economically insignificant.

[15] For example, Jacob and Lefgren (2003) find effects on property crimes that roughly range between 50% and 75% of a standard deviation.

2.5.1 Day-of-the-Week Results

Aggregating the NIBRS data to day-of-the-week, county, year level allows us to examine the effects of the four-day school week policy on specific days. Table 3 reports the corresponding summary statistics for this level of aggregation. Examining crime patterns at this level may help to disentangle the mechanism which underlies our main findings. Is crime increasing on the specific day that students on a four-day week have off? A crime spike on Fridays due to an increase in the number of students who are treated on Friday would be consistent with the direct effects of an incapacitation story. Alternatively, finding lower levels of crime on the day students have off would be consistent with concentration.

To investigate this, we run the following specification:

$$Y_{ctd} = \beta_0 + \beta_1 TF_{ct} + \beta_2 TM_{ct} + \sum_{d=1}^{7} \theta_d TF_{ct} * \psi_d + \sum_{d=1}^{7} \mu_d TM_{ct} * \psi_d + X_{ct}'\alpha + \gamma_c + \delta_t + \psi_d + \varepsilon_{ctd} \tag{2.2}$$

where ψ_d is a set of dummy variables indicating the day of the week (Wednesday is the omitted day). TF_{ct} and TM_{ct} are the percent of students in a county-year treated on Friday and Monday respectively. Both treatment variables are also interacted with each day of the week which are represented by the following vectors, $\sum_{d=1}^{7} \theta_d TF_{ct} * \psi_d$ and $\sum_{d=1}^{7} \mu_d TM_{ct} * \psi_d$. Other controls included are defined in the previous section and the results of the estimation are included in Table 4a (Table 4b reports the OLS results). The results show that although overall the policy increases larceny and property crimes, there is no evidence that it differentially increases on a particular day. That is, the policy does not cause crime to increase on Fridays (the day students have off) any more than any other day of the week. One possibility for this somewhat unintuitive result is that because we are analyzing rural counties, there may be too few observations in a given county/year/day cell to detect any differences across days. Another explanation is that students who have an "extra" day off per week not only increase their unlawful behavior on their day off (typically Friday), but this also spills-over to other days of the week. Either way, we find no evidence of a direct incapacitation story, although spill-overs from incapacitation may be an underlying mechanism. Clearly, a concentration story has been ruled out. If the concentration component of school attendance had dominated, then crime would have decreased on Fridays and/or overall, and we find the opposite – it increases. Similar to the county-year results in Table 2, we find no consistent changes in drug or violent crimes due to the policy, either overall or on any given day of the week. The varied findings across the different crime outcomes, however, highlights the importance of disaggregating crime by type for the analysis. Uncovering how the policy affects each type of crime, if at all, and through which channel helps provide a more comprehensive understanding of the impact of the four-day school week.

2.6 Robustness Checks

The results reported so far are robust to a variety of alternative specifications, including changing the treatment definition to percent of schools instead of students and varying the definition of rural.[16] Three additional robustness checks – a placebo test using adult crime, a county-year placebo test, and a check for the exogeneity of policy adoption, are examined in this section.

[16] We also restrict the sample to only summer months and find no effect.

2.6.1 Adult Crime and County Placebo Test

To ensure that the treatment variable is not picking up some underlying change, such as changes in the local economy or law enforcement practices, we run identical models using adult (age 25+) crime as the outcome.[17] If juvenile crime is only increasing in areas because of the treatment (the four-day week policy), then the relationship between the percent of juveniles on a four-day week in a given county-year and adult crimes should not be statistically significant. Table 5, which is formatted the same as Table 2, contains the results for adult arrests and adult offenses. Neither dataset (arrests or offenses) indicates a significant effect for adult crime, supporting the idea that a four-day school week policy is uniquely impacting juveniles and is not a proxy for other unobserved changes in the county.

Moreover, since we use many juvenile and adult crime outcomes in this analysis, there is the potential for a multiple testing problem. That is, increasing the number of outcomes used in the analysis increases the likelihood that some of the treatment coefficients will appear statistically significant even if a true relationship does not exist. To check for this concern, we test the joint significance of the estimated effect of the treatment obtained from the four juvenile crime outcomes: larceny, property, drug and violent. If our findings are an artifact of the multiple testing problem then we should reject that the four coefficients are jointly equal to zero. We find evidence against the null at the 2% level. On the other hand, we fail to reject that the estimated effect of the four-day week using the four adult crime outcomes are jointly statistically different from zero. Together these results alleviate the concern of a multiple testing problem.

[17] Individuals ages 18-25 are left out of these models. Some 18 year-olds are still in high school (and are therefore in the treated population) and those just out of high school are likely to socialize with treated individuals and may be indirectly affected by treatment.

As an additional sensitivity test we conduct a county level placebo analysis. Counties are randomly reassigned the percent on a four-day week (the treatment variable) of a different county. This reassignment maintains the characteristics of the treatment profile, the treatment turns on and can increase over time, but the timing of treatment has been randomly reassigned. If our main results are really capturing the effect of the policy on crime, the results for this specification should reveal no relationship between the treatment and crime outcome since the reassigned treatment does not reflect the true percent in that county-year who are treated. Figure 1 shows the distribution of the estimated coefficients of the treatment variable from estimating Equation 1 1,000 times, each time with random reassignment of the treatment profile. The actual estimate is indicated with a red dashed line. The median estimate of the effect and the associated statistics are both essentially 0, which strongly suggests our observed effect is not an artifact of the estimation strategy.

2.6.2 Leads of the Treatment Variables

A concern with our empirical approach which would threaten the identification of our estimates is if school districts are adopting this policy in response to the current crime rate in the county. For instance, if schools in counties with high crime rates adopt the four-day week policy as a means to reduce crime (and potentially are adopting other crime reducing policies concurrently), then the internal validity of our results will be jeopardized. We check for this type of adoption behavior by including leads of the treatment variable to Equation 2.1. We follow a similar approach to that of Gruber and Hanratty (1995), Friedberg (1998), and Bedard and Do (2005) and run the following model:

$$Y_{ct} = \beta_0 + \beta_1 T_{ct} + \beta_2 x_{ct} + \lambda_1 \triangle T_{ct+1} + \lambda_2 \triangle T_{ct+2} + X'_{ct}\alpha + \gamma_c + \delta_t + \varepsilon_{ct} \tag{2.3}$$

Note that the two lead terms $\triangle T_{ct+1}$ and $\triangle T_{ct+2}$ are the percent of students in a four-day school in a county between year t=0 and t=1 and then between t=1 and t=2, respectively. Thus, the estimated coefficients on these two lead terms represent the relationship between the change in four-day school week adoption between years t/t+1 and t+1/t+2 and crime in that county. If high crime rates in a county lead to the adoption of a four-day school week, we should observe school districts responding to high crime rates by adopting the four-day week in the following year and thus the estimates on the lead terms $\lambda_1$and λ_2 should be positive and significant. Table 6 contains the results for the estimates on the lead terms and shows that they are statistically insignificant and, even though they are occasionally large, vary widely in sign indicating that schools likely are not adopting this policy as a response to crime rates.[18] Since this specification relies on changes in policy adoption between years thus reducing the sample size, to provide context Table 6 also contains the baseline results for this restricted sample.

2.7 Conclusion

In this paper we show that the implementation of the four-day school week in rural areas leads to an increase in youth property crime, particularly larceny, while drug and violent crimes appear unaffected. Although our findings are not directly inline with the

[18]This test is not feasible in the arrests data because we drop the year of adoption due to it being partially treated.

standard incapacitation story often cited in the literature which shows spikes in crime at exactly the times when students are off from school, the estimates in this paper are consistent with spill-overs associated with incapacitation. That is, property crimes not only increases on days that they have off (Friday), but also on other days. Overall, these results are informative in that they highlight the fact that policymakers should consider the unintended consequences before implementing such a schedule.

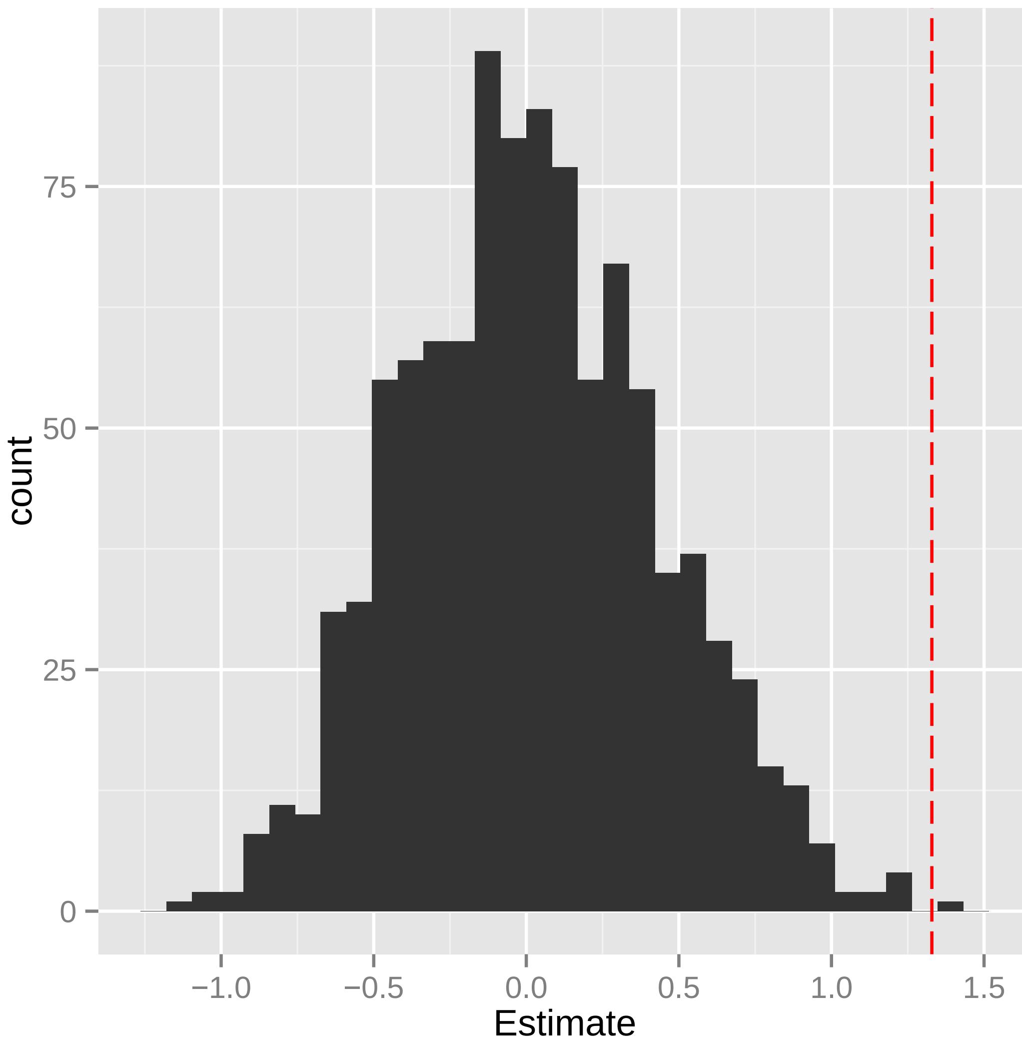

Figure 2.1: County Placebo Test

	Arrest Data	NIBRS Data: Offenses
Percent on 4-Day Week	0.23	0.21
	(0.37)	(0.35)
Percent Not Treated	0.48	0.48
	(0.50)	(0.50)
Percent Treated: 0.01 -49 %	0.31	0.34
	(0.46)	(0.47)
Percent Treated: 50 - 99%	0.05	0.04
	(0.23)	(0.19)
Percent Fully Treated	0.15	0.13
	(0.36)	(0.34)
Percent with Friday Off	0.21	0.18
	(0.36)	(0.33)
Percent with Monday Off	0.03	0.03
	(0.11)	(0.13)
Juv. Larceny Crime per 1,000	0.87	1.02
	(1.13)	(1.99)
Juv. Drug Crime per 1,000	0.30	0.59
	(0.41)	(1.05)
Juv. Property Crime per 1,000	1.17	2.13
	(1.35)	(3.24)
Juv. Violent Crime per 1,000	0.14	0.70
	(0.26)	(1.13)
Juv. Larceny Crime	17.53	17.992
	(27.80)	(26.527)
Juv. Drug Crime	5.58	6.867
	(8.85)	(9.730)
Juv. Property Crime	21.96	22.648
	(32.60)	(31.083)
Juv. Violent Crime	2.03	2.223
	(3.77)	(3.967)
Adult Larceny Crime per 1,000	1.85	8.42
	(1.72)	(12.04)
Adult Drug Crime per 1,000	1.81	1.34
	(1.83)	(1.48)
Adult Property Crime per 1,000	2.51	16.04
	(2.07)	(19.18)
Adult Violent Crime per 1,000	1.10	3.32
	(0.93)	(2.99)
Adult Larceny Crime	33.37	36.204
	(45.89)	(46.160)
Adult Drug Crime	30.26	36.358
	(42.85)	(46.310)
Adult Property Crime	42.76	46.290
	(55.60)	(55.870)
Adult Violent Crime	14.93	16.709
	(17.70)	(18.583)
Population	13,416	14,928
	12,544	13,500
Unemployment Rate	4.84	4.70
	(2.24)	(1.93)
Percent Free Lunch	0.28	0.29
	(0.15)	(0.15)
Percent White	0.75	0.74
	(0.18)	(0.18)
Student/Teacher Ratio	13.60	13.64
	(2.63)	(2.56)
Observations	721	526

Note: These are summary statistics including all years. The arrest data is from 1993-2009 and the NIBRS (offense) data is from 1996-2009. In general the offense data is a superset of the arrest data. However, the adult offense data is defined as ages 25+ and the adult arrest data includes those 18 and older.

