DISSERTATION

ESSAYS ON ECONOMICS OF EDUCATION

Submitted by

Kenese Io

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Anita Pena
Co-Advisor: Anders Fremstad

Ray Miller
Tobin Lopes

ABSTRACT

ESSAYS ON ECONOMICS OF EDUCATION

Chapter one analyzes the opt-out movement in Colorado and New York. In 2015, Congress passed the Every Student Succeeds Act (ESSA) and reinforced the focus on educational equity through the mechanism of standardized tests. The ESSA maintained a 95% participation requirement for grades 3-8 English and Language Arts (ELA) and Math state assessments. I utilize state education data from Colorado and New York to identify how standardized test protests, which are now referred to as the opt-out movement, impact the participation rates in both states. I employ fixed effects regressions to assess the participation rates before and after the protests by interacting the opt-out movement with racial composition, region, and free and reduced lunch status and find that White students are primary participants in the movement in both states. I provide visual estimates of the fixed effect regressions to demonstrate the decline in participation rates with time-varying controls. The decline in participation rates is persistent through 2018 in New York but trends back to pre opt-out levels in Colorado. I find a positive relationship between participation rates and performance in both states but this relationship is dampened after the opt-out protests. Finally, I calculate a counterfactual for school level performance to assess the relationship if schools maintained their pre opt-out levels. Results indicate that if policy makers use raw data to assess achievement gaps they could underestimate achievement gaps.

The second chapter utilizes regional codes from the National Center of Education Statistics (NCES), this study compiles district and school level data to provide additional insight on the relationship between learning modes and performance. Jack et al. (30) estimated the impact of different instruction modes during the 2020-21 academic year

on standardized test performance in 11 states and find that districts with full in-person learning experienced significantly smaller declines in pass rates. Colorado experienced a smaller performance decline relative to other states in the sample and appeared to be an outlier in their study. I use District-level data from their study to show a full transition to in-person learning would have reduced learning loss by 3-6 percentage points in Colorado. School-level analysis in Colorado indicates that the reduction in learning loss attributed to full in-person instruction is small and largely statistically insignificant apart from a few grades in Math. Analysis by racial subgroup indicates that increasing participation rates for minority students would positively impact performance.

Finally, the third chapter uses administrative data from Colorado State University's Institutional Research, Planning, and Effectiveness (IRPE) and Student Athlete Support Services to identify the relationship between support services and student athletes' semester GPA and credit earned ratio. I analyze the relationship for four types of support services at the extensive (meetings versus no meetings) and intensive margins (number of meetings). I find that the relationship between support services and GPA varies based on the nature of the support service and the time of reception. In semesters that student athletes receive intensive support services, like tutoring, they earn lower GPAs. Student athletes who receive less intensive support services, like mentoring, earn higher semester GPAs. I find that support services and student athletes credit earned ratio exhibit no statistically significant relationship across all specifications. Negative selection is present in the sample because student athletes who are academically unprepared are more likely to receive support services. More intensive support services like tutoring highlight this selection in the result. I attempt to isolate the effect of support services by using the first support service session provided by SASS. I differentiate between early intervention and general support services and find that the benefit to student athletes from support services comes from receiving services in the first four weeks of the semester.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Standardized Test Opt-Out: Evidence from Schools in Colorado and New York

1.1 Introduction

Standardized testing serves a fundamental role in measuring the quality of schools and is commonly used to evaluate teacher quality in the United States. The No Child Left Behind Act (NCLB) set the stage for continued expansion of standardized testing to ensure that schools are held accountable for their performance. The NCLB mandates that schools meet 95% participation rates on mandated tests. The strict 95% participation rate lower bound is set to ensure that the performance statistics derived from the standardized tests act as a representative sample of the student body. Prior to 2015 schools achieved and exceeded this threshold (11). The 2015 school year marks a sharp drop in the participation rates on state mandated tests in both Colorado and New York. The drop in participation rates reflects opt-out movements which consist of students and parents refusing to participate in state mandated tests as a form of protest.

In 2015, the Obama administration passed the Every Student Succeeds Act (ESSA) which returned accountability decisions to the state and local level and provided greater flexibility with performance level indicators and school improvement interventions. The ESSA maintained the essence of the NCLB with a focus on the idea of educational equity through assessment, requiring the reporting of student subgroup performance on two new student subgroups: gender and migrant status (National Conference of State Legislatures, 2015). The 95% participation rate at the school level is crucial for identifying low performing schools (40; 6). The 95% participation rate is also an important tool for identifying achievement gaps across racial groups. Failure to meet the ESSA requirements means that policymakers are faced with incomplete data. Incomplete data at the school level raises concerns for resource allocation, as underestimation could lead to low performing schools not receiving the necessary amount of assistance. These concerns are especially relevant for the goal of equitable education. Policymakers could draw false conclusions for schools with low participation rates by overestimating how much a school is closing the achievement gap between white and minority students. Alternatively, it could reduce policymakers' ability to identify the extent to which achievement gaps are widening.

Many Americans were aware of opt-out as it related to standardized tests and specifically the Common Core (47). NPR and the popular education news outlet, Chalkbeat, highlighted the movement in New York and Colorado as the epicenters of the opt-out movement (Kamenetz, 2015; Gorski, 2015). The opt-out movement is a demonstration of collective action. The implementation of the ESSA and the adoption of the Common Core standards increased testing requirements. These introductions coincided with a new teacher evaluation system which requires using student growth measures for 50% of the teachers' total score in both Colorado and New York (9; 31). The year 2015 marks the introduction of major educational policy changes that inspire collective action on behalf of parents and students. Some parents and students argue that high-stakes testing induces testing-fatigue, increased weight allocated to standardized test performance used in teacher evaluations, and a narrowed curriculum that focuses the material to "teach to the test" as enough of a reason to opt-out of testing. The Colorado State Board of Education requested waivers to allow schools and districts to fall below the 95% participation requirement, but the United States Department of Education rejected Colorado's request and stated districts would not be exempt from participation requirements (18). In 2016, the Department of Education issued the following proposed regulation:

> Failure to meet the 95 percent participation rate requirement is factored in the State's accountability system in a meaningful, publicly visible manner through a significant impact on a school's performance level or summative

rating, identification for targeted support and improvement, or another equally rigorous, State determined action, thus providing an incentive for the school to ensure that all students participate in annual State assessments (51).

In Colorado, schools who failed to meet the 95% requirement received a drop in their school rating meaning they could be "Performance Plan" schools in 2014 and "Improvement Plan" in 2016 and they would continue to drop in classification for each year they failed to meet 95%. The CDE reports in their School Performance Rating whether schools drop in classification is due to low participation after 2016. Some schools have maintained lower participation rates and received multiple drops in classification but parents can readily check whether this is due to low participation or poor performance. In New York, schools fell off the Reward School list and were not eligible for certain grants because of their low participation rate.

This paper identifies the relationship between demographic characteristics and participation rates at the school level for both states, while building off the analysis in Clayton et al. (12) and Chingos (11). I employ fixed effect regressions to demonstrate the "pre" and "post" opt-out movement participation rates and include interaction terms to highlight variation in regional outcomes and schools with different racial compositions. I then present estimates from a two way fixed effect model to compare the resilience of opt-out movement across states. To conclude, I present fixed effect estimates that identify the relationship between participation rates and school level performance. The paper's findings can be summarized as follows: (1) suburban schools in Colorado and New York saw the largest declines in participation rates; (2) White students opted-out at higher rates relative to their minority peers and the results hold following the inclusion of school and year fixed effects; (3) opt-out protests persisted throughout the sample in the state of New York but participation rates began to return to pre opt-out movement levels in the state of Colorado; (4) at the school level, the positive relationship between participation rates and proficiency is dampened in the post opt-out period. Counterfactual calculations indicate that schools could increase proficiency if they maintained their pre opt-out participation levels.

The opt-out movement in New York has been effective in that it validated the proponent's framing. The persistence of low participation rates in the state renders much of the performance data useless at the schools that experience high opt-out rates. The framework in which the proponent's categorize test scores as invalid measures of performance is somewhat self-perpetuating. With incomplete data school administrators cannot draw any robust conclusions with respect to achievement gaps in performance across income and racial groups. Twelve civil and human rights groups announced "the educational outcomes for the children we represent are unacceptable by almost every measurement. And we rely on the consistent, accurate, and reliable data provided by annual statewide assessments to advocate for better lives and outcomes for our children" (50). State tests are imperfect in their ability to assess learning, however, civil rights groups recognize consistent collection of performance data is imperative because state tests remain one of the only objective metrics to assess educational equity.

1.1.1 The Opt-Out Movement

The national opt-out survey highlighted dissatisfaction with the status quo. The frustration with state mandated tests illustrated by survey participants provided the building blocks for a protest (47).

Clayton et al. (12) utilize the framework of cultural capital developed by Bourdieu (7) to motivate the demographic characteristics of the opt-out movement in Colorado. Cultural capital develops over time and is present in knowledge about school, appropriate attitudes and beliefs, personal style, and linguistic competence sanctified by the dominant culture (37; 38). Cultural capital increases with socioeconomic class. Individuals with high levels of cultural capital are more likely to have access to information pertaining to changes in school policy. Bourdieu (8) further emphasizes the role of class in the use of cultural capital.

Individuals who preside in privileged classes can better exercise their cultural capital. In the context of schools, privileged parents interact with school administrators and educators differently by using their increased resources and often feel more comfortable with asking for exceptions (38).

Wang (52) analyzed the New York opt-out movement by using 221 press articles to apply a social network analysis of the opt-out movement that identified coalitions. Wang (52) finds that the opt-out coalition emphasizes increased anxiety on behalf of students, test scores as invalid measures of performance, and a dissatisfaction with the link between test scores and teacher evaluations. The pro-testing coalition expresses the importance of utilizing the test score data to evaluate achievement gaps across race, gender, and ability. Opponents of the opt-out movement frame the protest as "White, affluent families' irresponsible behavior" which could potentially harm minority students (52).

Bennett (5) utilizes a synthesis of news outlets, surveys, and contemporary studies to understand two of the mechanisms behind opt-out movements in these states. The first mechanism deals with the time allocated to the standardized tests; this mechanism is less convincing. In 2015, the Obama administration called for a 2% cap on the percentage of instructional time devoted to state-mandated tests (51). The second mechanism is derived from the link between standardized tests and teacher evaluation. Parents of students who do not participate in the mandated tests report disagreement with the increasing prevalence of mandated tests. They also disagree with the use of mandated tests to evaluate teacher quality. During the 2015 year, the test was a Common-Core-aligned assessment on which notably lower percentages of students were expected to achieve proficiency. The increase in difficulty coupled with the use of test performance to evaluate teachers served to further motivate parents having their children opt-out of the mandated tests as a form of protest to the use of mandated tests to evaluate teachers.

Chingos (11) explores the New York district participation rates in 2015 using cross-sectional analysis. The article uses district-level data and finds a positive relationship between wealth and test scores and opting-out. After controlling for free/reduced lunch, however, the relationship changed direction, showing that districts with lower scores had higher levels of opting-out (11).

Clayton et al. (12) analyze the opt-out movement in the context of Colorado. They identify demographic characteristics of the schools that have low participation rates. They utilize statewide panel data from 2012-2016 and find that the rate for students who opt-out is largest in suburban, rural, and high performing schools. They also explore participation rates by race and find that schools with higher proportions of white students have higher opt-out rates. This is consistent with Bennett (5) and Pizmony-Levy and Green Saraisky (48) which find that White, affluent, educated parents may be at the forefront of the opt-out movement, opposed to high levels of testing, and opposed to high stakes testing for teacher evaluation.

Clayton et al. (12) note that selectively encouraging a few students to opt-out could achieve similar aims as more complicated cheating programs. Traditional cheating involves artificially increasing scores but opting out involves additional concerns regarding discrimination if certain demographic groups are perceived to not have the capacity to meet test requirements (24; 3). Researchers using Ohio data demonstrated that school rankings could be swayed by a change in the participation of as few as 11 students in a single school (4).

The opt-out movement is important to analyze because state assessments are the only comparable measures of performance at the building level. The National Assessment of Educational Progress (NAEP), does not report at that level, nor is it aligned with state content standards. State assessments are the only measures of school level performance by demographic group. Opt-out creates distortion in performance preventing parents, educators, policymakers, and the public from understanding the extent to which schools are effectively educating all children (5). When schools fail to meet the 95% participation rate requirement, policymakers lose the ability to accurately assess achievement gaps and this could result in misallocation of resources (40; 6). This misallocation of resources means that policymakers cannot accurately assess the efficacy of their school or districts human capital investment. The returns on human capital investment in primary education (whether social or private) are the highest among all educational levels (54). Opt-out reduces policymakers' ability to identify the sufficient level of investment.

1.2 Data and Methods

1.2.1 Participation Data

The data sets are constructed separately from publicly available data provided by the Colorado Department of Education (CDE) and the New York State Education Department (NYSED). The data sets use school level data disaggregated by race. The participation rate variable is provided by the CDE and NYSED. Participation rates at the school level are separately merged with regional codes from the National Center for Education Statistics (NCES). The sample omits all online schools for both states and contains schools that serve grades 3-8. Analysis is run using the state mandated participation at the grade level and as a result the number of observations will vary depending on the grades that are served by each school. I focus the analysis on grades 3-8 across both states for ELA assessments. The panel begins in 2010 for both states and ends in 2018 for Colorado and New York. This sample period provides multiple years before and after the opt-out movement to analyze the opt-out movement. I omit the most recent years to avoid the effect of the pandemic. The students take state mandated assessments in the spring term and so including the 2019 school year would include the pandemic. The pandemic reflects a sharp decline in participation rates in the 2020 school year, however, this data is omitted because the decline in participation rates is not only due to the opt-out movement but also due to test administration differences in the context of the pandemic.

1.2.2 Performance Data

Performance data is gathered at the school level. The CDE and NYSED provide the percentage of students who test into respective categories. New York performance data is consistent in classification across the sample. The state categorizes performance into four categories with L1 representing unsatisfactory performance and L4 representing advanced performance. Students who test into L3 are considered proficient. Colorado's scoring categories changed simultaneously with the introduction of the protests as Colorado transitioned from the Colorado Student Assessment Program (CSAP)/Transitional Colorado Assessment Program (TCAP) to the Colorado Measures of Academic Success (CMAS) assessment (Colorado Department of Education, 2015). Before 2015, Colorado used a four category system like New York and following 2015 the state expanded to 5 categories and students who test into categories 4 and 5 are proficient and categories 1, 2, and 3 classify students as below or approaching proficiency. This is the same classification used in New York for students who test into L2 and L1 according to the NYSED and CDE. Throughout the performance analysis I simplify the categorization to proficient or not proficient by combining categories L3 and L4 in New York and categories 4 and 5 in Colorado into the proficient category and L1 and L2 in New York and categories 1-3 into not proficient in Colorado. State tests are ever changing with respect to content and this sample is not free of these changes. Across both states the CDE and NYSED report an increase in the difficulty of their state tests in the 2015 school year. I utilize a two-way fixed model in the performance analysis section to capture the change in test difficulty in the year fixed effects.

1.2.3 Participation Methods

In equation (1.1) I use pooled Ordinary Least Squares (OLS) to look at the relationship between participation rates by region. The outcome variable is participation rates in year t for school i in grade g. $Region_{is}$ is a categorical variable where city is the reference group and Opt_t is an indicator that takes on a value of 1 starting in 2015 for the opt-out protest. I utilize robust standard errors clustered at the school level represented by ϵ_{it}. I report the estimates for New York and Colorado in Table 1. The variable of interest is the interaction between region and the opt-out indicator.

$$Y_{igt} = \beta_0 + \beta_1 region_i * Opt_t + \beta_2 region_i + \beta_3 Opt_t + \epsilon_{it}. \tag{1.1}$$

I extend the demographic characteristic analysis of Clayton et al. (12) with the inclusion of school fixed effects. I estimate the following:

$$Y_{it} = \beta_1 X_{it} * Opt_t + \beta_2 X_{it} + \beta_3 Opt_t + \lambda_i + \varepsilon_{it}. \tag{1.2}$$

Equation (1.2) uses a school fixed effect model. School and grade are grouped in i. The outcome is the participation rate of specific grades in a school in year t. The coefficient of interest here is β_1 which is the interaction between the school level racial composition and the opt-out indicator Opt_t. The racial composition of the school is a continuous variable that measures the percentage of each racial group within a school. The estimates are presented with the percentage of White students as the reference group. The model includes school fixed effects with λ_i to capture unobserved time-invariant school characteristics that may affect participation rates. The opt-out indicator Opt_t turns on for all years after the opt-out movement in 2015. I use fixed effect estimation because the unobserved heterogeneity is assumed to be constant. I use robust standard errors clustered at the school level in ε_{it}. The standard errors are clustered to account for heteroskedasticity. Independent variables are not random and the model exhibits no multicollinearity.

$$Y_{it} = \beta_1 X_{it} + \lambda_i + \mu_t + \varepsilon_{it}. \tag{1.3}$$

In addition to the school fixed effects, Equation (1.3) estimates a two-way fixed model where I include μ_t which denotes the year fixed effects to control for time-varying shocks that may occur over the sample period, for example, potential state legislation that could impact all schools across the state. Both states in the sample are diverse in geography, racial composition, and economic composition. The addition of school and year fixed effects helps to control for some of the variation to school population caused by economic shocks, migration, and new state legislation that is not captured in Clayton et al. (12)'s pooled OLS analysis of the demographic characteristics.

1.2.4 Performance Methods

These concerns motivate the following two-way fixed effect model:

$$Y_{it} = \beta_1 Participation_{it} + \beta_2 (Participation_{it} * Opt_t) + \beta_3 X_{it} + \lambda_i + \mu_t + \varepsilon_{it}. \tag{1.4}$$

The outcome variable is the percentage of students at school i in year t that test into the proficient category. The coefficients of interest are β_1 and β_2 which capture the relationship between participation rates and proficiency.

1.2.5 Counterfactual Calculation

The coefficients estimated in equation (1.4) indicate a dampening positive relationship between participation rates and school level proficiency but the positive relationship is not completely eliminated. This allows me to calculate the relationship between participation rates if they had remained at their pre opt-out levels. I first calculate each states average participation rate from 2010-2014 and use this as their baseline. I then use a linear combination of the coefficients for the $Participation_{it}$ and $Participation_{it} * Opt_t$ in specification (1.4) and multiply this by the difference between average baseline participation rate and the participation rate for each post opt-out year. I use a stylized example to demonstrate the calculation:

$$(\beta_1 + \beta_2)(Average\ \ Participation_{2010-2014} - Participation_t) = CF \tag{1.5}$$

The counterfactual then demonstrates the percentage point increase in proficiency that would have occurred given participation rates remained at their average pre opt-out levels. I report 95% confidence intervals and use them to provide upper and lower bounds of the potential increase in proficiency.

1.3 Results

1.3.1 Participation Results

Table 1.1 presents the results of specification (1.1) and demonstrates the strength of the protest across grade levels in suburban schools in New York and Colorado. This table reports the interaction results for suburban and rural schools relative to the reference group, city schools. Relative to city schools, rural and suburban schools participated at higher rates for all grades in Colorado and grades 3-5 in New York before 2015. The interaction terms indicate that this relationship flips in 2015 with suburban and rural schools consistently decreasing relative to city schools for all grade levels. The fact that this relationship also holds for all grades in the analysis suggests that the coalition ties in New York across proponents for opt-outs play a significant role in inciting a school wide effect (52). In New York the interaction between suburban schools and the opt-out movement indicator shows a decline in participation rates of 17-28 percentage points relative to the city schools. The relationship strengthens as the grade level increases. In Colorado, the interaction with suburban schools shows a 6 percentage point decline in participation rates relative to city schools in the 8th grade, indicating that opt-out movement in Colorado may be focused in older students. The decline in participation rates for suburban schools is relatively small for elementary grades but it increases as the grade level increases. In the state of New York, the opt-out movement seems to be a school wide phenomenon that is prevalent in both suburban and rural schools but the decline is most prominent in suburban schools.