Table 2.1: Summary Statistics: Rural Counties in Colorado

	Larceny		Property		Drug		Violent	
	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	0.963*	0.824***	0.846**	0.699***	-0.351	-0.741	0.702	0.309
	(0.519)	(0.311)	(0.332)	(0.228)	(0.611)	(0.491)	(0.758)	(0.867)
Umemployment Rate	-0.0706*	-0.0806**	-0.044	-0.051	0.009	-0.012	0.018	0.028
	(0.038)	(0.035)	(0.034)	(0.034)	(0.049)	(0.049)	(0.064)	(0.063)
Percent on Free Lunch	-0.810	0.158	-0.929	0.052	-2.170*	-2.242**	4.355**	4.803**
	(0.923)	(1.115)	(0.740)	(0.826)	(1.124)	(1.133)	(2.076)	(2.222)
Percent White	-1.442**	-1.491***	-1.596**	-1.805***	-0.619	-0.571	-0.561	-0.513
	(0.695)	(0.498)	(0.763)	(0.465)	(1.202)	(0.916)	(2.622)	(2.081)
Student/Teacher Ratio	0.0703*	0.0973**	0.038	0.054	-0.030	-0.074	0.060	0.001
	(0.037)	(0.049)	(0.034)	(0.042)	(0.056)	(0.054)	(0.071)	(0.085)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	679	476	721	515	637	478	700	511

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is a count of crimes per county-year. Even columns report results using the arrest data from the Colorado DPS , odd columns report results using the NIBRS data. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.2: Poisson Model Base Specification – The Effect of Four-Day Week on Juvenile Crime

	Larceny		Property		Drug		Violent	
	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	1.329***	1.735**	1.507***	2.157	-0.040	0.376	0.050	0.206
	(0.465)	(0.847)	(0.511)	(1.300)	(0.108)	(0.529)	(0.091)	(0.584)
Umemployment Rate	-0.032	0.12	-0.014	0.142	0.002	0.0147	0.002	0.0527
	(0.078)	(0.120)	(0.082)	(0.216)	(0.020)	(0.053)	(0.015)	(0.061)
Percent on Free Lunch	-4.822***	2.771	-5.258***	6.318	-0.653	0.425	0.749	3.748*
	(1.564)	(4.769)	(1.851)	(6.584)	(0.422)	(1.923)	(0.464)	(2.217)
Percent White	-2.914*	3.778	-3.975	3.569	-0.265	0.361	-0.131	-0.389
	(1.542)	(3.580)	(1.809)	(5.524)	(0.598)	(1.377)	(0.430)	(1.617)
Student/Teacher Ratio	-2.368	-12.1	0.407	-2.399	0.248	8.843	-2.007	14.41
	(10.232)	(23.310)	(11.968)	(38.410)	(3.048)	(8.864)	(1.842)	(12.870)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	721	526	721	526	721	526	721	526

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. Even columns report results using the arrest data from the Colorado DPS , odd columns report results using the NIBRS data. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.3: OLS Base Specification – The Effect of Four-Day Week on Juvenile Crime

	Larceny			Property			Drug			Violent		
	No Treat	Off Fri	Off Mon	No Treat	Off Fri	Off Mon	No Treat	Off Fri	Off Mon	No Treat	Off Fri	Off Mon
Overall Crime	0.19	0.14	0.14	0.40	0.29	0.32	0.12	0.08	0.06	0.12	0.11	0.10
	(0.519)	(0.299)	(0.333)	(0.898)	(0.551)	(0.542)	(0.372)	(0.247)	(0.128)	(0.320)	(0.220)	(0.203)
Monday	0.22	0.12	0.11	0.45	0.24	0.26	0.09	0.07	0.05	0.11	0.10	0.11
	(0.689)	(0.219)	(0.242)	(0.998)	(0.414)	(0.457)	(0.221)	(0.158)	(0.109)	(0.311)	(0.143)	(0.178)
Tuesday	0.21	0.16	0.15	0.37	0.30	0.36	0.14	0.09	0.06	0.15	0.14	0.095
	(0.819)	(0.341)	(0.354)	(0.941)	(0.545)	(0.651)	(0.475)	(0.164)	(0.117)	(0.347)	(0.263)	(0.179)
Wednesday	0.20	0.14	0.16	0.40	0.27	0.29	0.12	0.11	0.08	0.12	0.11	0.10
	(0.431)	(0.278)	(0.394)	(0.793)	(0.442)	(0.529)	(0.361)	(0.343)	(0.164)	(0.278)	(0.252)	(0.153)
Thursday	0.18	0.11	0.14	0.33	0.27	0.29	0.11	0.08	0.07	0.13	0.13	0.10
	(0.409)	(0.194)	(0.291)	(0.568)	(0.449)	(0.495)	(0.312)	(0.177)	(0.143)	(0.407)	(0.235)	(0.214)
Friday	0.22	0.18	0.19	0.52	0.34	0.46	0.17	0.07	0.05	0.18	0.13	0.09
	(0.420)	(0.355)	(0.440)	(1.398)	(0.711)	(0.639)	(0.563)	(0.126)	(0.093)	(0.412)	(0.265)	(0.162)
Saturday	0.17	0.16	0.11	0.39	0.33	0.29	0.10	0.11	0.05	0.08	0.10	0.09
	(0.340)	(0.405)	(0.299)	(0.672)	(0.671)	(0.483)	(0.293)	(0.387)	(0.168)	(0.222)	(0.187)	(0.216)
Sunday	0.13	0.09	0.11	0.32	0.27	0.26	0.06	0.06	0.02	0.07	0.09	0.10
	(0.320)	(0.229)	(0.277)	(0.618)	(0.552)	(0.496)	(0.243)	(0.241)	(0.052)	(0.185)	(0.160)	(0.292)
Observations	1,688	1,240	586	1,688	1,240	586	1,688	1,240	586	1,688	1,240	586

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.4: Summary Stats by Day-of-the-Week (NIBRS Offense Data)

	Larceny	Property	Drug	Violent
	(1)	(2)	(3)	(4)
Percent Treated Friday	0.948***	0.895***	-0.698	0.610
	(0.289)	(0.212)	(0.538)	(0.822)
Percent Treated Monday	0.180	-0.516	-1.344	-0.132
	(1.791)	(0.784)	(1.289)	(1.644)
Percent Treated Fri X Fri	-0.0508**	-0.039	0.034	-0.139
	(0.025)	(0.040)	(0.027)	(0.157)
Percent Treated Fri X Sat	-0.039	-0.008	0.0714***	-0.016
	(0.059)	(0.053)	(0.025)	(0.056)
Percent Treated Fri X Sun	-0.016	0.000	0.021	-0.079
	(0.014)	(0.014)	(0.026)	(0.072)
Percent Treated Fri X Mon	-0.044	-0.056	0.0483*	-0.111
	(0.054)	(0.050)	(0.027)	(0.113)
Percent Treated Fri X Tues	-0.050	-0.009	0.0633***	-0.001
	(0.058)	(0.053)	(0.019)	(0.052)
Percent Treated Fri X Thurs	-0.021	0.023	0.0704*	-0.034
	(0.059)	(0.055)	(0.037)	(0.084)
Percent Treated Mon X Fri	-0.045	0.021	0.107	0.024
	(0.121)	(0.099)	(0.113)	(0.156)
Percent Treated Mon X Sat	-0.112	-0.146	-0.117*	-0.201
	(0.138)	(0.105)	(0.063)	(0.141)
Percent Treated Mon X Sun	0.062	0.052	0.102	-0.102
	(0.191)	(0.144)	(0.091)	(0.201)
Percent Treated Mon X Mon	-0.028	-0.052	-0.065	0.039
	(0.143)	(0.108)	(0.050)	(0.111)
Percent Treated Mon X Tues	0.083	0.133	0.090	-0.190
	(0.157)	(0.095)	(0.116)	(0.252)
Percent Treated Mon X Thurs	0.124	0.131	-0.020	-0.102
	(0.153)	(0.107)	(0.056)	(0.194)
Monday	0.005	0.00389*	-0.003	-0.002
	(0.003)	(0.002)	(0.003)	(0.007)
Tuesday	0.004	0.001	-0.004	-0.003
	(0.003)	(0.002)	(0.003)	(0.006)
Tursday	-0.002	-0.004	-0.003	-0.011
	(0.003)	(0.003)	(0.002)	(0.014)
Friday	0.011	0.006	-0.006	0.004
	(0.008)	(0.006)	(0.007)	(0.011)
Saturday	0.005	0.003	0.002	0.004
	(0.004)	(0.004)	(0.004)	(0.005)
Sunday	0.003	0.003	-0.002	0.002
	(0.003)	(0.002)	(0.003)	(0.006)
Time Varying County Charact.	X	X	X	X
County & Year Fixed Effects	X	X	X	X
Observations	3,216	3,414	3,222	3,369

Notes: The estimation sample is restricted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.5: Poisson Model Day-of-the-Week Analysis (NIBRS Data)

	Larceny	Property	Drug	Violent
	(1)	(2)	(3)	(4)
Percent Treated Friday	0.254*	0.317	0.104	0.018
	(0.148)	(0.189)	(0.085)	(0.101)
Percent Treated Monday	0.764*	0.990	0.098	0.454*
	(0.409)	(0.622)	(0.244)	(0.242)
Percent Treated Fri X Fri	0.046	-0.032	-0.145**	-0.080
	(0.054)	(0.111)	(0.062)	(0.054)
Percent Treated Fri X Sat	0.052	0.047	-0.020	0.054
	(0.049)	(0.101)	(0.070)	(0.034)
Percent Treated Fri X Sun	0.132**	0.185*	0.060	0.050
	(0.057)	(0.094)	(0.053)	(0.035)
Percent Treated Fri X Mon	0.006	-0.083	-0.007	-0.015
	(0.053)	(0.113)	(0.051)	(0.061)
Percent Treated Fri X Tues	-0.002	0.009	-0.081	-0.005
	(0.045)	(0.083)	(0.086)	(0.043)
Percent Treated Fri X Thurs	0.042	0.084	-0.059	0.006
	(0.043)	(0.093)	(0.065)	(0.075)
Percent Treated Mon X Fri	0.059	0.663**	-0.389**	-0.465**
	(0.192)	(0.314)	(0.186)	(0.213)
Percent Treated Mon X Sat	-0.255	-0.208	-0.299	-0.085
	(0.251)	(0.263)	(0.289)	(0.173)
Percent Treated Mon X Sun	0.189	0.199	-0.070	-0.022
	(0.242)	(0.361)	(0.237)	(0.210)
Percent Treated Mon X Mon	-0.309	-0.384	-0.028	-0.213
	(0.208)	(0.384)	(0.153)	(0.209)
Percent Treated Mon X Tues	-0.206	0.778	-0.251	-0.380*
	(0.252)	(0.614)	(0.203)	(0.192)
Percent Treated Mon X Thurs	0.169	0.327	-0.169	-0.085
	(0.276)	(0.260)	(0.251)	(0.263)
Monday	-0.004	0.047	-0.039**	0.018
	(0.024)	(0.047)	(0.018)	(0.032)
Tuesday	0.025	0.019	0.024	0.053**
	(0.028)	(0.040)	(0.035)	(0.025)
Tursday	-0.049**	-0.044	0.014	0.031
	(0.019)	(0.037)	(0.015)	(0.042)
Friday	-0.008	0.045	0.043**	0.07**
	(0.023)	(0.044)	(0.019)	(0.032)
Saturday	-0.012	0.029	0.000	-0.030
	(0.027)	(0.036)	(0.039)	(0.019)
Sunday	-0.108***	-0.098**	-0.075**	-0.039**
	(0.029)	(0.043)	(0.034)	(0.016)
Time Varying County Charact.	X	X	X	X
County & Year Fixed Effects	X	X	X	X
Observations	3,368	3,368	3,368	3,368

Notes: The estimation sample is restricted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.6: OLS Day-of-the-Week Analysis (NIBRS Data)

	Larceny		Property		Drug		Violent	
	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	0.066	0.160	0.028	0.128	-0.352	-0.121	-0.043	0.357
	(0.201)	(0.389)	(0.226)	(0.382)	(0.525)	(0.446)	(0.455)	(0.358)
Umemployment Rate	-0.025	-0.101**	-0.036	-0.102**	-0.019	-0.108**	-0.038	-0.027
	(0.036)	(0.045)	(0.037)	(0.041)	(0.046)	(0.049)	(0.053)	(0.052)
Percent on Free Lunch	-0.881	-2.572	-0.936	-2.564*	-0.579	-1.380	0.360	-2.157**
	(1.124)	(1.568)	(1.113)	(1.473)	(1.132)	(1.334)	(1.083)	(0.994)
Percent White	1.255	0.908	0.731	0.565	1.401	1.232	-0.376	-1.433
	(0.783)	(1.601)	(0.808)	(1.592)	(1.041)	(1.145)	(1.534)	(1.625)
Student/Teacher Ratio	0.002	-0.068	0.002	-0.0828*	0.016	-0.049	0.047	-0.006
	(0.034)	(0.049)	(0.031)	(0.045)	(0.030)	(0.045)	(0.034)	(0.044)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	721	525	721	525	721	522	721	525

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. Even columns report results using the arrest data from the Colorado DPS , odd columns report results using the NIBRS data. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.7: Poisson Base Specification for Adult Crime – Robustness Check

	Larceny		Property		Drug		Violent	
	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	0.266	6.403	0.691	5.619	-0.854	-0.0745	0.551	-0.667
	(0.466)	(4.072)	(0.300)	(7.064)	(0.714)	(0.425)	(0.010)	(1.242)
Umemployment Rate	-0.068	-0.79	-0.109	-0.533	-0.060	-0.11	-0.051	-0.188
	(0.100)	(0.832)	(0.129)	(1.250)	(0.090)	(0.093)	(0.068)	(0.215)
Percent on Free Lunch	-2.356	44.65	-3.195	58.9	0.094	2.645	0.314	-2.614
	(2.712)	(39.930)	(3.687)	(52.510)	(2.298)	(2.542)	(1.319)	(8.666)
Percent White	4.218	80.63	3.108	103	2.961	5.124	-0.345	1.47
	(2.850)	(53.080)	(3.642)	(69.220)	(2.767)	(3.698)	(1.889)	(11.750)
Student/Teacher Ratio	-7.032	-91.02	-12.861	-37.91	-9.711	22.19	-14.037	43.97
	(13.184)	(150.100)	(16.556)	(223.400)	(12.061)	(19.960)	(8.472)	(35.980)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	721	526	721	526	721	526	721	526

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. Even columns report results using the arrest data from the Colorado DPS , odd columns report results using the NIBRS data. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.8: OLS Base Specification for Adult Crime – Robustness Check

	Larceny		Property		Drug		Violent	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	1.731***	0.392	1.004**	0.576	-1.306*	0.072	0.037	0.394
	(0.484)	(0.628)	(0.450)	(0.539)	(0.753)	(0.565)	(0.888)	(0.666)
1-Year Lead in Adoption	--	0.399	--	0.362	--	(0.109)	--	0.897**
		(0.800)		(0.662)		(0.915)		(0.455)
2-Year Lead in Adoption	--	0.293	--	0.289	--	0.267	--	0.400
		(0.348)		(0.304)		(0.480)		(0.318)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	399	448	439	448	408	446	430	441

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.9: Poisson Model Leads of Treatments – Robustness Check (NIBRS Data)

	Larceny		Property		Drug		Violent	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	2.813**	2.916**	3.836*	3.927*	-0.020	-0.155	0.240	0.257
	(1.346)	(1.379)	(1.996)	(2.085)	(0.647)	(0.714)	(0.773)	(0.870)
1-Year Lead in Adoption	--	0.490	--	-0.541	--	-0.974	--	-0.225
		(1.171)		(2.529)		(0.685)		(1.157)
2-Year Lead in Adoption	--	0.518	--	1.245	--	-0.409	--	0.340
		(0.592)		(1.074)		(0.251)		(0.459)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	450	450	450	450	450	450	450	450

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. All regressions are weighted by county population.

Table 2.10: OLS Leads of Treatments – Robustness Check (NIBRS Data)

	Larceny		Property		Drug		Violent	
	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	1.652*	0.629*	1.349**	0.406	1.868	-0.00031	2.208**	0.463
	-0.945	-0.351	-0.643	-0.288	-1.364	-0.727	-1.036	-0.834
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	922	658	964	697	880	660	943	693

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. Even columns report results using the arrest data from the Colorado DPS , odd columns report results using the NIBRS data. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. Panel A regressions are weighted by county population.