Table 1.1: Effect of Opt-Out Movement on Test Participation Rates

NY	3rd	4th	5th	6th	7th	8th
Opt-Out	-6.641***	-6.483***	-6.559***	-4.400***	-4.337***	-4.288***
	(1.214)	(1.127)	(1.199)	(0.768)	(0.900)	(0.895)
Opt-out*Rural	-9.572***	-10.047***	-10.601***	-15.960***	-18.426***	-18.458***
	(1.341)	(1.269)	(1.347)	(1.067)	(1.223)	(1.216)
Opt-out*Suburb	-17.841***	-18.055***	-19.165***	-24.459***	-28.536***	-28.682***
	(1.388)	(1.322)	(1.393)	(1.275)	(1.509)	(1.509)
Rural	0.252*	0.274*	0.253*	0.128	0.198*	0.258**
	(0.140)	(0.140)	(0.133)	(0.089)	(0.108)	(0.110)
Suburb	0.100	0.091	0.037	-0.164*	-0.152	-0.141
	(0.140)	(0.141)	(0.133)	(0.098)	(0.123)	(0.127)
Observations	12597	12319	11841	8683	7149	7101
CO	3rd	4th	5th	6th	7th	8th
Opt-Out	-2.524***	-2.292***	-2.619***	-3.175***	-4.392***	-6.111***
	(0.240)	(0.254)	(0.303)	(0.416)	(0.560)	(0.815)
Opt-out*Rural	-1.390***	-2.251***	-2.551***	-3.710***	-4.078***	-4.420***
	(0.424)	(0.461)	(0.529)	(0.787)	(0.987)	(1.222)
Opt-out*Suburb	-0.812**	-1.603***	-2.021***	-3.449***	-5.268***	-6.791***
	(0.319)	(0.359)	(0.419)	(0.632)	(0.926)	(1.246)
Rural	0.039	0.192***	0.240**	0.135	0.222**	0.171
	(0.073)	(0.052)	(0.111)	(0.117)	(0.108)	(0.147)
Suburb	0.094**	0.146***	0.254**	0.036	0.045	-0.305
	(0.039)	(0.049)	(0.104)	(0.128)	(0.122)	(0.252)
Observations	9033	9037	8961	5358	4129	4053

* Results from specification (1.1). Robust Standard errors clustered at the school level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Schools grouped into 3 regions. Regions include Rural, Suburban, and reference group is City schools. Interaction results report pooled OLS estimates relative to the reference group, City schools.

The racial composition estimates from specification (1.2) are reported in Tables 1.2 and 1.3 for grades 3-8 state mandated ELA tests for New York and Colorado. White students are the reference group and the first coefficient depicts the scenario in which a school is composed of 100% White students. Estimates indicate that a 1 percentage point increase in Black students in the opt-out years corresponds with an increase in participation rates ranging from 0.24-0.31 percentage points in New York and a 0.05-0.27 percentage point increase in Colorado. In New York, the first column for grade 3 indicates that if a school were to transition from 100% White to 100% Black this would eliminate the decline in participation rates entirely. Estimates for both states show that changing a school's racial composition to include a higher percentage of black students will increase participation rates. Prior to the opt-out period, increases in the percentage of Black students correspond with lower participation rates and the relationship is largely insignificant, this relationship flips for both states following the introduction of the movement. The interaction between opt-out and the percentage of Hispanic students shows that states vary with respect to increasing the percentage of Hispanic students. In New York, increases in the percentage of Hispanic students consistently correspond with a decline in participation rates for lower grade levels and the relationship flips in 6th grade. In New York, however, the pre opt-out period indicates that increasing the percentage of Hispanic students corresponds with a decline in participation rates and the magnitude of the coefficient falls in the opt-out period signaling that Hispanic students are not the primary movers in the protest. In Colorado, increases in the percentage of Hispanic students in the post op-out period correspond with higher participation rates for all grades. Across both states the consistent result is that schools with larger proportions of Black students showed positive and significant increases in participation during the opt-out time period and this seems to increase with grade as well.

In order to visualize the fixed effects estimates I treat the opt-out movement as an event. I use year fixed effects from 2010 for New York and Colorado up until 2013 and omit

Table 1.2: Impact of a Change in Racial Composition on Test Participation in New York

New York Grades	3rd	4th	5th	6th	7th	8th
Opt-out	-24.729***	-24.565***	-25.230***	-28.468***	-29.368***	-29.519***
	(0.815)	(0.818)	(0.836)	(1.033)	(1.168)	(1.172)
Opt-out*%Black	0.247***	0.250***	0.274***	0.304***	0.304***	0.310***
	(0.018)	(0.017)	(0.019)	(0.020)	(0.025)	(0.025)
Opt-out*%Hisp	0.053**	0.055**	0.065**	0.153***	0.166***	0.162***
	(0.024)	(0.024)	(0.025)	(0.029)	(0.031)	(0.032)
Opt-out*%Asian	0.084*	0.088*	0.093*	0.169***	0.196***	0.197***
	(0.048)	(0.048)	(0.049)	(0.052)	(0.068)	(0.068)
Opt-out*%Native	-0.152	-0.137	-0.097	-0.224	-0.342*	-0.156
	(0.099)	(0.099)	(0.099)	(0.153)	(0.198)	(0.227)
Opt-out*%Multi	0.880***	0.820***	0.730***	0.624***	0.349	0.340
	(0.127)	(0.129)	(0.134)	(0.166)	(0.262)	(0.261)
%Black	-0.654***	-0.670***	-0.629***	-0.558***	-0.634***	-0.633***
	(0.078)	(0.080)	(0.086)	(0.112)	(0.135)	(0.136)
%Hispanic	-0.825***	-0.880***	-0.939***	-1.020***	-1.132***	-1.139***
	(0.069)	(0.070)	(0.075)	(0.105)	(0.129)	(0.130)
%Asian	-0.170	-0.168	-0.166	-0.468***	-0.534***	-0.589***
	(0.116)	(0.117)	(0.121)	(0.130)	(0.156)	(0.156)
%Native	-0.197	-0.185	0.061	0.456	0.542*	0.535*
	(0.278)	(0.277)	(0.299)	(0.282)	(0.328)	(0.315)
%Multiracial	-0.313***	-0.313***	-0.289**	-0.163***	-0.159***	-0.156***
	(0.114)	(0.117)	(0.113)	(0.051)	(0.051)	(0.051)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11811	11573	11072	8012	6582	6534

* Results from specification (1.2). Robust Standard errors clustered at the school level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are bound between 0-100 as percentages. Opt-out indicates a 100% White school and the interaction terms indicate the effect of a 1 percentage point increase in minority post opt-out. Data is masked to protect the identity of students if the number of participants is less than 16.

the penultimate year 2014 to develop a baseline for the pre opt-out movement period. I plot the year fixed effects from specification (1.3). Visual estimates with year fixed effects leading up until the opt-out movement show little variation in participation rates prior to the opt-out. I then use the year fixed effects for the post opt-out period from 2015-2018 for both states. The year fixed effects after the opt-out movement allows for inspection of the temporal nature of the opt-out movement. I also include time-varying controls for racial composition and free and reduced lunch status. This estimation procedure produces a visual illustration of the resilience of the protests in each state. The error term is clustered at the school level.

Table 1.3: Impact of a Change in Racial Composition on Test Participation in Colorado

Colorado Grades	3rd	4th	5th	6th	7th	8th
Opt-out	-5.739***	-7.387***	-7.870***	-10.147***	-13.743***	-17.322***
	(0.621)	(0.715)	(0.733)	(1.126)	(1.494)	(1.679)
Opt-out*%Black	0.051***	0.070***	0.093***	0.137***	0.192***	0.272***
	(0.012)	(0.012)	(0.015)	(0.030)	(0.039)	(0.048)
Opt-out*%Hisp	0.051***	0.081***	0.091***	0.112***	0.148***	0.186***
	(0.008)	(0.009)	(0.009)	(0.014)	(0.019)	(0.022)
Opt-out*%Asian	0.028	0.022	-0.007	-0.014	-0.022	-0.177
	(0.036)	(0.045)	(0.051)	(0.085)	(0.115)	(0.156)
Opt-out*%Native	0.014	0.014	0.035	-0.142	-0.034	-0.052
	(0.062)	(0.054)	(0.063)	(0.149)	(0.160)	(0.223)
Opt-out*%Multi	0.074	0.108	0.031	-0.011	-0.123	-0.177
	(0.065)	(0.076)	(0.083)	(0.147)	(0.205)	(0.242)
%Black	0.010	0.022	-0.060*	0.017	-0.053	-0.120
	(0.026)	(0.025)	(0.032)	(0.059)	(0.084)	(0.094)
%Hispanic	0.013	0.002	-0.025	0.001	-0.013	0.021
	(0.016)	(0.018)	(0.019)	(0.032)	(0.039)	(0.047)
%Asian	0.045	0.090	0.009	-0.027	0.015	-0.016
	(0.055)	(0.070)	(0.071)	(0.088)	(0.103)	(0.145)
%Native	-0.050	-0.125	-0.070	-0.099	-0.337	-0.161
	(0.089)	(0.095)	(0.094)	(0.212)	(0.266)	(0.269)
%Multiracial	0.041*	0.072***	0.018	0.059	0.141*	0.098
	(0.022)	(0.026)	(0.030)	(0.061)	(0.081)	(0.098)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9044	9046	8969	5376	4148	4072

* Results from specification (1.2). Robust Standard errors clustered at the school level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures 1.1 and 1.2 illustrate the estimates for New York and Colorado respectively. I include all grades 3-8 in the model and include school and year fixed effects as well as controls for time-varying school characteristics like racial composition and free and reduced lunch percentage. Both states exhibit little variation before the introduction of the

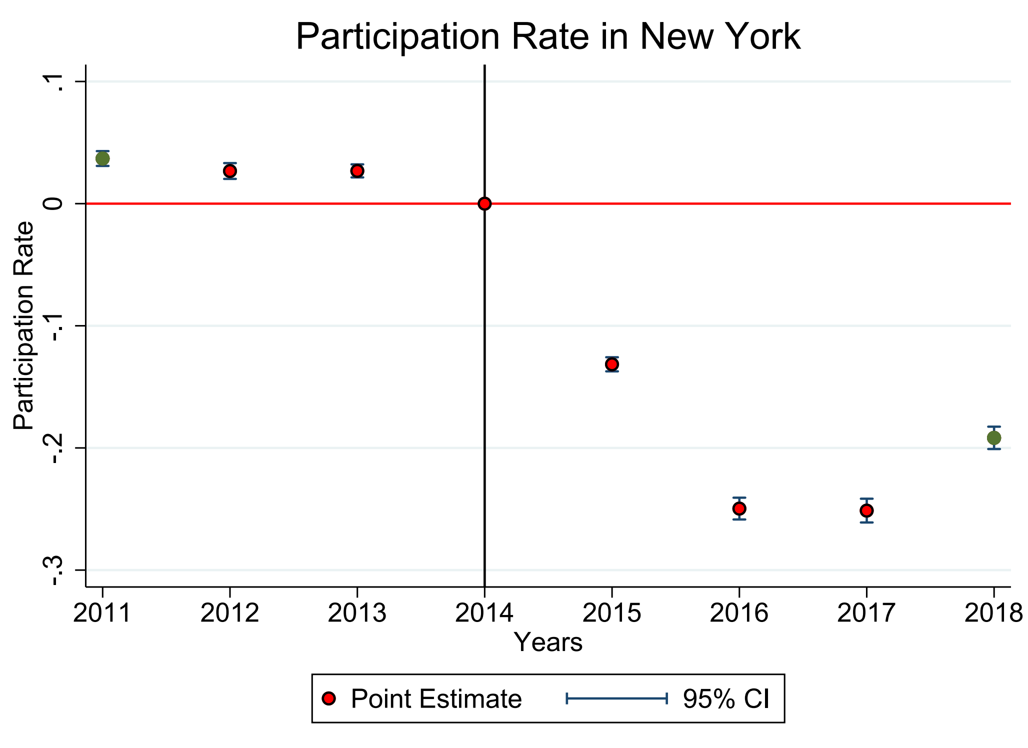

Figure 1.1: New York Grades 3-8 Fixed Effect Estimates from Specification (1.3)

movement and experience drops below the 95% participation requirement. The variation across states occurs in the year fixed effects post opt-out. As years pass from the initial protest there is greater resilience in the opt-out movement in New York. In the most recent year of the sample for New York (2018) the participation rate across all grades in schools remains below that of the introduction of the opt-out movement. In Colorado the opt-out movement seems to have lost much of the initial momentum that was built up in 2015 with participation rates climbing back to pre opt-out levels. The difference may lie in the motivation for the protest. In 2022, Colorado passed Senate Bill 70 which replaces

Senate Bill 191 and reduces the weight on job performance from 50% to 30%. Across the sample in the event study no reform to teacher evaluation occurs. This could indicate that Colorado parents and students may have been motivated to opt-out initially but the slow response on behalf of the state led to a decrease in motivation. In 2016, New York agreed to delay evaluations based on performance with 4 year lags. This meant that schools could not connect the classroom's performance to teacher evaluations until they had taught for 4 years at the school. Wang (52) finds through social network analysis that parents did not feel this was a solution to the issue but rather a temporary delay from the state. In 2019, New York removed the standardized test performance as a measure for teacher evaluation. The continued efforts illustrated by the parents and students in New York may be attributed to the real policy response that occurs on behalf of the state.

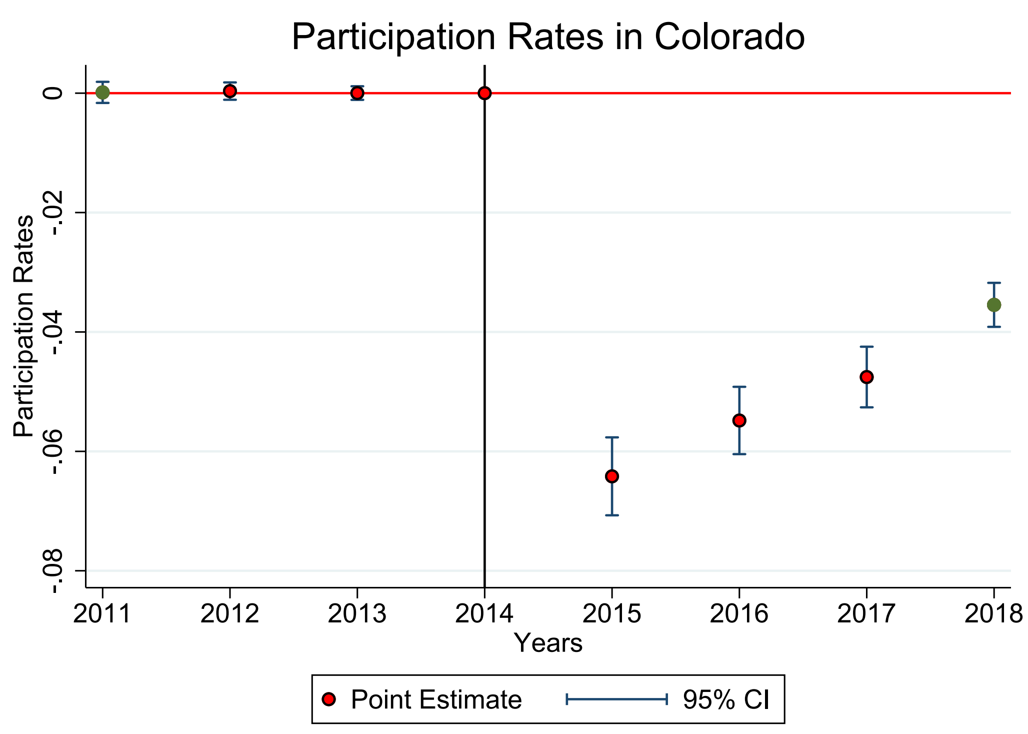

Figure 1.2: Colorado Grades 3-8 Fixed Effect Estimates from Specification (1.3)

1.3.2 Performance Results

To begin the performance analysis, I report the raw data for the school level proficiency over time. The goal of the NCLB is accountability. Low performing schools were identified and sanctions were set to incentivize schools to improve performance on state tests. The ESSA aims to provide equitable education to all students. The ESSA states a primary goal as "Advances equity by upholding critical protections for America's disadvantaged and high-need students" (2). School level achievement gaps are a good measure of equitable education.

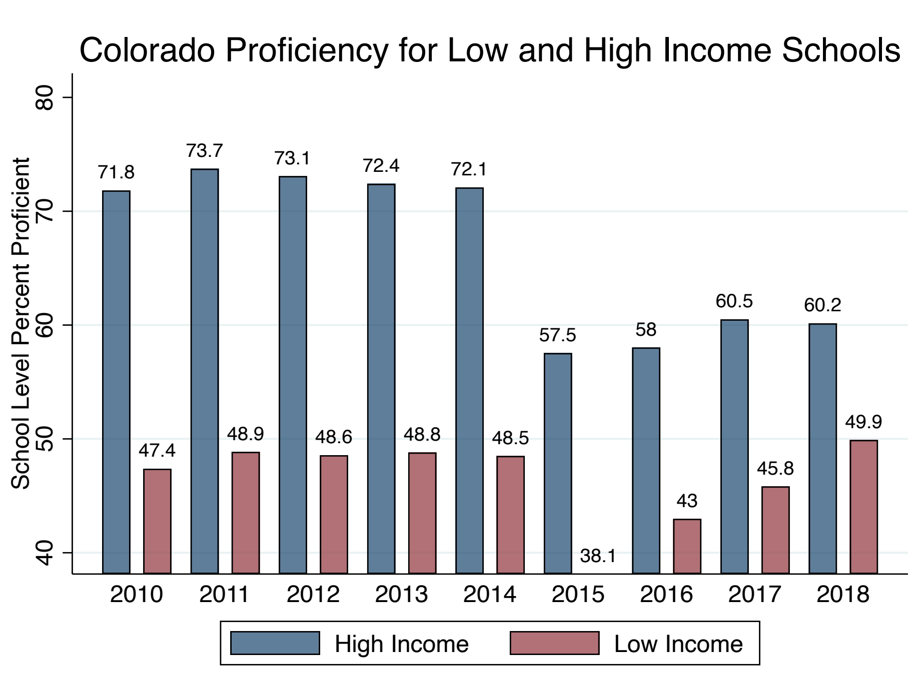

Figure 1.3: Proficiency Rates by School Income Composition in Colorado

I report the raw data for school level proficiency over time by income in Figures 1.3 and 1.4 for Colorado and New York respectively. Low income schools are defined by free and reduced lunch status. Schools that have more than 50% of their student body on free and reduced lunch are classified as low income schools. The achievement gap between high and low income schools in Colorado shrinks from 23.6% to 19.4% from 2014 to 2015. This

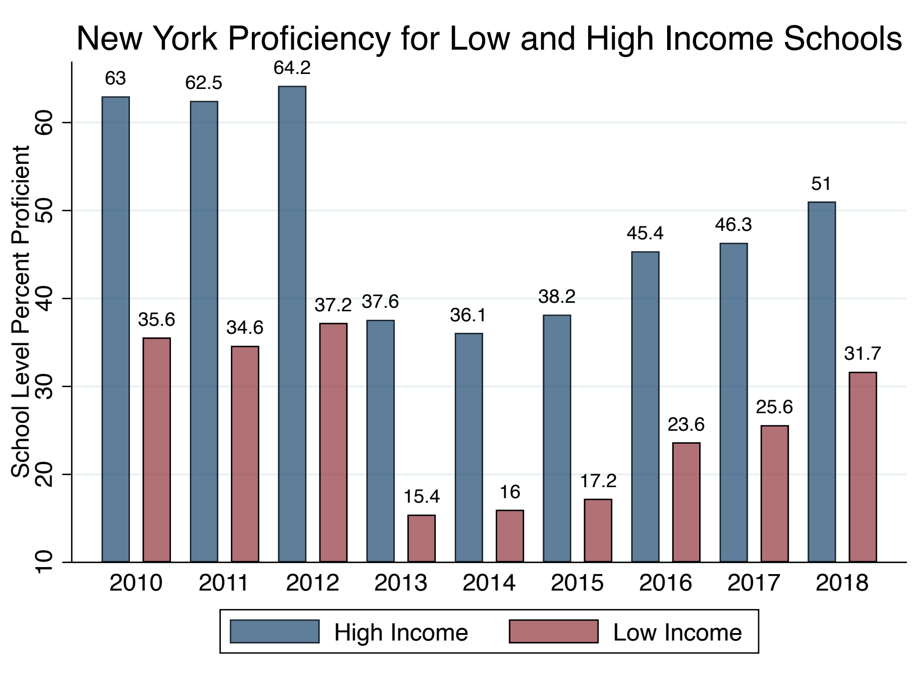

Figure 1.4: Proficiency Rates by School Income Composition in New York

gap continues to close over time. In New York the achievement gap shrinks in the year the protests begin but then begin to alternate in the following years. Figures 1.3 and 1.4 highlight the potential concern for policymakers ability to accurately assess achievement gaps. The data indicates that low and high income schools tend to fluctuate around a 20% achievement gap in proficiency.

The raw data indicates a sharp change in the test in 2015 in Colorado and 2013 in New York. This change in tests along with the large changes in participation rates makes it difficult to accurately assess the school level proficiency. I report the results from specification (1.4) in Table 1.4. The two-way fixed effect model shows the relationship between participation and proficiency before and after the opt-out movement. The relationship between free and reduced lunch status and performance is well documented in the literature and these estimates reinforce the finding that increasing the percentage of students on free and reduced lunch corresponds with lower performance. Increasing the proportion of Black and Hispanic students corresponds with lower rates of proficiency as well.