Table 2.11: Poisson Model Base Specification All Counties – Robustness Check

	Larceny		Property		Drug		Violent	
	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS	Arrests	NIBRS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent on 4-Day Week	2.256***	0.862	2.675***	0.509	0.423	0.226	0.326	0.00654
	(0.727)	(0.672)	(0.834)	(1.374)	(0.406)	(0.392)	(0.204)	(0.567)
County Controls	X	X	X	X	X	X	X	X
County & Year FEs	X	X	X	X	X	X	X	X
Observations	964	707	964	707	964	707	964	707

Notes: The estimation sample is restircted to only rural areas (those counties outside of an MSA). The outcome is measured as offenses per 1,000 population and is observed by academic year. Even columns report results using the arrest data from the Colorado DPS , odd columns report results using the NIBRS data. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are clustered at the county level in parentheses. Panel A regressions are weighted by county population.

Table 2.12: OLS Base Specification All Counties – Robustness Check

Chapter 3

The Academic Achievement of American Indians

3.1 Introduction

Disparities in educational achievement for various groups in the United States have been a concern of policy makers and economists for years. In a number of studies, significant portions of observed racial wage differentials can be explained by controlling for earlier academic achievement, raising concerns about long-term effects of relative school performance. [1] While the black-white achievement gap has been extensively studied, [2] the relative achievement of American Indians has received far less attention,

[1] For example, see OGorman (2010); O'Neill (1990); Johnson et al. (1998).

[2] For literature on the effect of socioeconomic conditions and family background on the black-white achievement gap, see Jencks and Phillips (1998); Fryer and Levitt (2004); Armor (1992); Brooks-Gunn et al. (1997); Mayer (1997).

especially by economists. [3] Part of this is due to the small number of American Indian students in most datasets. American Indians make up only about 1 percent of the student population in the United States. However, American Indians are one of the most disadvantaged groups in the United States, with the lowest employment rates and highest poverty rates of any racial or ethnic group (Census (2012)). Although it is widely known that academic achievement is lower for American Indians than white students, there is little work quantifying the impact of demographic and school characteristics or comparing their performance to other disadvantaged groups. To our knowledge, Clotfelter et al. (2009) is the only other economics study that reports on the American Indian-white test score gap, based on data from North Carolina.

The first contribution of this paper is to document, in detail, the current test scores of American Indian students relative to other groups, by describing mean gaps, the overall distribution of test scores, and the changes as students age. In recent years, the National Center for Education Statistics (NCES) has conducted the National Indian Education Survey (NIES) in connection with the National Assessment for Education Progress (NAEP), allowing for more robust analysis of this group than in the past. We use the restricted use data from this survey, together with data on school attributes and locations to document how observed individual, school, and geographic characteristics affect measures of the test score gap.

A second contribution of this paper is to explore how racial and cultural identification influences measures of achievement gaps. Most achievement gap research has taken a student's race to be a fixed category. This study is unique in estimating the gap

[3] We use the designation "American Indian" rather than "Native American" as it is the term used by NCES for racial and ethnic classification on the student and school records. "Native American" is also confusing as it can mean any student born in the United States, as opposed to an immigrant.

by using both the definition of the school and the definition of the student. The NAEP surveys used contain two measures of racial and ethnic groups: those reported by the school records and those reported by the student on the survey. As a result, gaps can be estimated for students who are only identified in a racial or ethnic minority group by themselves, only by their school, or by both.

This distinction is valuable because one proposed explanation for test score gaps relates to the interplay between academic achievement and social group identification. For example, the anthropologist John Ogbu developed the theory of *oppositional identity* among minority groups who found themselves involuntarily under the jurisdiction of the United States as opposed to voluntary migrants. He argues that poor academic performance is related to cultural opposition to the structural barriers and discrimination experienced by these groups: students that perceive school culture to be aligned with "white" ideals are theorized to be more likely to exhibit negative attitudes toward academic success.[4]

In a similar vein, Akerlof and Kranton (2002) theorize that if a student's social category prescriptions differ from the school's ideals, then a student suffers identity loss for diverging from her social category's ideal school effort and as a result may underinvest in human capital. Austen-Smith and Fryer (2005) present a two audience signaling model that incorporates the ideas of peer dynamics with returns to schooling. The model predicts that middle ability types will pool on low investment: some individuals will underinvest in education when the cost of forgone social acceptance is greater than the forgone high wage.

[4] See, for example, Fordham and Ogbu (1986).

A number of researchers have empirically examined the relationship between identity and academic and economic outcomes. Fordham and Ogbu (1986) find evidence in favor of the cultural opposition hypothesis, although Cook and Ludwig (1997) and Ainsworth-Darnell and Downey (1998) do not. Benjamin and Strickland (2010) use experimental data and find evidence that social identity affects fundamental economic preferences, specifically time and risk preference. Others have documented the relationship between student popularity and grades across racial and ethnic groups and find that some groups penalize individuals socially for *Acting White*. [5] Recent evidence also suggests that identification with school may be the mechanism behind some academic gains associated with policies like class size reductions (Fletcher (2009)).

The present study explicitly documents racial gaps in connection with a student's selection of race. American Indian students on average have more variation in this choice, making this population well suited for addressing questions related to identity. As described in detail below, the correspondence between a student's own classification and the school's classification of a student is lower for American Indians than for other groups. This lower correspondence is perhaps not surprising, as American Indian students may have more latitude for choosing from multiple groups with which to identify. American Indians have the highest rate of intermarriage of all race and ethnic groups in the United States (Lofquist et al. (2012)): only one half of American Indians marry another American Indian, compared to 95 percent of whites and 85 percent of blacks who marry within their own race.[6] Hispanic students may also have ancestry in indigenous groups in Latin America, which also provides multiple options for classification. Fur-

[5]For examples, see Fordham (1996); Fryer (2002); Steele and Aronson (1998); Austen-Smith and Fryer (2005); Fryer (2010); Fryer and Torelli. (2010).

[6]Previous literature documents the effect of intermarriage on the ethnic identification of children (Duncan and Trejo (2007)).

thermore, tribal policies allow an individual to be an enrolled tribal member based on ancestry which ranges from 1/32 (among the Kaw Nation) documented heritage to 5/8 (among the Ute), which may promote awareness of American Indian ancestry among individuals with many ancestral roots. This context provides more scope for examining the role of self-identification.

Furthermore, tension between group identification and investments in formal schooling may be particularly salient for American Indians given their historical experience. According to BIE (2012), in the early twentieth century it was "federal policy to acculturate and assimilate Indian people by eradicating their tribal cultures through a boarding school system." This history is described in more detail in the next section, and it suggests that cultural identity and formal education were perceived to be in conflict over a long period.

Section 3 describes the data. Section 4 presents mean differences in achievement by school reported race for 4th and 8th graders in math and reading and documents the impact of controlling for demographic and school characteristics. School reported race is used in this section to make the results most comparable with previous studies of test score gaps. Section 5 reports the quantile regression results, finding that the distribution of scores for American Indian students is quite different than for other groups. Section 6 examines gaps related to reservation location and gaps based on self- and school-reported race. Section 7 concludes.

3.2 Historical Background

In 1819, the Indian Civilization Act initiated federal support for educating native students. Initially, federal funds were provided mainly to missionary groups. Funding was also connected to other Indian-specific policies: after the 1830 Indian Removal Act forced the relocation of Cherokee and other tribes to what is now Oklahoma, federal funds for missionary schools east of the Mississippi (where the Cherokee had previously lived) were eliminated and generous subsidies were provided for western schools. In the 1870s, the government began to provide public schools through the Office of Indian Affairs, although missionary schools continued to receive funding through government contracts.[7]

Most of the early federal schools were boarding schools. The first boarding school for American Indian students, located in Carlisle, Pennsylvania, housed students from throughout the United States. The dominant view was that boarding schools were necessary to impede cultural transmission. The 1879 Annual Report of the Commissioner of Indian Affairs stated "The progress of the pupils in industrial boarding schools is far greater than day schools. The children being removed from the idle and corrupting habits of savage homes are more easily led to adopt the customs of civilized life and inspired with a desire to learn" (Quoted in Reyhner and Eder (2006), pg. 73). At this point, federal funds were restricted to schools that only provided instruction in English (a number of religious schools that provided some instruction in native languages in early grades no longer received funding). There are many accounts of students being punished for speaking native languages in any context, even in private conversations

[7]Much of this section draws on the histories provided in *American Indian Education: A History* (Reyhner and Eder (2006)) and *Education and the American Indian* Szasz (1999).

with each other; students were often given Anglo names, had to cut their hair, and were forbidden to engage in religious or cultural practices. Tribal groups were also often intermingled to impede native language use. (For example, the Carlisle School had 53 tribes represented in 1894.) The conflict between formal school and native languages and cultures was particularly acute as attendance was frequently involuntary. In 1891, Congress authorized compulsory education for Indians, with the Commissioner given the right to reduce rations and annuities to families whose children were not enrolled. Some of the most egregious examples of compulsion were the use of federal troops to enforce attendance among the Hopi and Navajo. Dana Coolidge's testimony in 1932 at the Senate Subcommittee of the Committee on Indian Affairs summarizes many of these conditions:

> "I am making a brief statement of my experience with what I consider the greatest shame of the Indian Service—the rounding up of Indian children to be sent away to government boarding schools. . . . The children are caught, often roped like cattle, and taken away from their parents, many times never to return. They are transferred from school to school, given white people's names, forbidden to speak their own tongue, and when sent to distant schools are not taken home for three years." (Quoted in DeJong 1993, pg. 117-18)

Boarding schools came under criticism in the 1920s and 1930s, and the use of day schools and traditional public schools subsequently increased. In 1934, the Johnson O'Malley Act authorized federal contracts with states and territories to provide education services. However, there were many reports that the new approach restricted educational opportunities: geographic isolation limited access to day schools for many students, and some public schools serving both American Indian and other students provided fewer opportunities and resources to the American Indian students. Accordingly, the House Select Committee to Investigate Indian Affairs and Conditions in 1944 again advocated the use of boarding schools, particularly for rural communities like those on the Navajo reservation. In 1969, the congressional report, *Indian Education: A National Tragedy, A National Challenge*, commonly known as the Kennedy Report, documented the continuing disparities in education. The Indian Self-Determination and Education Assistance Act of 1975 (P.L. 93-638) began contract programs for tribes to administer schools, and provided funds for supplemental programs for American Indian children in public schools both on and off reservations.

The latest revision to this act was part of the 2001 No Child Left Behind Act (NCLB), which reauthorized the program as Title VII Part A of the Elementary and Secondary Education Act. Today, there are 183 Bureau of Indian Education (BIE) schools, 59 of which are federally run and the other 124 are tribally operated but funded through BIE contracts or grants. The historical experience of conflict between native cultures and formal schooling continues to be widely discussed in American Indian communities. Given this experience, the issues of personal identity and investment in formal educational institutions are likely to be particularly salient for students with American Indian ancestry.

3.3 Data and Methods

The main data source for this analysis is the National Assessment for Educational Progress (NAEP), a math and reading exam administered by the National Center for Education Statistics (NCES) to a random sample of 4th, 8th, and 12th grade students throughout the United States. This is the largest nationally representative exam and provides the best measure to compare students across areas in the United States. Samples are designed to be representative of public school students and schools in each state, with a probability sampling design that over-samples some student groups. Approximately 3,000 students are sampled from each participating state per subject and per grade, with close to 150,000 students total for each subject and grade. On average, there are 30 students in each sampled school.

Only about 1 percent of students in the U.S. are American Indians or Alaska Natives. As a result, accurate estimates of their performance are limited in most studies. To remedy this limitation, Executive Order 13336 authorized a new component of the NAEP, the National Indian Education Study (NIES). This study provides a nationally representative sample of American Indians in 4th and 8th grade for 2005, 2007, and 2009. The surveys include the universe of Bureau of Indian Education funded schools in 2005, 2007 and 2009. These data also include state representative samples for states with large populations of American Indian students with the number of states varying over time.

The outcomes used in this analysis are math and reading scores in 4th and 8th grade pooled over the years 2007 and 2009.[8] We drop observations from Hawaii and Alaska because populations and institutions for Native students in these states are significantly different than those in the mainland United States.[9] To be consistent with the achievement gap literature, we exclude Asians from our sample as well. Unfortunately, 12th grade students are not included in this analysis because they are not included in the NIES sample.

[8]We drop 2005 because it was the first year of the oversample. The American Indians in 2005 differ on some observable characteristics from those in the 2007 and 2009 files. Results using the 2005 data are largely similar to those reported in this paper.

[9]Test scores for American Indian students in Alaska are lower on average than in most other states, particularly in reading. Native Hawaiian students are classified in the "Asian and Pacific Islander" category, making it more difficult to analyze the relative performance of this population.

One unique feature of the NAEP is that it contains two measures of student race. The first is the race that is designated in the school records, school-reported race. The second is the race category that the student selected on the NAEP exam, student reported race. Prior to 2002, the NAEP used self-reported race as the primary race variable; since 2002, school-reported race is the primary race variable used. The majority of this paper's initial analysis uses school-reported race since it is currently designated the primary measure. This measure also makes the estimates comparable with other studies that estimate racial and ethnic achievement gaps. Subsequent sections of this analysis will investigate and discuss the relative importance of self- and school-identified race.

The NAEP uses a complex sampling design. First, to achieve a nationally representative sample, students have different probabilities of being selected based on demographic characteristics and their state. In order to get accurate statistical inferences, each student is assigned a sampling weight based on their probability of selection; students with a higher probability of selection are assigned lower weights. The original sampling weights (ORIGWT) are used in all of the regression analysis in this study. Second, the sample design differs from a simple random sample because sampling involves the selection of clusters of students based on school and demographic variables. The consequence is that observations are not independent of one another and traditional standard errors will be under-estimated. The standard procedure to adjust for this is to use the NAEP provided replicate weights (REPWGT) to produce jackknifed standard errors.[10]

Finally, the NAEP uses itemized response theory to preserve efficiency. Each individual student is assigned a subset of the assessment questions for a given subject area,

[10]See NAEP Data Companion for more details on these weights and the recommended jackknife procedure.

which means raw scores are not directly comparable across students. Instead, NAEP assigns to each student five score estimates, or "plausible values," ranging between 0 and 500 points. These plausible values are constructed by the NAEP using marginal estimation scaling model techniques for latent variables (Mislevy and Sheehan 1987). Essentially, these are random draws from a student's posterior distribution conditional on answers to the items assigned and demographic and background variables. The recommended procedure to deal with this survey feature is to use a multiple imputation method. This method involves running a separate regressions for each of the plausible values. Following Rubin (1996), the final point estimates are an average of the point estimates from these five regressions. The estimated variance of the final estimate combines the standard errors of the five estimates (calculated using the jackknifing procedures described above) and the variance across the estimates:

$$V(\hat{\beta}) = \sum_{i=1}^{m} se_i^2 + (1+\tfrac{1}{m})^{m-1} \sum_{i=1}^{m} (\hat{\beta}_m - \bar{\hat{\beta}}).$$ [11]

Table 1 reports the mean characteristics of students by school-identified racial group. The table indicates that demographic and family characteristics for American Indian students are similar to those for black and Hispanic students, with more than three quarters eligible for free and reduced price lunches, about a quarter living in homes with less than 10 books, and with parental education levels that tend to be lower than those for black students but higher than those for Hispanic students. Relatively more American Indian students (about 1 out of every 8) are also classified as disabled compared to the other groups.

[11] For more detail, see Section 3 of the NAEP Data Companion. Practically, this was implemented using STATA's multiple imputation package. This procedure for using NAEP data is standard when individual student scores are used, but is not necessary in studies that draw on the NAEP reported state averages or in studies that examine specific item responses.