Table 1.4: Effect of Post Opt-Out Participation Rates on Proficiency

	NY	NY	CO	CO
Participation Rate	0.193***	0.170***	0.656***	0.657***
	(0.050)	(0.049)	(0.050)	(0.049)
Opt-Out*Participation	-0.144***	-0.136***	-0.423***	-0.421***
	(0.047)	(0.046)	(0.053)	(0.053)
%Free Reduced	-0.234***	-0.186***	-0.257***	-0.191***
	(0.023)	(0.019)	(0.030)	(0.026)
%Asian		0.530***		-0.064
		(0.057)		(0.092)
%Black		-0.253***		-0.043
		(0.063)		(0.073)
%Hispanic		-0.121***		-0.279***
		(0.033)		(0.036)
%Native		0.121***		-0.217
		(0.030)		(0.165)
%Multi		-0.021		0.010
		(0.039)		(0.069)
School FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	54115	54115	29867	29867

* Results from specification (1.4). Robust Standard errors clustered at the school level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Prior to the opt-out movement, the participation rate is positively related with percentage of students who test into the proficient category in New York and Colorado. The second coefficient in Table 1.4 illustrates the relationship between participation rates and the percentage of students who test into the proficient category post-op-out movement. The post-opt-out period indicates a dampening of the positive relationship reported in the pre-opt-out period. The coefficients see little to no change in magnitude when I include the racial composition variables.

I present the results of the calculation in Figures 1.5 and 1.6. New York demonstrates a gradual increase in the magnitude of the percentage point increase from 2015 to 2018 whereas Colorado experiences a gradual decline in the magnitude of the percentage point increase. The results indicate that across both states would see an average of 1-2 percentage point increase in proficiency if they maintained pre-opt-out levels.

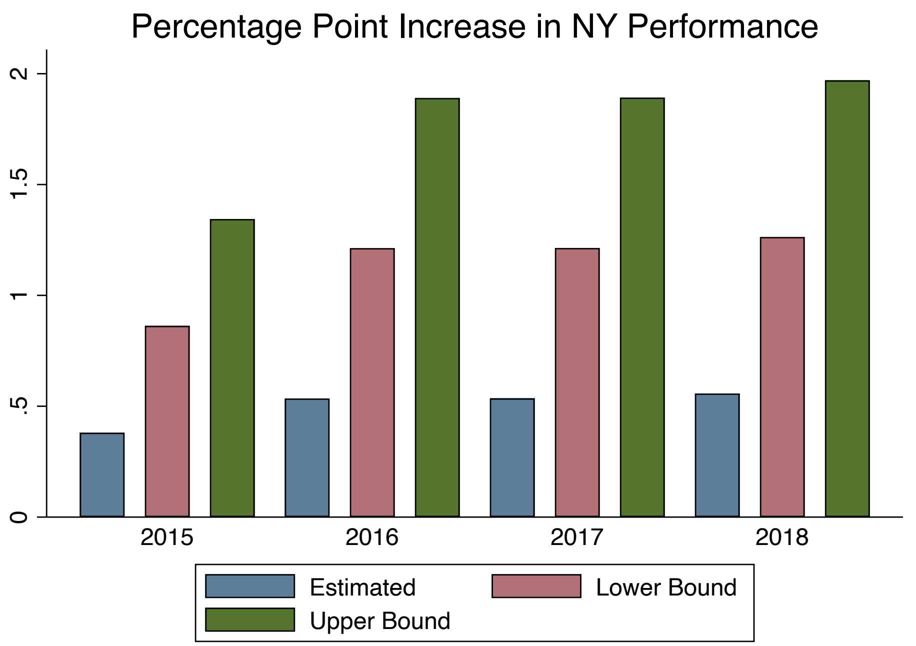

Figure 1.5: Percentage Point Increase in Proficiency NY with Pre-Opt Out Participation

The results indicate that if schools had maintained their pre-opt-out participation rates then school-level proficiency would be greater than the observed levels in the sample. As policymakers address the achievement gaps across various classifications reported in Figures 1.3 and 1.4 it is important to address the role of participation by each demographic group.

It is important to contextualize the performance results. In the participation analysis, I identify the suburban schools and White students as the primary participants in the opt-out movement for both states. The relationship between participation rates and school level proficiency is positive prior to the opt-out movement and this positive relationship is dampened post opt-out movement. Bennett (5) and Pizmony-Levy and Green Saraisky (48) highlight that White, affluent, educated parents may be at the forefront of the opt-out

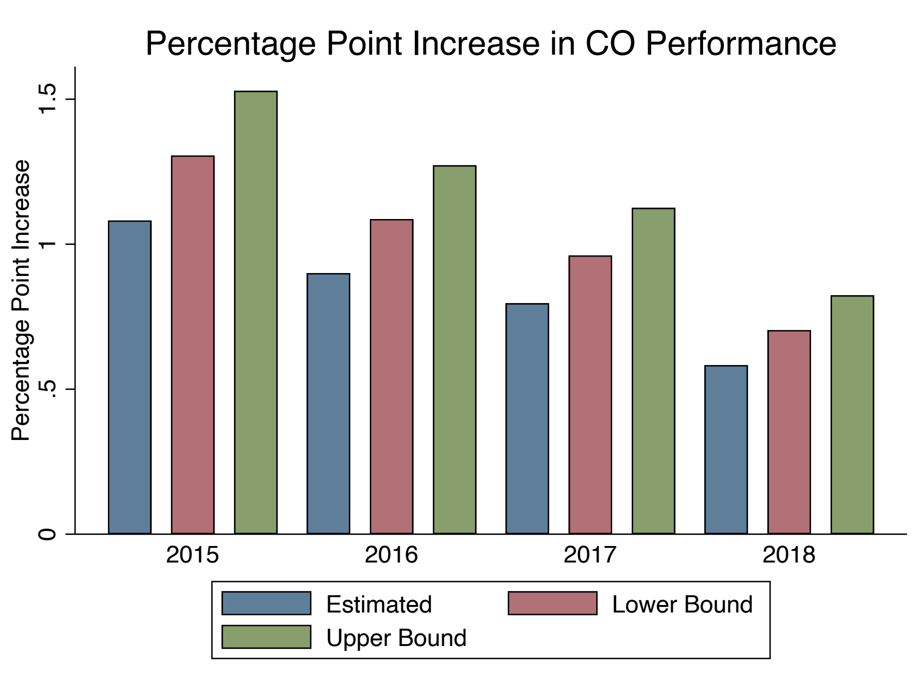

Figure 1.6: Percentage Point Increase in Proficiency NY with Pre-Opt-Out Participation

movement. Given the positive association between socioeconomic status and performance prevalent in the literature and demonstrated here, I can infer the mechanism behind the performance relationship. White affluent students in suburban schools are primary participants in the opt-out movement and as a result this causes the relationship between participation rates and school level proficiency to change. In the post-opt-out period the school's performance is missing students who have more affluent backgrounds and this means that if schools were to increase the participation rate with the sample of students who are taking the state tests this could reduce their rate of proficiency.

The dampened relationship means that policymakers could potentially overestimate the closure of achievement gaps between regional, racial, and income groups. The counterfactual calculations mean that if schools returned to their average pre-opt-out participation levels then schools could increase their level of proficiency. This indicates that policymakers who use the raw data reported in Figures 1.3 and 1.4 may be overestimating the closure in achievement gaps at the school level and potentially underestimate allocation of resources to low income and minority schools.

1.4 Conclusion

In this paper, I identify the concentration of the opt-out movement in suburban schools. The inclusion of school and year fixed effects illustrates the fact that schools with higher proportions of White students have higher opt-out percentages and that increasing the proportion of minority students has differential effects across states. Increasing the percentage of Black and Hispanic students correspond with a decline in participation rates in New York and the opposite is true for Colorado.

Figures 1.3 and 1.4 depict the fixed effect estimates for participation rates and indicate that both states satisfy the parallel trends assumption in the pre opt-out movement period and demonstrate the drop in participation rates following the movement. The year fixed effects post opt-out movement illustrate the variation in the resilience of collective action for parents and students, with New York remaining a strong opponent of standardized testing and Colorado returning to pre opt-out levels.

The performance analysis with the two-way fixed effect model demonstrated a negative relationship between participation rates and the percentage of students who test into the proficient category. At the school level, it appears that the affluent students who opt-out are higher performing and their lack of participation causes the positive relationship to drop substantially after the movement begins. The movement is centered in suburban schools and is driven by White students and this could lead to misinterpretation of achievement gaps between suburban, rural, and city schools and minority and White students.

1.4.1 Policy Implications and Future Research

The protests in New York seem to be effective with changes in legislation leading to the decoupling of teacher evaluations and standardized test scores. It is not clear whether the recent reduction in weight allocated to teacher's evaluation in Colorado is in response to the opt-out movement or other factors. The fixed effect model highlights the fact that certain demographic groups are more likely to engage in this type of collective action. State tests are imperfect measures of learning but they are currently the best measure we have at the school level to assess educational equity. Some argue that the achievement gap in and of itself is a racist idea (34; 33; 14). The opt-out movement pushes to eliminate the assessments because of testing anxiety and teacher evaluations. If the opt-out movement included minority groups then educational policy could change quickly and better measures of learning for grades 3-8 students could be developed. The centralized demographic participating in the opt-out movement and their motives do not align with the deconstruction of racially biased assessments.

Wang (52) highlights the friction between White and minority parents in New York. A similar friction exists in Colorado as the Twitter hashtag #optoutsowhite became popular in the state as the movement was picking up steam in 2015 (Gorski, 2015). The process of collective action efforts in both states are effective in dropping the overall participation rates of schools. This leads to incomplete data at the school and grade level and brings into question the coupling of teachers' evaluations and standardized tests. In New York the protest may very well have led to this decoupling.

The protests may be effective for decoupling evaluations but they also introduce issues of incomplete data at the school and grade level as a legitimate concern. Incomplete performance data at the school and grade level reduces policymakers ability to identify achievement gaps in schools and reduces their ability to create effective policies that could potentially reduce these inequities. The primary demographic of the opt-out movement is white children in suburban schools. As they continue to not participate in testing we lose valuable measurements to assess the achievement gaps between them and low-income students as well as the achievement gap between White and minority students.

Future research should incorporate student level data in affluent suburban schools where opt-out is prevalent and samples of low-income schools in neighboring areas. This would allow for an in-depth evaluation of the cheating concerns and it would also allow for analysis on the extent to which the incomplete data harms the assessment of achievement gaps. The analysis would include students who opt-out after taking at least one test. This would help to create the estimated performance of students who opt-out and compare that with the observed performance of classrooms. How many students would need to opt-out in a specific classroom to change the classroom's performance to the extent that it would cause a change the school's classification? This would allow for additional comparison for the observed classroom performance and estimated classroom performance with neighboring school classrooms to identify whether the opt-out leads to over or underestimation of the achievement gaps. This research would reinforce the benefits of complete data collection. Educational policy is imperfect but the ESSA requires consistent data collection to assess educational equity. State assessments will continue to change to provide a more holistic view of student learning but at the moment, the opt-out movement means incomplete data provides us with an imperfect picture to implement effective, equitable educational policy.

Chapter 2

The Pandemic and School Performance: Evidence from Colorado Schools

2.1 Introduction

The Coronavirus Disease 2019 (COVID-19) pandemic compelled schools to make difficult decisions. Administrators were faced with the daunting task of devising a mode of instruction that minimized the student, faculty, and staff exposure to COVID-19 while also meeting their students' learning needs. Throughout the year, school districts adopted various modes of instruction, including school closures with virtual learning options, full-time in-person instruction, and a combination of in-person and virtual learning known as "hybrid" schooling (32). The alteration in learning modes varied across districts, so students spent more or less time in-person depending on their district's decision.

National and international studies highlight how COVID-19 and virtual learning damaged educational success. Projections made by Kuhfeld et al. (35) indicated that learning losses would be most severe among the most vulnerable student populations across the United States. Their study utilized natural disasters like Hurricane Katrina to serve as proxies to project the learning losses for students due to lost instruction time (35). Kurmann and Lalé (36) find evidence of greater access to in-person instruction in private schools, less affluent areas, and schools with a larger share of white students, which they attributed partly to various regional differences. Domina et al. (21) found that marginalized populations were most severely affected by the pandemic's consequences, with losses in math and reading proficiency. Engzell et al. (23) look at the suspended in-person instruction in the Netherlands and find the average learning loss to be equivalent to a loss of 20% of a school year with lower income students seeing greater losses. To measure the effect of learning modes on state-mandated assessment pass rates, Jack et al. (30) collected district-level data from 11 different states. They created the COVID-19 School Data Hub (CSDH, 2022) to break down the mode of instruction into three categories: 1) "in-person," where all or most students had access to traditional, 5-day-per-week, in-person instruction; 2) "virtual," where all or most students received instruction online, five days a week; and 3) "hybrid," for schooling modes that did not fit into one of these categories. They used the following criteria for test score data: 1) at least 2 years of pre-pandemic test data; 2) no significant testing changes during this period; and 3) statewide participation rates in 2021 above 50%. They also included COVID-19 case rate data from USA Facts

Jack et al. (30) capitalize on the variation in learning modes across districts in 11 states to assess the impact of in-person instruction on district performance on state mandated tests. They find that moving a district from fully in-person to fully virtual learning would have reduced pass rate losses in Spring 2021 by 14 percentage points in Math and 8 percentage points in English Language Arts (ELA). In their summary analysis of the pass rate declines, Colorado stood out as an outlier relative to other states in the sample because it exhibited a smaller magnitude in declines. This paper analyzes the impact of the variation in instruction modes on Colorado's performance in more detail by incorporating school level data to explore the heterogeneity within districts. The school-level data allows me to address the projections in the literature that emphasize the differential impact of in-person instruction on marginalized populations. I use school-level data to analyze minority and white pass rates directly and provide additional insight on how the variation in learning modes impacted schools by regional classification.

2.2 Colorado Background Information

In the 2020-2021 school year, Colorado scaled back assessments for grades 3-8 by testing odd grades for ELA and even grades for Math. There was also a substantial decline in assessment participation rates across the state which varied from below 10% in multiple districts to 100% in other districts (10). The change in participation could bias the estimates of instruction mode on pass rates. If students who participate in the assessments are likely to perform better than those who did not, then estimates will understate the learning losses. If students who participated are likely to perform worse, then estimates will overstate the learning losses.

Jack et al. (30) highlight this issue and note that, "Based on state reports, participation declines during the pandemic appear to be larger among historically underserved student groups, such as students of color, students of lower socioeconomic status (SES) backgrounds, and students receiving special education services, among other student subgroups." Given that historically underserved student groups like minority students tend to lag behind white student peers, Jack et al. (30) argue that controlling for participation is necessary to avoid bias against finding a positive impact of in-person instruction. I incorporate school level participation to better capture the heterogeneity within districts. Table 2.1 shows the pass rates and participation rate by racial subgroups in 2019 which is the last year the test was administered prior to the COVID-19 pandemic. Colorado schools have fewer Black and Asian students and this lack of representation motivates my grouping of Minority students together for a Minority pass rate variable to estimate the school level regressions[1]. This rate is compared next to the 2021 year when changes in learning mode were present. For comparison, I restrict in Table 2.1 to only schools that had participation and pass rates in both years.

Across both student groups reported in Table 2.1, Minority students experienced slightly larger participation declines around 20% relative to White students who experienced 13% declines in participation in both assessments. Minority students see pass rate declines of 6% and 9% whereas White students see 3% and 19% pass rates declines for ELA and Math respectively.

[1]The minority variable is largely comprised of Hispanic students. Asian students have higher average proficiency rates and Black students have lower average proficiency rates relative to White students but both demographic groups only have unmuted data for 6 additional schools in the sample

Table 2.1: Pass and Participation Percentage Rates by Race

ELA	Pass 2019	Pass 2021	Participation '19	Participation '21	N
Minority	35.06	29.01	96.95	77.71	155
White	55.32	52.79	96.36	83.07	382
Math	Pass 2019	Pass 2021	Participation '19	Participation '21	N
Minority	40.74	21.52	95.75	78.20	88
White	57.90	38.75	96.29	83.09	268

* Table illustrates school-level data for the mean pass and participation rate by student subgroup. Here I restrict the sample to include only schools that have data for each subgroup in 2019 and 2021. I report only 2 student subgroups as Asian and Black student subgroups only have 3 and 2 schools where data is not muted for both years.

The magnitude of Jack et al. (30)'s estimates indicate that in future pandemic situations school districts should focus on providing in-person instruction to reduce learning losses. I utilize the district level data from the full sample in the Jack et al. (30) paper and estimate the following by state:

$$\%pass_{it} = \alpha + \beta_1(\%InPerson_{it}) + \beta_2(\%Hybrid_{it}) + \beta_3 X_{it} + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.1}$$

where i represents districts, t is the year, $pass_{it}$ is the pass rate for a specific subject in district i during year t, $\%InPerson_{it}$ is the share of days in-person in 2021, X_{dt} contains district level demographics (share White, share Black, share Hispanic, and share FRPL) in year t. γ_i includes district fixed effects to capture unobserved time-invariant district characteristics that may affect pass rates and I include δ_t which is a time fixed effect to control for time-varying shocks that may occur over the sample period and affect all districts. I present the results from Jack et al. (30) with the results from specification (2.1) in Table 2.2. Table 2.2 allows me to compare the magnitude of the coefficients for each state in Jack et al. (30)'s sample. Colorado is an outlier in that the magnitude of the coefficient is much smaller. I run the analysis by state and there are a few states that drive these results. Mississippi would see a potential 25 percentage point reduction in pass rate losses in Math and Ohio, Virginia, and Wisconsin would all see a 13 percentage point loss reduction or more by transitioning from fully virtual to fully in-person learning. Colorado appears to

Table 2.2: The Effect of In-person Learning on District Pass Rates by State

ELA	Jack et al.	CO	CT	MA	MN	MS	OH	RI	VA	WI	WV	WY
%In Person	0.081***	0.029**	0.137***	0.032*	-0.073***	0.105***	0.106***	0.056	0.027**	0.089***	0.067	-0.013
	(0.010)	(0.013)	(0.019)	(0.014)	(0.011)	(0.010)	(0.010)	(0.049)	(0.012)	(0.005)	(0.050)	(0.085)
%Hybrid	0.061***	0.011	0.108***	0.040***	-0.058***	0.060***	0.077***	0.048	0.061***	0.054***	0.044	-0.029
	(0.012)	(0.016)	(0.021)	(0.007)	(0.010)	(0.009)	(0.009)	(0.043)	(0.008)	(0.011)	(0.048)	(0.090)
Math	Jack et al.	CO	CT	MA	MN	MS	OH	RI	VA	WI	WV	WY
%In Person	0.140***	0.063***	0.099***	0.033**	-0.001	0.250***	0.151***	-0.040	0.153***	0.137***	0.138**	0.017
	(0.014)	(0.008)	(0.009)	(0.015)	(0.008)	(0.010)	(0.005)	(0.033)	(0.016)	(0.006)	(0.058)	(0.022)
%Hybrid	0.078***	0.016	0.044	0.033***	-0.025**	0.149***	0.089***	0.026	.070***	.043***	0.117**	-0.029
	(0.014)	(0.016)	(0.027)	(0.008)	(0.013)	(0.016)	(0.012)	(0.048)	(0.013)	(0.013)	(0.055)	(0.102)
N	10756	614	786	1136	1659	656	2742	108	660	1976	275	144

Results are from specification (2.1). Table 2.2 All regressions are weighted by district enrollment and include: a) district fixed effects; b) year fixed effects demographic controls (race/ethnicity shares, share of students eligible for free or reduced price lunch [FRPL], and share of English language learners [ELL]); d) county-level unemployment rates from the U.S. Bureau of Labor Statistics averaged by school year from June-May for 2016-2021; e) district enrollment; and f) test participation controls.; Robust Standard errors clustered at the District level in parentheses. Variables are bound between 0 and 1 and coefficients indicate the percentage point reduction in learning loss by going from 100% virtual instruction to 100% in-person or hybrid.

show a much more modest loss reduction with 6 percentage points in mathematics and 3 percentage points in ELA testing. In states like Mississippi and Virginia policymakers have a strong argument to transition to in-person learning to avoid significant learning losses. Policymakers in states like Colorado and Rhode Island would have a more difficult time arguing for a full transition to in-person learning modes. States like Minnesota spent approximately 85% of the year in hybrid or virtual learning modes. The negative coefficients for Minnesota is difficult to interpret as 66% of that school year was spent in hybrid learning modes.

Tables 2.1 and 2.2 motivate the school level analysis to explore the heterogeneity that exists within districts. The impact of in-person instruction is largely driven by a few states in the sample. School-level analysis will provide better insight to policymakers in Colorado and encourage this unit of analysis for each state in the future. In-person instruction may be the best policy option for some states in future pandemic situations but this may not be the case for every state in the sample.

I report coefficients on participation rate to address the relationship between instruction mode and performance while controlling for participation. The small coefficients in Table

2.2 imply that the policy push for in-person instruction in Colorado may not be the most appropriate however it might be the case that certain regions and demographic groups are more sensitive to changes in learning modes and this would mean that some school boards may have stronger arguments to transition to in-person learning within districts. Jack et al. (30)'s concern regarding the biased estimates due to the large declines in participation present in Table 2.1 motivate my inclusion of minority and white participation rates in the school-level analysis.

I analyze the role of in-person instruction at the school and district level in Colorado. The findings are as follows: (1) School-level analysis across all grades show no statistical or economic significance for the impact of in-person instruction. While grade-level analysis suggests in-person instruction can reduce learning losses in specific test grades like 6th grade Math, in the aggregate I find little statistical/economic significance. (2) Regional results suggest that suburban schools may experience reduced learning losses by 3-6 percentage points in Math, but no impact in ELA. (3) School-level results for disaggregated performance indicate that minority and white pass rates would not see a reduction in learning loss by moving to fully in-person instruction, but the coefficient for participation rates is positive and significant indicating that higher performing minority students has lower participation rates in the 2020-2021 school year.