Table 1 also shows that the mismatch between school- and self-identified race is greatest for American Indian students. Students who are identified to be white by their school also report themselves to be white most of the time: 80 percent of school-identified white fourth graders and 86 percent of school-identified white eighth graders self report the same race. For black students, the mismatch is marginally higher, with 75 percent of school-identified black fourth graders and 80 percent of school-identified black eighth graders also reporting themselves to be black. Hispanic students have the lowest rate of mismatch: 90 percent of the school-designated Hispanic 4th graders also self report as being Hispanic, and this match is 97 percent in 8th grade. However for American Indian students, only 55 percent of school-identified American Indian fourth graders and only 60 percent of school-identified American Indian eighth graders choose the same designation. The reverse mismatch (percentage of all self-identified students who are assigned the same category by the school) is similar; this finding is reported in Appendix Table A.1. Nearly all self-designated white and black students are assigned the same race by the school. About 60 percent of self-identified Hispanic students are similarly identified by their schools. However, among student who designate themselves to be American Indian, only 44 percent of 4th graders and 65 percent of 8th graders are also reported to be American Indian by the school. (Means of all other characteristics by self-reported race are also reported in Appendix Table A.1.)

Why is the mismatch so large for American Indian students? Mismatches, in part, stem from the fact that schools are much more likely to report students to be of the majority group. Nearly all students who call themselves white are also called white by their schools, but schools designate many students to be white who self report a different race. Because the white population is large, this mismatch looks small as a percentage of school-identified white students but large in comparison to the self-reported other race group. For example, in 4th grade math, there are about 4,460 students who self report as American Indian whose schools designate them as white. This is only 2 percent of the school-designated white sample but comprises 40% of the self-identifying American Indian sample. Because American Indian student populations are small in many states, mismatches are highest for this group. Furthermore, the rate of mismatch for self-identifying American Indian students is twice as high outside Western states than within the West where American Indian populations are largest.

Schools are also more likely to report the same race/ethnicity as the student when the student is limited English proficient. This is one reason why school-designated Hispanic students nearly universally self identify as Hispanic. However, there are about twice as many self-identifying Hispanic students as school-identified Hispanic students: many of these students speak English fluently and these students are more likely to be classified as white by their schools (or occasionally as American Indian). American Indian students have lower rates of limited English proficiency than Hispanic students, and therefore do not have as clear of a "marker" for the school. Hispanic populations are much larger than American Indian populations, and students who self designate as American Indian are more likely to be classified by their schools as Hispanic than the reverse.

3.4 Mean Differences in Achievement by School Reported Racial Groups

The first part of this analysis documents the mean differences in achievement by school reported race. Table 2 reports the mean math test score gap relative to whites in standard deviation units for 4th and 8th graders; Table 3 contains the results for reading. The first column of each set of results reports the raw gap with no controls. On average, school-identified American Indians score 60 percent of a standard deviation lower than school-identified whites on the NAEP math exam in 4th grade, increasing to three quarters of a standard deviation by 8th grade. This gap is about 15 percent smaller than the gap for school-identified black students. Gaps for black students are consistent with other estimates in the literature. School-identified Hispanic students' math gaps are nearly the same as those for American Indians.

The raw gaps in reading for all three groups are more similar, between 60 and 70 percent of a standard deviation. In addition, the math gaps widen with age for all three groups. However, the reading gaps in 8th grade are smaller or remain the same as in 4th grade for all three racial and ethnic groups.

The remaining columns of Tables 2 and 3 report regression adjusted mean differences in test scores by school reported race. Following the literature, covariates are added first for personal and family characteristics. Family and individual level characteristics include age, gender, disability status, English proficiency, eligibility for free lunch and number of books in home. Eighth grade results also include parental education measures (unavailable for fourth grade students). Adjusting for these factors reduces the gap considerably. About 40 percent of the raw math test score gap and about half of the raw reading gap for American Indians is explained by these factors. This leaves a gap relative to whites of about a third of a standard deviation in both math and reading. In unreported regressions examining subsets of these variables and in regressions run separately for each racial sub-group, relatively more of the gap for Hispanics is explained by English Language Learner status than for other groups. Socioeconomic characteristics (free lunch status, number of books in the home, and parental education) explain relatively more of the gap for American Indians than for black students or for Hispanic students.

In the next column (column 3 for 4th grade and column 6 for 8th grade), school fixed effects are included in addition to the individual and family characteristics. Including these fixed effects means that the gaps are average comparisons of students of different races and ethnicity within the same schools. The inclusion of school fixed effects further reduces the math gap for American Indians by about 10 percent of a standard deviation. School fixed effects have an even larger impact on the reading gap, reducing this gap by 13 to 18 percent of a standard deviation. Similar results are obtained using specific school characteristics such as level of urbanization, school size, and resource measures in place of fixed effects.[12] Surprisingly, these fixed effects have a negligible impact on the black and Hispanic gaps. This may suggest that school characteristics play a larger role for American Indians or that American Indians disproportionately select into different types of schools. We investigate the specific effect of school location on a reservation in Section 6. Taken together, family, individual and school characteristics explain roughly one half to two-thirds of the American Indian-white test score gap.

[12]Note that we are not claiming that schools matter less than socioeconomic factors in explaining achievement gaps: the order covariates are introduced matters **?**. We choose the order to be comparable with other results in the achievement gap literature.

These same factors account for two-thirds to three-fourths of the Hispanic-white gap and only about a third of the black-white gap.

The last columns of Table 2 and 3 include the same covariates as the previous columns as well as controls for parent education level. The NAEP data for 4th graders do not include parental education. The education level of a student's parent however, matters very little in explaining the gap for American Indians and blacks after conditioning on other covariates. The remainder of the analysis will exclude parental education to make 4th and 8th grade results comparable.

3.5 Racial Gaps at Different points in the Achievement Distribution

Examining test score behavior at the mean is useful. However, the conditional mean only captures shifts in the distribution of the outcome. If the overall shapes of the distribution of test scores are different by race/ethnicity, then assuming the adjusted mean difference in test scores is constant across the conditional distribution will be misleading. Quantile regression analysis allows for statistical inference at the median as well as non-central locations. In this context, it allows for the estimation of gaps across the test score distribution, relative to white students at the same point in the distribution of white students. For example, if the distribution of scores for one group had the same mean but was flatter than the distribution for another group, there would be a positive gap at the 90th percentile and a negative gap at the 10th percentile. Likewise, if one distribution had a thicker left tail, gaps at the 90th percentile might be small, while those at the 10th percentile would be large. However, some care is needed in the interpretation of the coefficients when including covariates. A quantile coefficient is the effect of the variable of interest on the distribution of the outcome variable, holding all else constant. However, the coefficient does not measure the marginal effect on individuals. This is a critical distinction: there is no guarantee that an individual will remain in the pth quantile after a covariate is marginally altered.

Figure 3.1 shows that the unadjusted test score gap for all races relative to whites for math and reading is largest at the 10th percentile and gets smaller monotonically across the distribution. Each point on the graph represents a test score gap relative to whites that corresponds to a given location in the distribution. For example, the upper left graph in Figure 1 shows that American Indians in 4th grade at the 10 percentile have a raw math gap of nearly 0.9 of a standard deviation relative to whites at that decile. This falls to less than 0.6 for American Indian students in the highest decile.

Figure 1 also shows that the difference in the unadjusted test score gap between the 10th percentile and the 90th percentile is the largest for American Indians. The gradient is steepest for American Indians in all grades and subjects, with a particularly sharp difference in the gradients in 8th grade. The gradient for black students is relatively flat. This implies that the overall shape of the distribution for black students is more similar to that for white students, although the overall distribution is shifted. For American Indian students, however, the distribution of raw scores has a much different shape.

This general pattern of larger gaps at lower deciles persists even after controlling for other characteristics. Table 4 reports quantile regression results with bootstrapped standard errors in parentheses, adjusting for individual, family, and school covariates. American Indians in the 90th percentile have a gap of 12 percent of a standard deviation in math in 4th grade and about a fourth of a standard deviation in 8th grade. However, at the 10th percentile these gaps are about twice as large. The reading results are even more stark: virtually no gap exists at the 90th percentile, while the gap at the 10th percentile is about 30 percent of a standard deviation in both 4th and 8th grade.

As in the raw scores, this pattern of much larger gaps at lower deciles is strikingly different for American Indian students than for other racial and ethnic groups. School-identified black and Hispanic students at the lowest end of the achievement distribution also do worse relative to their white counterparts than those at the highest end, but the difference is much smaller. For black students, the gap at the 10th percentile is only about 6 percentage points larger than at the 90th percentile, and for Hispanic students it is usually less than 5 percentage points larger.

One concern might be that the quantile results are driven by students attending schools located on reservations. In unreported results, we have estimated these quantile regressions for only students attending schools located outside of reservation boundaries. We find nearly identical estimates with similarly larger gaps at lower deciles than at higher deciles. (These results are available on request.) We also estimated the quantile regressions separately by race to examine the relative role of covariates on the quantiles. We find that the magnitude of impact of socioeconomic variables (like the number of books at home) has a relatively constant coefficient across the percentiles for black and hispanic students. However, the magnitude of the effect on American Indian students tends to be larger for the highest percentiles than for the lowest percentiles.

3.6 Geography, Identity and the Achievement Gap

The previous two sections have established that American Indians score lower than whites on the NAEP math and reading exams. In particular, the evidence suggests that the math gap increases as students progress in school and that the gradient with respect to the ability distribution is particularly steep for American Indians, with much larger gaps for the poorest performing students. Disadvantages in family background explain slightly less than half of the math gap and half of the reading gap. School characteristics further reduce these gaps in math and reading by another 5-18 percent of a standard deviation. This contribution of school effects is larger than for other racial and ethnic groups in the sample. The following two subsections investigate the unique geography and identity for American Indians as possible additional contributors.

3.6.1 Effect of Reservations on the American Indian Achievement Gap

About a fourth of American Indians live on reservations, which in many cases are remote areas with thin labor markets and few amenities. The NAEP oversamples these areas, with about 40 percent of the American Indian sample coming from schools located on reservations. The larger estimated impact of schools on test scores for American Indians relative to other groups may be in part related to reservation related factors, but using school fixed effects in the earlier analysis obscures the specific role of reservation location on American Indian students.

We identify schools located on reservations using GIS maps of reservation boundaries and each school's longitude and latitude. Reservation boundaries in some areas are quite large and include schools that serve a substantial number of non-American Indian students as well. Roughly 4,000 students in each sample attend schools on a reservation, about 25 percent of which are identified by the school as non-American Indian. Most of these students are reported to be white or Hispanic; very few are black. [13]

We examine the effect of reservation location by estimating a separate effect of reservations on American Indian and other students as in the following regression equation:

$$y = \beta_0 + \beta_1 Onres + \beta_2 AI + \beta_3 Onres * AI + \alpha Race + \sigma X + \gamma Z + \delta State + \varepsilon \qquad (3.1)$$

The variable *Race* includes controls for black, Hispanic and other. The vector X includes the standard set of individual and family characteristics as described in section 4. The vector Z is comprised of school characteristics such as level of urbanization and school enrollment.[14] State fixed effects are also included in this specification. Results are reported in Table 5.

Table 5 indicates that the estimated effect of attending a school located on a reservation, β_1, is negative, but is also small and indistinguishable from zero. This coefficient captures the average effect of attending a school on a reservation for students of all races and ethnicities. The coefficient on the variable AI, β_2, measures the test score gap for American Indians relative to white students in schools located outside reservation boundaries. The estimates indicate that school-identified American Indians attending

[13] Similar results are obtained when we used distance to the center of a reservation and tribal or BIE control of a school.

[14] We did not include expenditures as these were not available for BIE schools in 2009. However, results for 2007 that also include expenditures are similar to those reported here.

schools off of a reservation score on average about a quarter of a standard deviation lower in math and reading in 4th grade than white students attending schools off of a reservation. The gap in 8th grade math scores is slightly larger, while the gap in 8th grade reading scores is slightly smaller.

Table 5 reports that the gaps are larger in schools located on reservations. The interaction term, β_3, measures the additional gap for American Indian students attending a school on a reservation. The gap for an American Indian student on a reservation relative to a white student on a reservation is the sum of coefficients β_2 and β_3, and these gaps are uniformly larger than the off reservations gaps. Gaps on reservations are about 16 to 20 percent of a standard deviation larger than off a reservation (with the exception of reading in 8th grade); stated otherwise, on-reservation gaps are about 40 percent larger than off-reservation gaps. Due to data restrictions, these results can only control for a limited set of school characteristics, and it may be that schools located on reservations differ in systematic ways from schools in other locations. One key difference is that these schools have student bodies that are much more racially segregated. The Bureau of Indian Education schools have student bodies that are nearly exclusively American Indian. Of the roughly 250 other schools located on reservations, about 40 percent have student bodies that are 95 percent or more American Indian. Another 20 percent of schools on reservations have student bodies that are less than 20 percent American Indian. As a result, estimated test gaps on reservations compare American Indian and white students who on average attend different schools. The high correlation between location and the makeup of the schools makes it difficult to disentangle the effects of school characteristics, school location, and student race: given the small size of many of these schools and the limited variation in student characteristics within them, using school fixed effects with this sub-sample leads to imprecisely estimated coefficients.

3.6.2 Effect of Self Identification on the Racial Achievement Gap

As noted in the introduction, there are several reasons to believe that self-identity may be particularly salient for American Indian students. Predictions from the "single ideal" school model, put forth in Akerlof and Kranton (2002), suggest that students may make lower investments in formal schooling if the behaviors and norms associated with their social group do not align with those promoted by public schools. Given the long history of explicit opposition to native languages and cultures by schools, this theory may be particularly relevant to native populations. Furthermore, the high levels of intermarriage provide more scope for individually deciding among groups with which to identify.

As noted, the NAEP data include a student's self-reported race/ethnicity in addition to school reported race/ethnicity. This allows for investigation of how self-identification and school-identification separately relate to test scores. As Table 1 and Appendix Table A.1 indicated, the data reveal a significant discrepancy between school- and self-reported race for American Indian students, far more than for any other group.

The choice of race and ethnicity, both by students and by schools, is likely endogenous to student performance. Because of this self-selection, the coefficient estimates should not be interpreted as causal effect of racial identification, but the relationships nonetheless are revealing of the association between academic outcomes and self-identification or school-identification. Table 6 reports the test score gap for three mutually exclusive identification categories: school-reported-only, self-reported-only, and both school- and self-reported. In 4th and 8th grade, the unexplained math gap for students classified both by themselves and their schools as American Indian is substantially larger than the gap for students who are only self- or only school-identified. A similar pattern appears in the reading results. Furthermore, in 4th grade, both the math and the reading gap for students that are school-only identified are larger than the gaps for self-only identified American Indians. As students progress in school, however, this reverses. By 8th grade, the gap for self-only identified American Indians is much larger than the gap for school-only identified students. In fact, in 8th grade the reading test score gap among those who self-identify as American Indian is more than 10 times as large as the gap for students who are only school-identified as American Indian: school-only-identified American Indians do not experience a reading gap relative to whites by 8th grade, while self-identified students have gaps of a quarter of a standard deviation. The 8th grade math test score gap among self-only identified students is also larger than the gap for school-only identified students, although the difference in math gaps is not as dramatic as it is in the reading results.