The following sections outline the data and methods used in the analysis. I then present district level summary statistics and results for the district and school level analysis and conclude with policy implications and areas for future research.

2.3 Data and Methods

2.3.1 District-Level Data

This paper utilizes three sources of data from the Colorado Department of Education (CDE), CSDH, and the National Center of Education Statistics (NCES). Schooling mode data are drawn from the COVID-19 School Data Hub (Jack et al.). District demographic data is from the NCES. The CSDH updates Colorado learning modes monthly for the 2020-2021 school year. The learning mode data is constructed using each time period's schooling mode classification, the length of the time period, and the district. The total number of student-days spent in each schooling mode is then calculated for each district for the entire 2020-2021 school year. The days for each mode are then divided by the total number of district-level student-days to generate shares by schooling mode. In-person instruction indicates that schools were open for in-person attendance, but students may still have had the opportunity to attend virtually. The primary outcomes are pass rates for students in Grades 3-8 in ELA and Math, as measured by the share of students who meet or exceed expectations in ELA or Math on state assessments. The district level sample includes Colorado schools that administer grades 3-8 ELA and Math assessments from 2015-2022. Colorado tested odd grades 3,5, and 7 in ELA and even grades 4, 6, and 8 in Math for the 2021 school year. I follow Jack et al. (30) in using district-level participation data to show robustness to variation in participation. Demographic data from the NCES includes district-level information on the share of enrolled students by race and ethnicity, English language learner (ELL) status, and eligibility for free and reduced price lunch (FRPL). I utilize pairwise correlations that incorporate the average case rate from the COVID-19 School Data Hub to capture the possible role of variation in COVID-19 case rates in driving district's decisions to adopt a particular learning mode. I also run correlations between in-person schooling and the share of Republican votes by county in the 2020 national elections to control for political explanations for some of the variation in in-person instruction.

2.3.2 School-Level Data

The CDE provides school-level performance data for the Colorado Measures of Academic Success (CMAS) by race and FRPL status. This allows for direct estimation of the impact of in-person learning on specific demographic group performance. In the CDE, disaggregated school-level data, pass rates are muted if the number of valid scores is less than 16. This means that schools with less than 16 students of a specific subgroup will have their group's scores muted but they are still included in the school-level data. Schools may also have muted performance data with greater than 16 students of a specific subgroup but less than 16 valid scores due to lack of participation. Minority representation in Colorado schools is low, and minority students exhibited much lower participation rates in 2021. In order to produce estimates to identify the impact on minority students, all minority groups are combined into a single group. This allows me to test the vulnerability of subgroups to in-person treatment. The school-level data is merged with regional codes from the NCES. The NCES classification, which relies on the standard urban and rural definitions developed by the U.S. Census Bureau, is used to classify regions into four categories: city, suburban, town, and rural. I combine town and rural classifications because both classifications are at minimum 10 miles from an urbanized area and neither have population requirements. This combination creates three distinct categories of city, suburb, and rural. The regional codes allow for further exploration of demographic subgroups. When participation rates are sufficiently high, I estimate Minority and White pass rates by region as well. I follow Jack et al. (30) and weight each school-level regression by school enrollment to address the potential bias that could arise from enrollment declines and the possibility that these declines could be more prevalent with specific learning modes and contain students or specific performance levels.

2.3.3 District-Level Methods

I first estimate the determinants of in-person instruction mode. Grossmann et al. (25) find that decisions to reopen schools to in-person instruction were more tied to local political partisanship. I check correlations by regressing the share of in-person instruction on various demographic characteristics at the district level with the following:

$$\%InPerson_i = \beta_1 X_i + \varepsilon_i \tag{2.2}$$

Where $\%InPerson_i$ is the share of in-person instruction days for district i and X_i includes potential determinants of in-person schooling like Republican vote-share in the 2020 election, percent of free and reduced students, minority student shares, covid-case rates by county, and the previous pass rate by district.

I identify the impact of variation in learning modes on performance at the district level. I estimate the following:

$$\%pass_{it} = \alpha + \beta_1(\%InPerson_{it}) + \beta_2(\%Hybrid_{it}) + X_{it} + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.3}$$

where i represents districts, t is the year, $pass_{ict}$ is the pass rate for a specific subject in district i during year t, $\%InPerson_{it}$ is the share of days in-person in 2021. I code each year prior to the 2020-2021 school year as 100% in-person. I follow Jack et al. (30) and scale each variable between 0 and 1. This allows for a comparison of estimates. The bound between 0 and 1 means that the coefficients of interest are interpreted as the reduction in learning losses by transitioning from 100% virtual to 100% in-person or 100% hybrid. X_{it} contains district level demographics (share White, share Black, share Hispanic, and share FRPL) in year t. γ_i includes district fixed effects to capture unobserved time-invariant district characteristics that may affect pass rates and I include δ_t which is a time fixed effect to control for time-varying shocks that may occur over the sample period and affect all districts. I use fixed effect estimation because the unobserved heterogeneity is assumed to be constant. I use robust standard errors clustered at the district level in ε_{it}. The standard errors are clustered to account for heteroskedasticity. Independent variables are not random and the model exhibits no multicollinearity. Enrollment declines are a concern in the district and school level analysis. If the declines are larger in areas with more virtual learning, and the group who exits the system has systematically higher or lower test scores, this could bias the results. All regressions are weighted by district enrollment in the district analysis section and by school enrollment in the school section each regression.

I use grade level analysis by test to provide a robustness check which incorporates a similar estimation strategy to that in equation (2.2) with the following:

$$\%pass_{igt} = \alpha + \beta_1(\%InPerson_{it}) + \beta_2(\%Hybrid_{it}) + X_{it} + \gamma_i + \delta_t + \phi_{ct} + \varepsilon_{it} \tag{2.4}$$

where the only change is the dependent variables $pass_{igt}$ which is the pass rate for a specific subject in district i in grade g during year t.

The grade level analysis provides a robustness check by analyzing variations in the value of in-person learning by age group and, additionally, to the extent that there are changes in the size of these grades as a result of the pandemic, it is possible that these changes could be driving the overall impact.

To explore demographic variation in the effect of in-person learning, I estimate regressions with interactions between demographics and schooling mode and the treatment year for the pandemic. The regression takes the following form:

$$\%pass_{it} = \alpha + \beta_1(\%InPerson_{it} * X_{it}) + \beta_2(\%Hybrid_{it} * X_{it}) + X_{it} + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.5}$$

The regressions include interactions between race and ethnicity or FRPL shares and the in-person and hybrid variables. The regression includes baseline interactions between demographic variables and schooling mode. Specification (2.4) allows effects of in-person shares to vary by race. For example, the impact of in-person instruction may vary depending on the district's Black student composition. I estimate this at the district and school level and results are reported in Table 3.7 in the appendix[2].

[2] I estimate the (2.5) at the school level and report results in Table 3.8 in the appendix. Estimates are statistically insignificant and much smaller in magnitude compared to Jack et al. (30)

2.3.4 School-Level Methods

I first estimate the impact of district level learning modes on school-level pass rates by region with the following:

$$\%pass_{sgt} = \alpha + \beta_1(\%InPerson_{it}) + \beta_2(\%Hybrid_{it}) + \beta_3 X_{st} + \gamma_s + \delta_t + \varepsilon_{it} \tag{2.6}$$

where $pass_{sgt}$ is the pass rate for school s in grade g in year t. The $\%InPerson$ and $\%Hybrid$ variable still accounts for district level policies as Colorado only reported changes in learning modes at the district level. The variable X_{st} contains school level demographics (share White, share Black, share Hispanic, and share FRPL) in year t. I follow Jack et al. (30) and include the school level participation rates in X_{st} as well. The decline in participation rates presents selection issues but I observe school level participation rates and control for them. Each regression is weighted by school enrollment to address the concern that enrollment declines may be larger in areas with more virtual learning, and to the extent that the group who left the system has systematically higher or lower test scores, this could bias the results (17; 30). I include γ_s for school fixed effects to capture unobserved time-invariant school characteristics that may affect pass rates and I include δ_t which is a time fixed effect to control for time-varying shocks that may occur over the sample period and affect all schools. The standard errors are clustered at the district level because the schools are nested within the districts and the policy variable is at the district level. Policy changes impact all schools within that district and as a result the within-district errors are likely not independent of one another. Each regression here is weighted by school enrollment.

In order to address the differential impact of the COVID-19 pandemic on marginalized students I utilize the disaggregated data by demographic group to estimate equation (2.7). I estimate the impact of the instruction mode on Minority and White pass rates. I include the variable $part_{sgt}$ which controls for the Minority and White student participation rate. I run the following:

$$\%pass_{sgt} = \alpha + \beta_1(\%InPerson_{it}) + \beta_2(\%Hybrid_{it}) + \beta_3 part_{sgt} + \beta_4 X_{st} + \gamma_s + \delta_t + \varepsilon_{it} \quad (2.7)$$

where $pass_{igt}$ is the pass rate for the racial groups (Minority and White pass rates) in school s in year t. The variable $\%InPerson$ is the share of days in-person in 2021 and $\%Hybrid$ is the share of days spent in hybrid learning mode. I utilize the student subgroup participation rate as a control in this regression. The variable X_{st} contains school-level demographics (share White, share Black, share Hispanic, and share FRPL) in year t and I include γ_s for school fixed effects to capture unobserved time-invariant school characteristics that may affect pass rates and I include δ_t which is a time fixed effect to control for time-varying shocks that may occur over the sample period and affect all schools. Standard errors are clustered at the district level.

Kurmann and Lalé (36) find evidence of greater access to in-person instruction in schools with a larger share of white students, which they partially attribute to various regional differences. I explore the relationship between in-person instruction and performance by school regional classification in (2.8). I estimate the impact of district level learning modes on school level pass rates by region with the following:

$$\%pass_{st} = \alpha + \beta_1(\%InPerson_{it}) + \beta_2(\%Hybrid_{it}) + \beta_3 X_{st} + \gamma_s + \delta_t + \varepsilon_{it} \quad (2.8)$$

where $pass_{sgt}$ is the pass rate for school s in year t. The $\%InPerson$ and $\%Hybrid$ variable still accounts for district level policies as Colorado only reported changes in learning modes at the district level. The variable X_{st} contains school-level demographics (share White, share Black, share Hispanic, participation rates, and share FRPL) in year t.