A few of these patterns are similar for black students. The math and reading test score gaps are also largest among those that are classified as both self- and school- identified. Similarly, in 4th grade school-only identified students have larger gaps than self-only-identified students. Furthermore, the gaps for self-only-identified students grow from 4th to 8th grade. However, unlike American Indians, by 8th grade the math gap is still larger for school-only-identified blacks compared to self-only-identified blacks. The reading gap by 8th grade is slightly larger for self-only-identified blacks compared to school-only-identified blacks, but only by 8 percent of a standard deviation. This difference pales in comparison to the 25 percent of a standard deviation difference in the gap between these two race classifications for American Indians.

In contrast, out of the three mutually exclusive race categories, in 4th grade Hispanics that are school-only-identified experience the largest gap in math and reading, but by 8th grade, in both reading and math, the gap for self-only-identified Hispanics is largest. Like with American Indians and black students, the gap for self-only-identified students grows from 4th to 8th grade.

Appendix Table A.2, a companion table to Table 6, reports the effects for each type of race category at the 10th and 90th percentiles. These results are similar to the ones reported in Table 6. For all race classifications – self-only-identified, school-only-identified, and both self- and school-identified – American Indians, blacks and Hispanics at the 10th percentile experience a larger gap than those students at 90th percentile in the achievement distribution. As before, gradients are steeper for American Indians than other groups, although the gradient for self-only-identified American Indians is the flattest of the three categories.

Theoretical and empirical evidence from the identity literature suggests that there will be differences in individuals' behavior depending on the salience of group membership. One hypothesis is that the concentration of a group in a school affects the salience of group membership. Another possibility is that in locations where American Indians are better known by the general population, there may be more defined stereotypes or students may have a different propensities for how they classify themselves. In these data, the mismatch between school- and self-identification is much larger in states outside the West and in states with smaller percentages of American Indian students.

Accordingly, we disaggregate students in three different ways: by reservation location, by percent American Indian in the school, and by geographic location.

Table 7 reports regression results using the same specification used in Table 6. Panel A is a sub-sample of only those who attend a school on a reservation. As noted, about 25 percent of students on reservations are identified by their schools as a classification other than American Indian. Of these students, a little over a third self-identify as American Indian. Panel B is the complement of the Panel A sub-sample. It includes all the students that attend a school off a reservation; American Indians comprise roughly 3% of this sample.

Table 7 shows that, as before, students on a reservation have larger gaps on average than those off reservations in math, although the reading point estimates are more similar across the two locations. Gaps are also generally largest for students jointly identified as American Indian both by their schools and by themselves. Similarly, the pattern of the relative impacts of school and self identification are parallel across the two sub-samples, with generally declining gaps for school-only-identified students and rising gaps for self-identified students moving from 4th to 8th grade. However, the coefficient for self-only-identified students is not statistically significant at conventional levels on reservations. (This is due to the fact that the standard errors are much larger in the small sample of reservation students).

The next two panels disaggregate instead on the basis of whether American Indian students are a majority (Panel C) or a minority (Panel D) in the school. The average gaps vary across these two sub-samples in a similar way as they do when comparing reservation and off-reservation schools (Panel A and Panel B). Again, the gaps for students who are school-only-identified decline from 4th to 8th grade, while those for self-only-identified students rise. Students who are both school- and self-identified have the largest gaps, with point estimates that are similar across the sub-samples. Standard errors are again larger in the sample of students in highly segregated schools, but the point estimates are similar. Appendix Table A.3 replicates these results by Census region, again with similar results.

The analysis presents compelling evidence that even in eighth grade school- and self-racial identification of a student is remarkably inconsistent for American Indians, more so than for any other racial or ethnic group in the sample. The gap for American Indians who are self-only-identified is substantially larger than those that are school-only-identified in 8th grade. It also appears that the gaps for self-only-identified students become relatively larger as students progress in school while the gaps for school-only-identified students declines, and this is true across schools of varying compositions and locations.

3.7 Conclusion

American Indian academic achievement has been little studied, in spite of unique socioeconomic, geographic, and historical factors than may influence student outcomes. This study establishes a number of facts about the current state of American Indian academic achievement in comparison with other racial and ethnic groups in the United States. We investigate three general classes of explanations for racial and ethnic gaps. First, American Indian students on average come from more disadvantaged family and socioeconomic backgrounds than many other students. Second, schools for American Indian students are more likely to be located in more geographically isolated communities and on reservations. These and other school characteristics may affect performance. Third, the unique history of formal education for American Indians makes it more probable that conflicts related to group identification and academic achievement are salient for this population.

We find a raw achievement gap of about 60 to 70 percent of a standard deviation for both math and reading. These gaps are smaller than the gaps for blacks and similar to the gap for Hispanics. As with other groups, the gaps also appear to be larger in 8th grade than in 4th grade for math. For reading, gaps are somewhat smaller in 8th grade than in 4th grade for American Indian and Hispanic students, while remaining relatively unchanged for black students.

Although the raw gaps are similar to those of other racial and ethnic groups, there are a number of important differences. First, more of the gap can be explained with observable characteristics for American Indian and Hispanic students than for black students. School fixed effects explain more of the remaining gap for American Indians than for both other groups. Demographic and school characteristics explain about two thirds of the raw gap for American Indians: after controlling for individual, family, and school characteristics, unexplained gaps for reading in 4th and 8th grade and for math in 4th grade are about 20 percent of a standard deviation, while the unexplained gap for 8th grade math is about 40 percent of a standard deviation. (For black students, two-thirds of the raw gap remains unexplained after controlling for observed characteristics.) Location on a reservation also matters: a gap exists off reservations, but the achievement gap for American Indians students is somewhat larger on reservations than in other areas.

Second, we find that the distribution of test scores for American Indians is much different than for other racial and ethnic groups. Black and Hispanic students have score gaps that are comparatively uniform across the distribution: while students in the top quintile obviously perform better than those in the lowest quintile, the gap relative to white students in the same quintile is fairly similar across the distribution. In contrast, for American Indians the gaps relative to whites are much larger for students at the bottom of the achievement distribution. In reading, gaps at the 90th percentile are negligible and insignificant after adjusting for student and school characteristics, while at the 10th percentile, gaps are about a third of a standard deviation. In math, the gaps at the 10th percentile are about twice as large as the gaps at the 90th percentile.

Finally, the different historical context of education for American Indians suggests that there may be perceived conflicts between formal schooling and group identity. Although we cannot test for this directly, we exploit an unusual feature of the NAEP data: it contains measures of both self- and school-identified race and ethnicity. These data provide a unique opportunity to examine how self-identification and school-identification may have different associations with achievement. The match between self- and school-identification is lower for American Indians than for any other ethnic group. Gaps for students only identified as American Indian by the school decline between 4th and 8th grade, while gaps for self-identified students increase as students age. For example, there is no statistically significant gap in 8th grade reading for students who are only identified by their school as American Indian, while self-identifying American Indians (both with or without school identification) have a gap of one fourth of a standard deviation. In contrast, for black and Hispanic students, the pattern across these three mutually exclusive identification categories is less clear, with gaps that are more similar in magnitude. We do not find evidence that American Indian self-identity is more salient by region of the country, by the concentration of American Indians in a school, or in schools located on reservations. In all of these areas, the general pattern is a declining gap from 4th to 8th grade for school-only-identified American Indians. In contrast, the gaps for self-only-identified American Indians grow with age and are universally larger than those gaps for school-only-identified students by 8th grade.

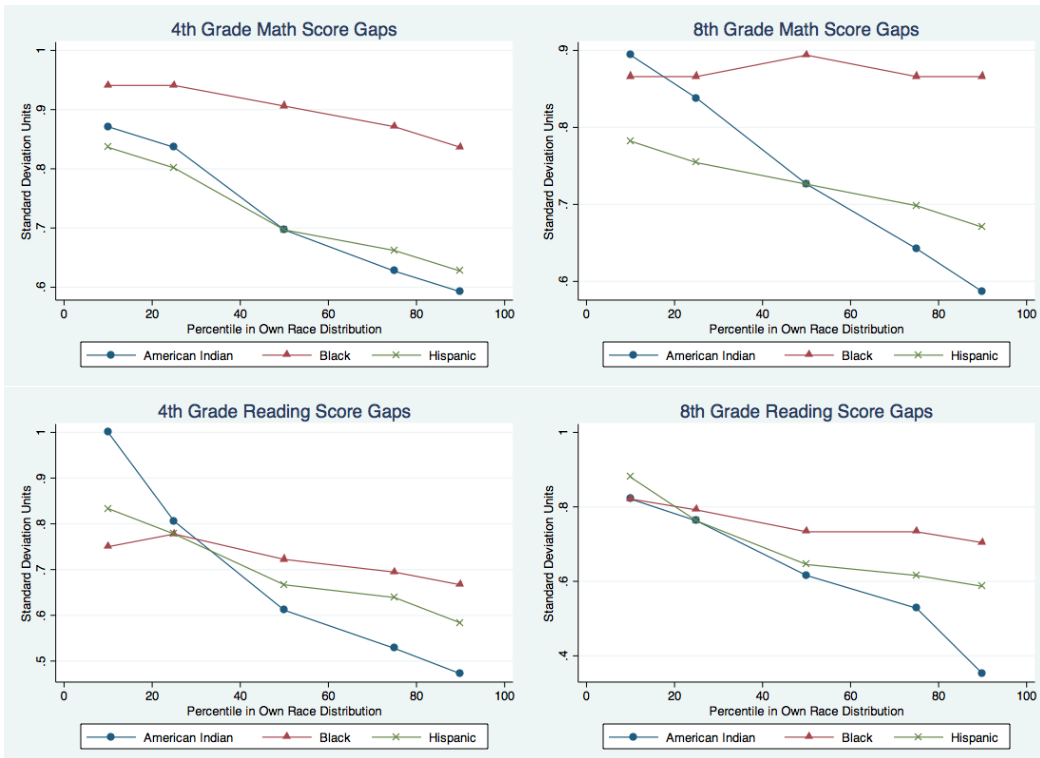

Figure 3.1: Raw Racial and Ethnic Test Score Gaps Relative to White Students in Same Percentile, weighted by sampling weights

	Math									
	4th					8th				
	Full Sample	White	American Indian	Black	Hispanic	Full Sample	White	American Indian	Black	Hispanic
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Test Scores										
Math	0.000	0.247	-0.458	-0.490	-0.314	0.000	0.291	-0.577	-0.604	-0.409
	(1.000)	(0.704)	(0.818)	(0.742)	(0.766)	(1.000)	(0.899)	(1.016)	(0.897)	(0.928)
Race- School Identified										
White	0.599	1.000	0.000	0.000	0.000	0.620	1.000	0.000	0.000	0.000
	(0.490)					(0.485)				
American Indian	0.027	0.000	1.000	0.000	0.000	0.027	0.000	1.000	0.000	0.000
	(0.163)					(0.162)				
Black	0.182	0.000	0.000	1.000	0.000	0.186	0.000	0.000	1.000	0.000
	(0.386)					(0.389)				
Hispanic	0.183	0.000	0.000	0.000	1.000	0.162	0.000	0.000	0.000	1.000
	(0.387)					(0.368)				
Other	0.009	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000
	(0.093)					(0.076)				
Race- Student Identified										
White	0.495	0.796	0.109	0.013	0.061	0.539	0.859	0.082	0.007	0.013
	(0.500)	(0.403)	(0.312)	(0.115)	(0.239)	(0.499)	(0.348)	(0.274)	(0.081)	(0.115)
American Indian	0.034	0.023	0.553	0.018	0.010	0.025	0.011	0.606	0.008	0.003
	(0.182)	(0.150)	(0.497)	(0.131)	(0.100)	(0.156)	(0.103)	(0.489)	(0.091)	(0.053)
Black	0.146	0.007	0.037	0.743	0.020	0.156	0.009	0.020	0.799	0.008
	(0.353)	(0.084)	(0.188)	(0.437)	(0.139)	(0.363)	(0.921)	(0.140)	(0.401)	(0.091)
Hispanic	0.279	0.125	0.202	0.165	0.904	0.227	0.075	0.149	0.093	0.971
	(0.448)	(0.330)	(0.402)	(0.371)	(0.294)	(0.419)	(0.263)	(0.356)	(0.291)	(0.166)
Other	0.046	0.049	0.099	0.061	0.005	0.053	0.047	0.143	0.093	0.004
	(0.209)	(0.216)	(0.298)	(0.238)	(0.071)	(0.224)	(0.212)	(0.351)	(0.291)	(0.063)
Other Controls										
Free/reduced lunch	0.492	0.299	0.786	0.775	0.797	0.443	0.263	0.748	0.728	0.751
	(0.500)	(0.458)	(0.410)	(0.418)	(0.402)	(0.497)	(0.440)	(0.434)	(0.445)	(0.433)
Books 0-10	0.124	0.062	0.223	0.208	0.229	0.145	0.089	0.230	0.188	0.296
	(0.330)	(0.242)	(0.416)	(0.406)	(0.420)	(0.352)	(0.284)	(0.421)	(0.391)	(0.457)
Books 11-25	0.215	0.152	0.289	0.300	0.327	0.214	0.158	0.292	0.296	0.318
	(0.411)	(0.359)	(0.453)	(0.458)	(0.469)	(0.410)	(0.365)	(0.454)	(0.456)	(0.466)
Books 26-100	0.336	0.378	0.283	0.272	0.268	0.349	0.372	0.320	0.340	0.274
	(0.472)	(0.485)	(0.451)	(0.445)	(0.443)	(0.477)	(0.483)	(0.466)	(0.474)	(0.446)
Books 100 plus	0.325	0.408	0.205	0.220	0.176	0.293	0.381	0.159	0.176	0.112
	(0.468)	(0.491)	(0.403)	(0.414)	(0.381)	(0.455)	(0.486)	(0.365)	(0.380)	(0.315)
Male	0.508	0.512	0.502	0.500	0.503	0.502	0.507	0.497	0.485	0.506
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Disabled	0.119	0.124	0.135	0.120	0.100	0.102	0.100	0.137	0.114	0.093
	(0.324)	(0.330)	(0.341)	(0.325)	(0.300)	(0.303)	(0.300)	(0.344)	(0.318)	(0.290)
English language learner	0.083	0.007	0.142	0.018	0.392	0.044	0.004	0.132	0.011	0.220
	(0.277)	(0.085)	(0.349)	(0.133)	(0.488)	(0.205)	(0.064)	(0.339)	(0.103)	(0.414)
Age	7.512	7.513	7.581	7.518	7.497	11.435	11.428	11.619	11.461	11.402
	(1.161)	(1.141)	(1.158)	(1.194)	(1.190)	(1.162)	(1.148)	(1.177)	(1.201)	(1.166)
Mother-H.S drop out	-	-	-	-	-	0.104	0.068	0.137	0.085	0.259
						(0.305)	(0.251)	(0.344)	(0.279)	(0.438)
Mother- H.S. grad	-	-	-	-	-	0.208	0.208	0.230	0.233	0.179
						(0.406)	(0.406)	(0.421)	(0.423)	(0.383)
Mother- some college	-	-	-	-	-	0.173	0.182	0.174	0.192	0.119
						(0.379)	(0.386)	(0.379)	(0.394)	(0.324)
Mother- college grad	-	-	-	-	-	0.364	0.433	0.249	0.334	0.152
						(0.481)	(0.495)	(0.432)	(0.472)	(0.359)
Mother- don't know	-	-	-	-	-	0.151	0.110	0.210	0.156	0.291
						(0.358)	(0.313)	(0.408)	(0.363)	(0.454)
Father- H.S. drop out	-	-	-	-	-	0.108	0.083	0.128	0.082	0.231
						(0.310)	(0.276)	(0.334)	(0.275)	(0.422)
Father- H.S. grad	-	-	-	-	-	0.207	0.215	0.240	0.219	0.160
						(0.405)	(0.411)	(0.427)	(0.414)	(0.366)
Father- some college	-	-	-	-	-	0.136	0.150	0.136	0.131	0.091
						(0.343)	(0.357)	(0.343)	(0.337)	(0.288)
Father- college grad	-	-	-	-	-	0.306	0.382	0.169	0.231	0.126
						(0.461)	(0.486)	(0.375)	(0.421)	(0.331)
Father- don't know	-	-	-	-	-	0.242	0.170	0.327	0.337	0.392
						(0.428)	(0.376)	(0.469)	(0.473)	(0.488)
Observations	322,992	193,506	8,871	58,727	59,066	278,096	172,324	7,492	51,583	45,064

Notes: Means of student characeristics are based on math sample. Means for reading sample are nearly identical, and are available on request.
Standard Deviations in parentheses. Parental education not available for 4th grade.

Table 3.1: Summary Statistics by Race: Means of Student Characteristics by School Reported Race

	Math						
	4th	**4th**	**4th**	**8th**	**8th**	**8th**	**8th**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
American Indian- School Identified	-0.598***	-0.340***	-0.203***	-0.740***	-0.418***	-0.366***	-0.344***
	(0.016)	(0.014)	(0.018)	(0.023)	(0.021)	(0.024)	(0.024)
Black- School Identified	-0.735***	-0.482***	-0.427***	-0.879***	-0.580***	-0.554***	-0.564***
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)
Hispanic- School Identified	-0.599***	-0.182***	-0.157***	-0.724***	-0.237***	-0.237***	-0.188***
	(0.009)	(0.007)	(0.007)	(0.013)	(0.011)	(0.012)	(0.011)
Other- School Identified		-0.085***	-0.127***		-0.097***	-0.111***	-0.106***
		(0.018)	(0.018)		-0.031	(0.030)	(0.030)
Year 2009		-0.058***	-0.110***		-0.212***	-0.149**	-0.138**
		(0.010)	(0.011)		-0.015	(0.058)	(0.054)
Age		-0.033***	-0.049***		-0.134***	-0.137***	-0.125***
		(0.004)	(0.003)		-0.007	(0.007)	(0.007)
Male		0.109***	0.110***		0.155***	0.156***	0.147***
		(0.004)	(0.004)		-0.006	(0.006)	(0.006)
Disabled		-0.544***	-0.561***		-0.890***	-0.882***	-0.851***
		(0.007)	(0.007)		-0.012	(0.012)	(0.012)
Limited English Proficiency		-0.421***	-0.385***		-0.703***	-0.677***	-0.653***
		(0.012)	(0.010)		-0.019	(0.019)	(0.019)
Free Lunch		-0.293***	-0.188***		-0.298***	-0.261***	-0.194***
		(0.006)	(0.005)		-0.008	(0.008)	(0.008)
Books 0-10		-0.396***	-0.335***		-0.655***	-0.626***	-0.525***
		(0.008)	(0.008)		-0.010	(0.009)	(0.009)
Books 11-25		-0.308***	-0.258***		-0.509***	-0.483***	-0.409***
		-0.006	(0.006)		-0.009	(0.008)	(0.008)
Books 26-100		-0.062***	-0.043***		-0.263***	-0.247***	-0.207***
		(0.005)	(0.005)		(0.006)	(0.006)	(0.006)
Mother- H.S. grad		-	-		-	-	-0.0160
							(0.011)
Mother- some college		-	-		-	-	0.135***
							(0.010)
Mother- college grad		-	-		-	-	0.125***
							(0.012)
Mother- don't know		-	-		-	-	-0.085***
							(0.011)
Father- H.S. grad		-	-		-	-	0.013
							(0.009)
Father- some college		-	-		-	-	0.133***
							(0.010)
Father- college grad		-	-		-	-	0.198***
							(0.009)
Father- don't know		-	-		-	-	0.006
							(0.008)
School Fixed Effects	No	No	Yes	No	No	Yes	Yes
Observations	343,245	322,992	322,992	282,242	278,096	278,096	278,096

The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table 3.2: Math Standardized Test Score Gaps in Standard Deviation Units

	Reading						
	4th	4th	4th	8th	8th	8th	8th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
American Indian- School Identified	-0.642***	-0.364***	-0.186***	-0.575***	-0.280***	-0.155***	-0.138***
	(0.024)	(0.021)	(0.026)	(0.023)	(0.022)	(0.028)	(0.028)
Black- School Identified	-0.692***	-0.415***	-0.340***	-0.699***	-0.435***	-0.362***	-0.372***
	(0.010)	(0.010)	(0.010)	(0.008)	(0.007)	(0.008)	(0.008)
Hispanic- School Identified	-0.676***	-0.163***	-0.104***	-0.640***	-0.154***	-0.098***	-0.065***
	(0.012)	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)
Other- School Identified		-0.035	-0.055**		-0.108***	-0.117***	-0.113***
		(0.021)	(0.019)		(0.031)	(0.031)	(0.030)
Year 2009		-0.203***	-0.264***		-0.299***	-0.321***	-0.301***
		(0.014)	(0.013)		(0.014)	(0.013)	(0.013)
Age		-0.067***	-0.0751***		-0.129***	-0.125***	-0.116***
		(0.006)	(0.006)		(0.006)	(0.006)	(0.005)
Male		-0.146***	-0.138***		-0.206***	-0.198***	-0.199***
		(0.005)	(0.004)		(0.006)	(0.006)	(0.006)
Disabled		-0.807***	-0.762***		-0.860***	-0.790***	-0.758***
		(0.014)	(0.012)		(0.015)	(0.014)	(0.014)
Limited English Proficiency		-0.672***	-0.587***		-0.851***	-0.767***	-0.734***
		(0.015)	(0.013)		(0.021)	(0.020)	(0.019)
Free Lunch		-0.387***	-0.263***		-0.283***	-0.195***	-0.154***
		(0.006)	(0.006)		(0.007)	(0.007)	(0.007)
Books 0-10		-0.455***	-0.374***		-0.674***	-0.592***	-0.523***
		(0.012)	(0.010)		(0.011)	(0.010)	(0.010)
Books 11-25		-0.331***	-0.273***		-0.476***	-0.417***	-0.372***
		(0.007)	(0.007)		(0.009)	(0.009)	(0.009)
Books 26-100		-0.084***	-0.061***		-0.227***	-0.194***	-0.171***
		(0.006)	(0.005)		(0.008)	(0.007)	(0.007)
Mother- H.S. grad		-	-		-	-	-0.027**
							(0.011)
Mother- some college		-	-		-	-	0.122***
							(0.011)
Mother- college grad		-	-		-	-	0.065***
							(0.010)
Mother- don't know		-	-		-	-	-0.141***
							(0.012)
Father- H.S. grad		-	-		-	-	0.037***
							(0.009)
Father- some college		-	-		-	-	0.120***
							(0.011)
Father- college grad		-	-		-	-	0.136***
							(0.009)
Father- don't know		-	-		-	-	0.012
							(0.010)
School Fixed Effects	No	No	Yes	No	No	Yes	Yes
Observations	329,873	325,780	325,780	288,246	284,175	284,175	284,175

The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table 3.3: Reading Standardized Test Score Gaps in Standard Deviation Units

	Math		Reading	
	4th	**8th**	**4th**	**8th**
	(1)	**(2)**	**(3)**	**(4)**
American Indian- School Identified				
10th-Percentile	-0.281***	-0.485***	-0.327***	-0.318***
	(0.042)	(0.042)	(0.049)	(0.049)
25th-Percentile	-0.232***	-0.415***	-0.249***	-0.236***
	(0.032)	(0.040)	(0.033)	(0.031)
50th Percentile	-0.194***	-0.348***	-0.164***	-0.158***
	(0.024)	(0.028)	(0.029)	(0.028)
75th Percentile	-0.165***	-0.300***	-0.111***	-0.083**
	(0.028)	(0.027)	(0.034)	(0.035)
90th Percentile	-0.121***	-0.267***	-0.068	0.016
	(0.027)	(0.032)	(0.050)	(0.047)
Black- School Identified				
10th-Percentile	-0.460***	-0.590***	-0.365***	-0.393***
	(0.012)	(0.013)	(0.018)	(0.016)
25th-Percentile	-0.439***	-0.565***	-0.351***	-0.381***
	(0.009)	(0.012)	(0.016)	(0.012)
50th Percentile	-0.427***	-0.550***	-0.337***	-0.359***
	(0.009)	(0.008)	(0.011)	(0.011)
75th Percentile	-0.411***	-0.530***	-0.324***	-0.343***
	(0.009)	(0.010)	(0.011)	(0.009)
90th Percentile	-0.390***	-0.511***	-0.313***	-0.331***
	(0.010)	(0.014)	(0.014)	(0.017)
Hispanic- School Identified				
10th-Percentile	-0.157***	-0.276***	-0.107***	-0.102***
	(0.017)	(0.021)	(0.021)	(0.017)
25th-Percentile	-0.159***	-0.247***	-0.111***	-0.106***
	(0.013)	(0.012)	(0.014)	(0.015)
50th Percentile	-0.156***	-0.233***	-0.107***	-0.102***
	(0.010)	(0.010)	(0.016)	(0.015)
75th Percentile	-0.157***	-0.209***	-0.101***	-0.096***
	(0.010)	(0.015)	(0.013)	(0.012)
90th Percentile	-0.144***	-0.194***	-0.089***	-0.085***
	(0.011)	(0.018)	(0.021)	(0.019)
Observations	322,992	278,096	325,780	284,175

Notes: Racial subgroups also include a dummy variable for other. Additionally, the specification includes family/individual controls (see Table 2) and school fixed effects. The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table 3.4: Standardized Test Score Gaps by Quantile

	Math		Reading	
	4th	**8th**	**4th**	**8th**
	(1)	**(2)**	**(3)**	**(4)**
On Reservation	-0.057	-0.0103	-0.0585	-0.0207
	(0.051)	(0.044)	(0.066)	(0.039)
American Indian- School Identified	-0.244***	-0.327***	-0.242***	-0.208***
	(0.018)	(0.024)	(0.026)	(0.027)
On Reservation*American Indian-School Identified	-0.162**	-0.208***	-0.214**	-0.0542
	(0.066)	(0.064)	(0.089)	(0.061)
Observations	312,160	268,520	314,806	274,728

Notes: Racial subgroups also include a dummy variable for other. Additionally, the specification includes family/individual controls (see Table 2) as well as state fixed effects and school controls (enrollment and level of urbanization). The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table 3.5: Standardized Test Score Gaps by Reservation Location

	Math		Reading	
	4th	**8th**	**4th**	**8th**
	(1)	**(2)**	**(3)**	**(4)**
American Indian				
School Identified Only	-0.169***	-0.193***	-0.147***	-0.022
	(0.022)	(0.031)	(0.031)	(0.031)
Self Identified Only	-0.154***	-0.288***	-0.113***	-0.251***
	(0.017)	(0.028)	(0.021)	(0.028)
School and Self Identified	-0.238***	-0.495***	-0.216***	-0.251***
	(0.025)	(0.032)	(0.029)	(0.036)
Black				
School Identified Only	-0.402***	-0.388***	-0.345***	-0.258***
	(0.010)	(0.016)	(0.013)	(0.017)
Self Identified Only	-0.249***	-0.365***	-0.262***	-0.344***
	(0.020)	(0.029)	(0.024)	(0.026)
School and Self Identified	-0.440***	-0.603***	-0.342***	-0.400***
	(0.007)	(0.010)	(0.011)	(0.008)
Hispanic				
School Identified Only	-0.270***	-0.253***	-0.284***	-0.270***
	(0.015)	(0.046)	(0.021)	(0.051)
Self Identified Only	-0.162***	-0.308***	-0.201***	-0.294***
	(0.006)	(0.011)	(0.009)	(0.010)
School and Self Identified	-0.171***	-0.281***	-0.122***	-0.145***
	(0.008)	(0.012)	(0.012)	(0.010)
Observations	322,992	278,096	325,780	284,175

Notes: Racial subgroups also include a dummy variable for other. Additionally, the specification includes family/individual controls (See Table 2) and school fixed effects. The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table 3.6: Standardized Test Score Gaps by Group Identification

	Math		Reading	
	4th	**8th**	**4th**	**8th**
	(1)	(2)	(3)	(4)
Panel A: School Located On Reservation				
AI School Identified Only	-0.400***	-0.372***	-0.215	-0.187
	(0.101)	(0.099)	(0.185)	(0.122)
AI Self Identified Only	-0.208	-0.668*	-0.308	-0.222
	(0.204)	(0.329)	(0.303)	(0.232)
AI School and Self Identified	-0.480***	-0.659***	-0.201	-0.331*
	(0.119)	(0.070)	(0.120)	(0.092)
Observations	4,555	3,632	4,782	3,850
AI Observations*	3,855	3,075	4,233	3,390
Panel B:School Located Off Reservation				
AI School Identified Only	-0.158***	-0.180***	-0.135***	-0.015
	(0.024)	(0.034)	(0.034)	(0.034)
AI Self Identified Only	-0.155***	-0.292***	-0.111***	-0.243***
	(0.017)	(0.028)	(0.021)	(0.028)
AI School and Self Identified	-0.235***	-0.446***	-0.213***	-0.245***
	(0.029)	(0.035)	(0.032)	(0.040)
Observations	311,538	268,394	313,985	274,307
AI Observations*	11,568	7,185	11,523	7,371
Panel C: More than 50% AI in School				
AI School Identified Only	-0.272***	-0.0695	-0.244***	-0.134
	(0.077)	(0.083)	(0.068)	(0.093)
AI Self Identified Only	-0.109	-0.196	-0.202	-0.333
	(0.097)	(0.280)	(0.116)	(0.173)
AI School and Self Identified	-0.330***	-0.287***	-0.269***	-0.359
	(0.061)	(0.074)	(0.081)	(0.068)
Observations	8,301	6,842	8,124	6,934
AI Observations*	5,008	4,036	5,145	4,229
Panel D: Less than 50% AI in School				
AI School Identified Only	-0.198***	-0.179***	-0.138***	-0.017
	(0.031)	(0.034)	(0.035)	(0.038)
AI Self Identified Only	-0.192***	-0.289***	-0.117***	-0.274***
	(0.021)	(0.028)	(0.022)	(0.031)
AI School and Self Identified	-0.306***	-0.459***	-0.235***	-0.276***
	(0.038)	(0.040)	(0.034)	(0.046)
Observations	314,691	271,254	317,656	277,241
AI Observations*	10,693	6,457	10,892	6,792

*AI observations includes school identified American Indians, self identified American Indians and those that are school and self identified as American Indian.

Notes: Controls include all racial dummy variables, family/individual controls (see Table 2) and school fixed effects. The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table 3.7: Standardized Test Score Gaps by School Characteristics and Group Identification

Bibliography

AINSWORTH-DARNELL, J. W. AND D. DOWNEY, "Assessing the Oppositional Cultural Explanation for Racial/Ethnic Differences in School Performance," *American Sociological Review* 63 (1998), 536–553.

AKEE, R. Q., T. J. HALLIDAY AND S. KWAK, "Investigating the Effects of Furloughing Public School Teachers on Juvenile Crime in Hawaii," *Economics of Education Review* 42 (2014), 1–11.

AKERLOF, G. A. AND R. E. KRANTON, "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature* 40 (2002), 1167–1201.

ANDERSON, D. M., "In school and out of trouble? The minimum dropout age and juvenile crime," *Review of Economics and Statistics* (2012).

ANDERSON, D. M., B. HANSEN AND M. B. WALKER, "The minimum dropout age and student victimization," *Economics of Education Review* 35 (2013), 66–74.

ANDERSON, M. AND M. B. WALKER, "Does Shortening the School Week Impact Student Performance? Evidence from the Four-Day School Week," *Education Finance and Policy* (2014).

ARMOR, D., "Why is Black Educational Achievement Rising?," *Public Interest:* (1992), 65–80.

AUSTEN-SMITH, D. AND R. FRYER, "An Economic Analysis of 'Acting White'," *Quarterly Journal of Economics* (May 2005), 551–583.

BEDARD, K. AND C. DO, "Are middle schools more effective? The impact of school structure on student outcomes," *Journal of Human Resources* 40 (2005), 660–682.

BENJAMIN, C. J., D. AND J. STRICKLAND, "Social identity and preferences," *American Economic Review* 100 (2010), 1913–1928.

BERTHELON, M. E. AND D. I. KRUGER, "Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in Chile," *Journal of Public Economics* 95 (2011), 41–53.

BIE, "The Bureau of Indian Education," http://www.bia.gov/WhatWeDo (June 2012).

BLACK, S. E., P. J. DEVEREUX AND K. G. SALVANES, "Staying in the Classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births*," *The Economic Journal* 118 (2008), 1025–1054.

BOOZER, M. AND S. E. CACCIOLA, "Inside the'Black Box'of Project STAR: Estimation of peer effects using experimental data," *Yale Economic Growth Center Discussion Paper* (2001).

BROOKS-GUNN, J., G. DUNCAN AND J. ABER, *Neighborhood poverty: Context and consequences for children*, volume 1 (Russell Sage Foundation Publications, 1997).

BROWN, C. AND M. CORCORAN, “Sex-Based Differences in School Content and the Male-Female Wage Gap,” *Journal of Labor Economics* 15 (1997), 431–65.

BROWN, J., “Quitters never win: The (adverse) incentive effects of competing with superstars,” *Journal of Political Economy* 119 (2011), 982–1013.

CARRELL, S. E., R. L. FULLERTON AND J. E. WEST, “Does Your Cohort Matter? Measuring Peer Effects in College Achievement,” *Journal of Labor Economics* 27 (2009), 439–464.

CARRELL, S. E., M. E. PAGE AND J. E. WEST, “Sex and science: How professor gender perpetuates the gender gap,” *The Quarterly Journal of Economics* 125 (2010), 1101–1144.

CENSUS, U. S., “Poverty Status in the Past 12 Months, 2006-2010 American Community Survey 5-Year Estimates (American Community Survey). S1701,” (2012).

CLOTFELTER, C. T., H. F. LADD AND J. L. VIGDOR, “The academic achievement gap in grades 3 to 8,” *The Review of Economics and Statistics* 91 (2009), 398–419.

COOK, P. AND J. LUDWIG, “Weighing the Burden of ’Acting White’: Are There Race Differences in Attitudes Towards Education?,” *Journal of Public Policy and Analysis* 16 (1997), 256–278.

CROSS, A. B., D. C. GOTTFREDSON, D. M. WILSON, M. RORIE AND N. CONNELL, “The impact of after-school programs on the routine activities of middle-school students: Results from a randomized, controlled trial*,” *Criminology & Public Policy* 8 (2009), 391–412.

DONIS-KELLER, C. AND D. L. SILVERNAIL, *Research brief: A review of the evidence on the four-day school week* (Center for Education Policy, Applied Research and Evaluation, University of Southern Maine, 2009).

DUNCAN, B. AND S. TREJO, “Ethnic Identification, Intermarriage, and Unmeasured Progress by Mexican Americans,” in G. J. Borjas, ed., *Mexican Immigration to the United States* (Chicago: National Bureau of Economic Research and University of Chicago Press, 2007), 229–267.

FLETCHER, J., “Is identification with school the key component in the Black Box of education outcomes? Evidence from a randomized experiment,” *Economics of Education Review* 28 (2009), 662–671.

FORDHAM, S., *Blacked Out: Dilemmas of Race, Identity and Success at Capital High* (Chicago: University of Chicago Press, 1996).

FORDHAM, S. AND J. OGBU, “Black Students’ School Success: Coping with the "Burden of ’Acting White’",” *Urban Review* 18 (1986), 176–206.

FOSTER, G., “It’s not your peers, and it’s not your friends: Some progress toward understanding the educational peer effect mechanism,” *Journal of public Economics* 90 (2006), 1455–1475.

FRIEDBERG, L., “Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data,” *The American Economic Review* 88 (1998), 608–627.

FRYER, R., “An Economic Approach to Cultural Capital,” working paper, University of Chicago, 2002.

———, “The Importance of Segregation Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap,” *Handbook of Social Economics* 1 (March 2010).

FRYER, R. AND S. LEVITT, “Understanding the Black-White Test Score Gap in the First Two Years of School,” *Review of Economics and Statistics* 86 (2004), 447–464.

FRYER, R. AND P. TORELLI., “An Empirical Analysis of ’Acting White’,” *Journal of Public Economics* 94 (2010), 380–396.

GARRATT, R. J., C. WEINBERGER AND N. JOHNSON, “The state street mile: Age and gender differences in competition aversion in the field,” *Economic Inquiry* 51 (2013), 806–815.

GNEEZY, U., M. NIEDERLE AND A. RUSTICHINI, “Performance in competitive environments: Gender differences,” *QUARTERLY JOURNAL OF ECONOMICS-CAMBRIDGE MASSACHUSETTS-* 118 (2003), 1049–1074.

GRAU, E. AND M. F. SHAUGHNESSY, “The Four Day School Week: An Investigation and Analysis,” (1987).

GRIFFITH, M., “What Savings Are Produced by Moving to a Four-Day School Week?,” *Education Commission of the States (NJ3)* (2011).

GROGGER, J. AND E. EIDE, “Changes in college skills and the rise in the college wage premium,” *Journal of Human Resources* (1995), 280–310.

GRUBER, J. AND M. HANRATTY, "The Labor-Marker Effects of Introducing National Health Insurance: Evidence From Canada," *Journal of Business & economic statistics* 13 (1995), 163–173.

HAN, L. AND T. LI, "The gender difference of peer influence in higher education," *Economics of Education review* 28 (2009), 129–134.

HANUSHEK, E. A., J. F. KAIN, J. M. MARKMAN AND S. G. RIVKIN, "Does peer ability affect student achievement?," *Journal of applied econometrics* 18 (2003), 527–544.

HERRING, C., "School's New Math: The Four-Day Week," *Wall Street Journal* (2010).

HOFFMANN, F. AND P. OREOPOULOS, "A professor like me the influence of instructor gender on college achievement," *Journal of Human Resources* 44 (2009), 479–494.

HOXBY, C., "Peer effects in the classroom: Learning from gender and race variation," Technical Report, National Bureau of Economic Research, 2000.

JACOB, B. A. AND L. LEFGREN, "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime," *American Economic Review* 93 (2003), 1560–1577.

JENCKS, C. AND M. PHILLIPS, "The Black-White Test Score Gap: An Introduction," in *The Black-White Test Score Gap* (1998), 1–51.

JOHNSON, W., D. NEAL ET AL., "Basic skills and the black-white earnings gap," *The Black-White test score gap* (1998), 480–500.

KINGSBURY, K., "Four-Day School Weeks," *TIMES* (2008).

KUZIEMKO, I., R. W. BUELL, T. REICH AND M. I. NORTON, "Last-place Aversion: Evidence and Redistributive Implications," *The Quarterly Journal of Economics* 129 (2014), 105–149.

LOCHNER, L. AND E. MORETTI, "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *The American Economic Review* 94 (2004), 155–189.

LOFQUIST, D. ET AL., "Households and Families: 2010," http://www.census.gov/hhes/families/ (2012).

LUALLEN, J., "School's out... forever: A study of juvenile crime, at-risk youths and teacher strikes," *Journal of urban economics* 59 (2006), 75–103.

LYLE, D. S., "Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point," *The Review of Economics and Statistics* 89 (2007), 289–299.

MAYER, S., *What Money Can't Buy: Family Income and Children's Life Chances* (Harvard University Press, 1997).

MOFFITT, R. A. ET AL., "Policy interventions, low-level equilibria, and social interactions," *Social dynamics* 4 (2001), 45–82.

MURNANE, R., J. WILLETT, Y. DUHALDEBORDE AND J. TYLER, "How important are the cognitive skills of teenagers in predicting subsequent earnings," *Journal of Policy Analysis and Management* 19 (2000), 547–568.

MURNANE, R., J. WILLETT AND F. LEVY, "The growing importance of cognitive skills in wage determination," Technical Report, National Bureau of Economic Research, 1995.

NIEDERLE, M. AND L. VESTERLUND, "Do women shy away from competition? Do men compete too much?," *Quarterly Journal of Economics* 122 (2007), 1067–1101.

OGORMAN, M., "Educational disparity and the persistence of the black-white wage gap in the US," *Economics of Education Review* 29 (2010), 526–542.

O'NEILL, J., "The role of human capital in earnings differences between black and white men," *The Journal of Economic Perspectives* 4 (1990), 25–45.

OST, B., "The role of peers and grades in determining major persistence in the sciences," *Economics of Education Review* 29 (2010), 923–934.

PAGLIN, M. AND A. RUFOLO, "Heterogeneous human capital, occupational choice, and male-female earnings differences," *Journal of Labor Economics* (1990), 123–144.

POP-ELECHES, C. AND M. URQUIOLA, "Going to a better school: Effects and behavioral responses," Technical Report, National Bureau of Economic Research, 2011.

RASK, K. AND J. TIEFENTHALER, "The role of grade sensitivity in explaining the gender imbalance in undergraduate economics," *Economics of Education Review* 27 (2008), 676–687.

REYHNER, J. AND J. EDER, *American Indian education: A history* (University of Oklahoma Press, 2006).

RODRÍGUEZ-PLANAS, N., "School and drugs: Closing the gap-Evidence from a randomized trial in the US," Technical Report, Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit, 2012.

ROSE, H. AND J. BETTS, "The effect of high school courses on earnings," *Review of Economics and Statistics* 86 (2004), 497–513.

SACERDOTE, B. I., "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics* 116 (2001), 681–704.

SNYDER, H. N. AND M. SICKMUND, "Juvenile offenders and victims: 2006 national report.," *Office of juvenile justice and delinquency prevention* (2006).

STEELE, C. AND J. ARONSON, *Stereotype Threat and the Test Performance of Academically Successful African Americans* (Washington D.C.: The Brookings Institution Press, 1998).

STINEBRICKNER, R. AND T. R. STINEBRICKNER, "What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds," *Journal of public Economics* 90 (2006), 1435–1454.

SZASZ, M., *Education and the American Indian: The road to self-determination since 1928* (UNM Press, 1999).

WEINBERGER, C., "Is Teaching More Girls More Math the Key to Higher Wages," *Squaring Up: Policy Strategies to Raise Womens Income in the US* (2001).

WEINBERGER, C. J., "Mathematical college majors and the gender gap in wages," *Industrial Relations: A Journal of Economy and Society* 38 (1999), 407–413.

WHITMORE, D., "Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment," *American Economic Review* (2005), 199–203.

WOOLDRIDGE, J. M., "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics* 90 (1999), 77–97.

ZIMMERMAN, D. J., "Peer effects in academic outcomes: Evidence from a natural experiment," *Review of Economics and Statistics* 85 (2003), 9–23.

Appendices

Appendix A

The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence

Appendix A. The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence

Table A1: STEM Majors

Major Requires CHEM1A	**STEM Majors**
X	Biology
X	Biochemistry
X	Biopsychology
X	Chemistry
X	Engineering
X	Computer Engineering
	Computer Science
X	Earth Science
X	Ecology
X	Environmental Science
	Mathematics
	Statistics
X	Geophysics
X	Hydrology
X	Zoology
X	Pharmacology
X	Physics
X	Physiology
X	Physical Geography

Table A2: Balance Test – Orientation Attendees vs. Non-Attendees

	Summer Orientation Group	Did Not Attend Orientation	Diff. (1) - (2)
	(1)	(2)	(3)
Student Background and Class Characteristics			
No. On-track is above median	0.48	0.46	0.02
	(0.01)	(0.04)	(0.04)
Lecture is at 8 or 9 a.m.	0.35	0.39	-0.04
	(0.01)	(0.04)	(0.04)
Female	0.49	0.52	-0.03
	(0.01)	(0.04)	(0.04)
URM (underrepresented minority)	0.38	0.48	-0.10**
	(0.01)	(0.04)	(0.04)
High school GPA is above median	0.51	0.46	-0.05
	(0.01)	(0.04)	(0.04)
SAT math score is above median	0.51	0.51	0.00
	(0.01)	(0.04)	(0.04)
SAT verbal score is above median	0.54	0.48	0.06
	(0.01)	(0.04)	(0.04)
Attended public high school	0.92	0.93	-0.01
	(0.01)	(0.02)	(0.02)
English is only language spoken in home	0.54	0.48	0.06
	(0.01)	(0.04)	(0.04)
No parent graduated from college	0.37	0.43	-0.05
	(0.01)	(0.04)	(0.04)
Observations	1,624	178	1,802

Note: URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian. Column 3 uses an asterisk system to denote whether the differences in means are significant. Level of significance is indicated as follows: *** p<0.01, ** p<0.05, * p<0.1. Standard deviations are in parentheses. The sample includes only those students enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB for the year 2013.

Table A3: Major Composition by Gender (%)

	Women		Men	
	Intended Major at Entry	Major at Grad.	Intended Major at Entry	Major at Grad.
Hard Science	15.07	14.29	49.37	33.77
Bio/Environ. Sci.	51.81	29.88	24.84	17.45
Social Science	3.48	19.27	2.79	15.79
Human./Arts/Interd.	3.08	36.57	2.55	32.97
Undeclared	26.55	0.00	20.45	0.01

Note: See Appendix Table A4 for the majors that fall into each major category: hard science, biology/environmental studies, social sciences, and humanities/arts/interdisciplinary. The sample includes only those students enrolled in the first quarter of General Chemistry (CHEM1A) in the fall quarter of their freshman year at UCSB between the years 1997 and 2007.

Table A4: Majors by Category

Hard Sci.	**Bio & Env. Studies**	**Social Sci.**	**Human., Arts, & Interdis.**
Biochemistry	Biochemistry-Molecular Biology	Anthropology	Art History
Chemistry	Biological Sciences	Geography	Art Studio
Chemical Engineering	Biopsychology	Geophysics	Asian & American Studies
Computer Engineering	Physiology	Physical Geography	Asian Studies
Electrical Engineering	Biology	Economics-Accounting	Black Studies
Earth Science	Cell & Develp. Biology	Economics-Mathematics	Chicana and Chicano Studies
Hydrological Sciences	Microbiology	Economics	Chinese
Mechanical Engineering	Environmental Studies	Political Science	Classics
Pharmacology		Psychology	Communication Studies
Physics		Sociology	Comparative Literature
Computer Science		Business Economics	Creative Studies
Mathematics			Dance
Financial Math & Stats			Dramatic Art
Statistics			English
Zoology			Feminist Studies
Aquatic Biology			Film & Media Studies
Ecology and Evolution			Financial Mathematics & Statistics
Computer Science			French
			Germanic Languages
			Global Studies
			History or History of Public Policy
			Interdisciplinary Studies
			Italian Cultural Studies
			Japanese
			Language, Culture & Society
			Latin Am/Iberian Studies
			Law & Society
			Linguistics
			Medieval Studies
			Middle Eastern Studies
			Music & Music Composition
			Philosophy
			Portuguese
			Religious Studies
			Slavic Languages & Literatures
			Spanish

Table A5: The Effect of the Number of On-Track Students on STEM Major Completion for Various Groups

	Late-Track	Late-Track & On-Track	Bottom 1/3	Middle 1/3	Top 1/3
	(1)	(2)	(3)	(4)	(5)
Panel A: Point Estimates					
Ln(no. of on-track)	-0.279	-0.143	-0.206	-0.135	-0.0805
	(0.190)	(0.088)	(0.152)	(0.291)	(0.133)
Ln(no. of on-track) X Fem.	-0.150*	-0.117***	-0.162***	-0.0126	-0.115*
	(0.083)	(0.031)	(0.048)	(0.076)	(0.065)
Instructor, Year, Time of day FE	X	X	X	X	X
Student Characteristics	X	X	X	X	X
Panel B: Estimated effects in %-pts. associated with a 15% increase in no. of on-track					
Women	-6.00**	-3.64***	-5.15**	-2.07	-2.74
Men	-3.91	-2.00	-2.89	-1.90	-1.13
Observations	1,935	14,165	4,770	3,389	5,556

Note: Each column is a separate specification. Controls include percent female in a class, log class size, year and instructor fixed effects, whether the lecture was held in the morning, a vector of student background characteristics, and a student's declared major at entry. Student background characteristics include: gender, race (black, Hispanic, Asian, American Indian, Filipino, Indian and white is the omitted group), if a student went to public high school, if English is the only language spoken in the home, the highest level of education attainment of the parent with the highest level of education, high school grade point average, SAT math and verbal scores and age. Clustered standard errors are in parentheses , *** p<0.01, ** p<0.05, * p<0.1. Clusters are by CHEM1A lecture (class). A 15% increase in number of on-track students in a class is the equivalent of increasing the number of on-track students by about 1 standard deviation (44 students). URM stands for underrepresented minorities and includes all race categories except white, Asian and Indian.

Appendix B

The Academic Achievement of American Indians

	Math									
	4th					8th				
Variable	**Full Sample (1)**	**White (2)**	**American Indian (3)**	**Black (4)**	**Hispanic (5)**	**Full Sample (6)**	**White (7)**	**American Indian (8)**	**Black (9)**	**Hispanic (10)**
Test Scores										
Math	0.000	0.264	-0.302	-0.464	-0.249	0.000	0.336	-0.576	-0.603	-0.375
	(1.000)	(0.708)	(0.784)	(0.721)	(0.793)	(1.000)	(0.890)	(0.986)	(0.890)	(0.948)
Race- School Identified										
White	0.599	0.964	0.401	0.029	0.268	0.620	0.988	0.263	0.034	0.204
	(0.490)	(0.187)	(0.490)	(0.168)	(0.443)	(0.485)	(0.107)	(0.440)	(0.181)	(0.403)
American Indian	0.027	0.006	0.441	0.007	0.020	0.027	0.004	0.651	0.003	0.018
	(0.163)	(0.078)	(0.497)	(0.083)	(0.140)	(0.162)	(0.064)	(0.477)	(0.059)	(0.132)
Black	0.182	0.005	0.093	0.925	0.108	0.186	0.002	0.061	0.947	0.076
	(0.386)	(0.070)	(0.290)	(0.263)	(0.310)	(0.389)	(0.048)	(0.240)	(0.223)	(0.265)
Hispanic	0.183	0.022	0.054	0.025	0.594	0.162	0.004	0.018	0.009	0.694
	(0.387)	(0.148)	(0.226)	(0.155)	(0.491)	(0.369)	(0.063)	(0.134)	(0.093)	(0.461)
Other	0.009	0.003	0.012	0.014	0.010	0.006	0.001	0.006	0.007	0.008
	(0.092)	(0.054)	(0.108)	(0.117)	(0.101)	(0.076)	(0.353)	(0.079)	(0.080)	(0.091)
Race- Student Identified										
White	0.495	1.000	0.000	0.000	0.000	0.539	1.000	0.000	0.000	0.000
	(0.500)					(0.499)				
American Indian	0.034	0.000	1.000	0.000	0.000	0.025	0.000	1.000	0.000	0.000
	(0.182)					(0.156)				
Black	0.146	0.000	0.000	1.000	0.000	0.156	0.000	0.000	1.000	0.000
	(0.353)					(0.363)				
Hispanic	0.279	0.000	0.000	0.000	1.000	0.227	0.000	0.000	0.000	1.000
	(0.448)					(0.419)				
Other	0.046	0.000	0.000	0.000	0.000	0.053	0.000	0.000	0.000	0.000
	(0.209)					(0.224)				
Other Controls										
Free/reduced lunch	0.492	0.291	0.667	0.768	0.686	0.443	0.249	0.713	0.726	0.670
	(0.500)	(0.454)	(0.471)	(0.422)	(0.464)	(0.497)	(0.433)	(0.452)	(0.446)	(0.470)
Books 0-10	0.124	0.064	0.175	0.213	0.185	0.145	0.081	0.217	0.196	0.258
	(0.330)	(0.245)	(0.380)	(0.409)	(0.388)	(0.352)	(0.273)	(0.412)	(0.397)	(0.437)
Books 11-25	0.215	0.155	0.259	0.308	0.276	0.214	0.156	0.286	0.302	0.287
	(0.411)	(0.362)	(0.438)	(0.462)	(0.447)	(0.410)	(0.363)	(0.452)	(0.459)	(0.452)
Books 26-100	0.336	0.379	0.305	0.271	0.293	0.349	0.376	0.318	0.335	0.292
	(0.472)	(0.485)	(0.460)	(0.444)	(0.455)	(0.477)	(0.484)	(0.466)	(0.472)	(0.455)
Books 100 plus	0.325	0.402	0.262	0.209	0.247	0.293	0.387	0.179	0.166	0.163
	(0.468)	(0.490)	(0.440)	(0.406)	(0.431)	(0.455)	(0.487)	(0.383)	(0.372)	(0.369)
Male	0.508	0.512	0.516	0.506	0.505	0.502	0.500	0.521	0.504	0.509
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	0.500
Disabled	0.119	0.123	0.139	0.116	0.111	0.102	0.095	0.157	0.112	0.106
	(0.324)	(0.329)	(0.346)	(0.321)	(0.315)	(0.303)	(0.294)	(0.364)	(0.315)	(0.307)
English language learner	0.083	0.014	0.098	0.027	0.246	0.044	0.004	0.127	0.012	0.159
	(0.277)	(0.117)	(0.297)	(0.162)	(0.431)	(0.205)	(0.063)	(0.333)	(0.111)	(0.365)
Age	7.512	7.505	7.562	7.509	7.518	11.435	11.420	11.582	11.472	11.432
	(1.161)	(1.141)	(1.159)	(1.192)	(1.184)	(1.162)	(1.148)	(1.174)	(1.201)	(1.167)
Mother-H.S drop out	-	-	-	-	-	0.104	0.061	0.134	0.085	0.217
						(0.305)	(0.239)	(0.340)	(0.279)	(0.412)
Mother- H.S. grad	-	-	-	-	-	0.208	0.207	0.233	0.237	0.189
						(0.406)	(0.405)	(0.423)	(0.425)	(0.391)
Mother- some college	-	-	-	-	-	0.173	0.182	0.171	0.189	0.136
						(0.379)	(0.386)	(0.377)	(0.391)	(0.343)
Mother- college grad	-	-	-	-	-	0.364	0.445	0.255	0.331	0.207
						(0.481)	(0.497)	(0.436)	(0.471)	(0.405)
Mother- don't know	-	-	-	-	-	0.151	0.105	0.207	0.158	0.251
						(0.358)	(0.307)	(0.405)	(0.365)	(0.433)
Father- H.S. drop out	-	-	-	-	-	0.108	0.077	0.132	0.084	0.198
						(0.310)	(0.267)	(0.339)	(0.277)	(0.399)
Father- H.S. grad	-	-	-	-	-	0.207	0.215	0.239	0.223	0.174
						(0.405)	(0.411)	(0.427)	(0.416)	(0.379)
Father- some college	-	-	-	-	-	0.136	0.151	0.133	0.127	0.104
						(0.343)	(0.358)	(0.339)	(0.333)	(0.305)
Father- college grad	-	-	-	-	-	0.306	0.394	0.182	0.230	0.171
						(0.461)	(0.489)	(0.386)	(0.421)	(0.376)
Father- don't know	-	-	-	-	-	0.242	0.162	0.314	0.337	0.354
						(0.428)	(0.368)	(0.464)	(0.473)	(0.478)
Observations	322,939	159,877	11,131	47,152	89,969	278,074	149,765	6,969	43,478	63,084

Notes: Means of student characeristics are based on math sample. Means for reading sample are nearly identical, and are available on request.

Standard Deviations in parentheses. Parental education not available for 4th grade.

Table A1: Summary Statistics by Race: Means of Student Characteristics by Student Reported Race

	Math				Reading			
	4th		8th		4th		8th	
	10th	90th	10th	90th	10th	90th	10th	90th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
American Indian								
School Identified Only	-0.231***	-0.0966**	-0.282***	-0.105*	-0.314***	-0.0053	-0.175**	0.122*
	(0.057)	(0.033)	(0.046)	(0.050)	(0.063)	(0.080)	(0.059)	(0.057)
Self Identified Only	-0.189***	-0.127***	-0.317***	-0.267***	-0.124**	-0.103***	-0.335***	-0.158***
	(0.030)	(0.020)	(0.046)	(0.040)	(0.041)	(0.028)	(0.048)	(0.040)
School and Self Identifie	-0.346***	-0.158***	-0.591***	-0.392***	-0.309***	-0.139**	-0.401***	-0.106*
	(0.057)	(0.030)	(0.055)	(0.042)	(0.053)	(0.041)	(0.067)	(0.051)
Black								
School Identified Only	-0.448***	-0.353***	-0.411***	-0.369***	-0.382***	-0.307***	-0.285***	-0.231***
	(0.023)	(0.017)	(0.027)	(0.023)	(0.024)	(0.021)	(0.027)	(0.026)
Self Identified Only	-0.284***	-0.234***	-0.400***	-0.364***	-0.307***	-0.227***	-0.417***	-0.262***
	(0.040)	(0.030)	(0.056)	(0.045)	(0.051)	(0.046)	(0.049)	(0.034)
School and Self Identifie	-0.466***	-0.410***	-0.645***	-0.556***	-0.358***	-0.323***	-0.433***	-0.365***
	(0.015)	(0.011)	(0.015)	(0.014)	(0.022)	(0.014)	(0.015)	(0.014)
Hispanic								
School Identified Only	-0.260***	-0.266***	-0.297***	-0.107	-0.287***	-0.288***	-0.427***	-0.203***
	(0.034)	(0.025)	(0.078)	(0.065)	(0.039)	(0.033)	(0.102)	(0.058)
Self Identified Only	-0.172***	-0.141***	-0.365***	-0.243***	-0.223***	-0.180***	-0.388***	-0.227***
	(0.011)	(0.011)	(0.021)	(0.021)	(0.018)	(0.012)	(0.016)	(0.020)
School and Self Identifie	-0.178***	-0.155***	-0.331***	-0.229***	-0.130***	-0.102***	-0.168***	-0.118***
	(0.017)	(0.012)	(0.019)	(0.018)	(0.019)	(0.022)	(0.018)	(0.020)
Observations	322,992	322,992	278,096	278,096	325,780	325,780	284,175	284,175

Notes: Racial subgroups also include a dummy variable for other. Additionally, the specification includes family/individual controls (see Table 2) and school fixed effects. The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table A2: Standardized Test Score Gaps in Standard Deviation Units at 10-percentile and 90-percentile

	Math		Reading	
	4th	8th	4th	8th
	(1)	(2)	(3)	(4)
Panel A: American Indian-Northeast				
School Identified Only	-0.113	-0.119	-0.155	0.057
	(0.082)	(0.144)	(0.088)	(0.114)
Self Identified Only	-0.103**	-0.251***	-0.125**	-0.192**
	(0.038)	(0.070)	(0.052)	(0.067)
School and Self Identified	-0.346	-0.730***	-0.152	-0.466
	(0.245)	(0.222)	(0.250)	(0.300)
Observations	54,296	47,990	54,989	49,149
AI Observations*	990	602	1,039	619
Panel C: American Indian-South				
School Identified Only	-0.099***	-0.232***	-0.124***	-0.039
	(0.031)	(0.041)	(0.041)	(0.051)
Self Identified Only	-0.118***	-0.262***	-0.119***	-0.168***
	(0.024)	(0.048)	(0.028)	(0.052)
School and Self Identified	-0.117***	-0.417***	-0.087*	-0.102*
	(0.034)	(0.046)	(0.045)	(0.048)
Observations	119,018	100,655	118,667	102,774
AI Observations*	4,539	2,628	4,604	2,769
Panel B: American Indian-Midwest				
School Identified Only	-0.147***	-0.121**	-0.239***	0.041
	(0.049)	(0.055)	(0.058)	(0.066)
Self Identified Only	-0.120***	-0.283***	-0.118***	-0.297***
	(0.026)	(0.043)	(0.036)	(0.047)
School and Self Identified	-0.239***	-0.402***	-0.320***	-0.202***
	(0.045)	(0.072)	(0.057)	(0.066)
Observations	74,336	66,699	75,954	67,974
AI Observations*	3,872	2,775	3,973	2,887
Panel D: American Indian-West				
School Identified Only	-0.276***	-0.189**	-0.125	-0.059
	(0.043)	(0.058)	(0.071)	(0.054)
Self Identified Only	-0.272***	-0.340***	-0.097**	-0.356***
	(0.036)	(0.070)	(0.033)	(0.071)
School and Self Identified	-0.344***	-0.580***	-0.283***	-0.380***
	(0.045)	(0.046)	(0.051)	(0.063)
Observations	75,342	62,752	76,170	64,278
AI Observations*	6,300	4,488	6,421	4,746

*AI observations includes school identified American Indians, self identified American Indians and those that are school and self identified as American Indian.

Notes: Controls include all racial dummy variables, family/individual controls (see Table 2) and school fixed effects. The level of significance is indicated as follows: * p< 0.10, ** p<0.05, *** p<0.01.

Table A3: Standardized Test Score Gaps in Standard Deviation Units by Census Region