2.4 Results

Colorado districts experienced high levels of alternative instruction modes in 2020-2021. I present averages across all districts for instruction mode in Table 2.3. The largest share of learning mode days are allocated to the hybrid learning mode at 43.8%. Hybrid learning modes could vary from school to school, with some providing 4 days of in-person instruction and some providing 1. Grossmann et al. (25) find that decisions to

Table 2.3: Instruction Mode Averages for Colorado

	Districts	% In-Person	% Hybrid	% Virtual
CO	136	28.9	43.8	27.3

reopen schools to in-person instruction were more tied to local political partisanship. I present pairwise correlations from specification (2.2) in Table 2.4. Table 2.4 illustrates the

Table 2.4: Pairwise Correlations between In-Person Learning on District Demographic and Pandemic Variables

	Correlation	Std. Errors
Prev Pass Rate	-0.693	(0.362)
Share Black	-2.193	(0.373)
Share Hispanic	-0.104	(0.217)
Share FRPL	-0.021	(0.168)
Share ELL	-0.756	(0.323)
Avg Case Rate	1.073	(0.367)
Repub Vote Share	0.012	(0.002)

* Correlations across Colorado school districts. The share of in-person measures the share of time during the 2020-21 school year that the district offered full time in-person instruction (rather than hybrid or virtual instruction). "Prev Pass Rate" represents the average pass rate on state standardized assessments for students in Grades 3-8 between 2016-2019. Demographic variables include: a) the share of students who are Black (based on NCES 2020-21 data); b) the share of students who are Hispanic (based on NCES 2020-21 data); c) the share of students who are eligible for free and reduced price lunch (FRPL) (based on NCES 2019-20 data due to changes in reporting requirements in 2020-21. Results from specification (2.2). All variables are bound by 0 and 1.

relationship between various demographic characteristics and Colorado districts in-person learning. The table demonstrates the raw pairwise correlations in the first column. Higher case rates and Republican vote shares are positively correlated with higher shares of in-person instruction whereas higher shares of minority students are negatively correlated with in-person instruction. Districts with higher shares of low-income students are also negatively correlated with in-person instruction.

2.4.1 District-Level Results

Table 2.5 illustrates the effect of instruction mode on performance at the district level in Colorado. Moving across columns within subjects, I add area-year fixed effects for progressively smaller areas. The coefficients demonstrate no significance and less stability relative to Jack et al. (30) with the inclusion of the county year fixed effects. The results are

Table 2.5: The Effect of In-Person Schooling on District-Level Pass Rates

Test Format	Math		ELA	
%In Person	0.040	-0.007	0.018	-0.026
	(0.013)	(0.008)	(0.028)	(0.038)
%Hybrid	0.008	-0.036	0.012	-0.024
	(0.029)	(0.029)	(0.030)	(0.048)
Commuting Zone X Year FE	Yes	No	Yes	No
County X Year FE	No	Yes	No	Yes
Observations	614	614	634	634

* Results from specification (2.3).Regression replicates A from Jack et al. (30). Table 2.4 shows the relationship between district in-person share, hybrid share, and pass rates in Math and English Language Arts (ELA) on state standardized assessments for students in Grades 3-8. Virtual share is the reference group. I present Columns 1 and 3 with commuting zone fixed effects, and county-year fixed effects in Columns 2 and 4. All regressions are weighted by district enrollment and include: a) district fixed effects; b) year fixed effects demographic controls (race/ethnicity shares, share of students eligible for free or reduced price lunch (FRPL), and share of English language learners (ELL); d) county-level unemployment rates from the U.S. Bureau of Labor Statistics averaged by school year from June-May for 2016-2021; e) district enrollment; and f) test participation controls. Robust Standard errors clustered at the District level in parentheses.

statistically insignificant and the magnitude is small relative to the findings from Jack et al.

(30) who found with a full sample of 11 states that moving a district from 100% virtual to 100% access to in-person learning would have reduced pass rate losses in Spring 2021 by 13 to 14 percentage points in Math and about 8 percentage points in ELA. The results here indicate that Colorado district's pass rates declined less relative to the other states in their sample to changes in learning mode at the district level and further reinforce Colorado as an outlier in the sample.

I use grade level analysis to provide a robustness check from (2.4). Jack et al. (30)'s analysis finds that when they run analysis by grade the significance and magnitude remains largely unchanged with all states in the sample. Colorado only tested odd grades in ELA and even grades in Math in the 2021 school year and this could contribute to the smaller effects present in Table 2.6. The in-person and hybrid learning mode exhibited no statistical significance in Table 2.5 but the grade level analysis shows that 4th grade Math may exhibit some reduction in learning loss by transitioning to in-person learning but the magnitude is smaller indicating a 6 percentage point reduction in learning loss for 4th grade Math. The literature suggests that in-person learning is particularly important for students in Grades K-3 given that they are "still developing the skills needed to regulate their own behavior and emotions, maintain attention, and monitor their own learning" (45), which are each critical aspects of in-person learning. Colorado's summary stats indicate that hybrid learning modes were most prevalent in the state and this could suggest that the mixed learning modes in Colorado incorporated enough in-person instruction to avoid significant learning losses. The stratified testing method could also hide the impact because early 3rd and 5th grade Math was not tested.

Table 2.6: The Effect of In-Person Schooling on Changes in Pass Rates by Grade Level

	Math In-Person	Hybrid	ELA In-Person	Hybrid	N
Grade 3			0.020 (0.030)	0.035(0.032)	553
Grade 4	0.066**(.027)	0.046(0.030)			512
Grade 5			0.056 (0.035)	0.041 (0.030)	558
Grade 6	0.044 (0.033)	-0.047(0.039)			501
Grade 7			-0.009 (0.045)	-0.030 (0.054)	558
Grade 8	0.053(0.091)	0.107(0.094)			460

* Table 2.5 estimates the impacts of in-person learning on student pass rates on state standardized assessments by grade. Results from specification (2.4). Odd grades in ELA and even in Math. Robust Standard errors clustered at the District level in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.4.2 School-Level Results

Table 2.7 reports the results of specification (2.6). The first column assesses the overall impact on all grades and the subsequent columns measure the impact by grade. At the school level there is no statistical or economically significant impact of learning modes on ELA pass rates but Math results indicate that transitioning from 100% virtual to 100% in-person or 100% hybrid could reduce learning losses by 5 percentage points for all grades. The grade level results indicate that this effect would largely be present in the 6th grader's performance. These results reinforce the district level findings and suggest that for Colorado schools, it may not be necessary to transition full to in-person instruction as most grades do not see a substantial reduction in learning loss. Jack et al. (30) note that the hybrid learning mode has significant variation in how they measure this in the COVID-19 School Data Hub (CSDH, 2022). These could be four days in-person and one virtual or one day in-person and four virtual. Colorado districts spent about 44% of the pandemic school year in hybrid learning modes and districts likely altered between 1 and 4 days a week in in-person schooling.

The school and district level demonstrate a small marginal benefit in terms of reducing learning losses by incorporating more in-person instruction, but Table 2.8 provides the

Table 2.7: The Effect of In-Person Schooling on School-Level Pass Rates

Pass Rates	All	ELA 3rd	5th	7th	All	Math 4th	6th	8th
% In-Person	-0.006	0.016	-0.019	-0.027	0.046	0.042	0.121***	-0.002
	(0.035)	(0.044)	(0.063)	(0.051)	(0.031)	(0.044)	(0.041)	(0.096)
% Hybrid	-0.016	-0.020	0.024	-0.022	0.051*	0.043	0.096*	0.053
	(0.041)	(0.050)	(0.073)	(0.073)	(0.029)	(0.050)	(0.058)	(0.111)
Participation Rate	0.050***	0.061	0.126	0.123*	0.082	0.136**	-0.011	0.104
	(0.019)	(0.041)	(0.087)	(0.062)	(0.057)	(0.060)	(0.036)	(0.082)
Observations	9128	2003	2035	931	6659	1489	830	612

Results from specification (2.6). All regressions are weighted by school enrollment and include: a) school fixed effects; b) year fixed effects demographic controls at the school level from the Colorado Department of Education (CDE) (race/ethnicity shares, share of students eligible for free or reduced price lunch [FRPL]). Robust Standard errors clustered at the District level in parentheses.

impact of instruction mode by Minority and White pass rates. Looking at the observations, much of the minority performance data suffers from having less than 16 students reported for that school. This is a result of less diversity within schools as well as the decline in participation rates. The first column illustrates the impact of moving from fully virtual to fully in-person or hybrid on minority pass rates and the second column highlights the impact on white pass rates. The instruction mode exhibits no statistical significance for either group's pass rates. The minority student pass rate has a positive and significant relationship on the minority pass rate. This could mean that higher performing minority students participated at lower rates relative to their lower performing peers. If this is the case, then Colorado could improve pass rates by ensuring that participation levels remained as close as possible to the pre COVID-19 levels. It is important to note that lack of significance for the in-person instruction does not imply that COVID-19 did not disproportionately impact marginalized populations. Colorado might exhibit less minority representation relative to the other states in Jack et al. (30)s' sample. It also incorporated stratified testing and the CDE mutes data with less than 16 observations. In order to fully understand the impact of instruction by demographic group administrative data would be required but Table 2.8 highlights the concerns presented by Jack et al. (30) in that the

Minority student participation decline is likely to bias the estimates in favor of finding a positive impact of in-person instruction.

Table 2.8: Effect of In-Person Schooling on Pass Rates by Racial Group in Colorado Schools

ELA Pass Rates	Minority	White
% In-Person	-0.023	-0.015
	(0.036)	(0.037)
% Hybrid	0.018	-0.039
	(0.056)	(0.046)
Participation Rate	0.146***	0.061
	(0.048)	(0.049)
Observations	2476	5007
Math Pass Rates	Minority	White
% In-Person	0.008	0.021
	(0.034)	(0.059)
% Hybrid	0.026	0.038
	(0.042)	(0.060)
Participation Rate	0.107**	0.079
	(0.051)	(0.066)
Observations	1703	4025

Results from specification (2.7). Table 2.8 incorporates the same controls used in Table 2.6. Table 2.8 alters the dependent variable to capture White and Minority pass rates. Standard errors are clustered by district and are reported in parentheses. All regressions are weighted by school enrollment. Robust Standard errors clustered at the District level in parentheses.

I conclude the results by exploring the regional concerns. It may be the case that schools under specific regional classifications dealt with changes in learning modes differently. I present the results from equation (2.8) in Table 2.9. I see some of the heterogeneity at the school level. Suburban schools saw a reduction in learning loss by 6 percentage points in Math by transitioning from fully virtual to 100% in-person instruction and a 3 percentage point reduction by moving from fully virtual to hybrid. Rural schools demonstrate a similar impact of transitioning from virtual to hybrid. City schools demonstrate no significance but if a large share of the minority students are concentrated in the city, then decline in pass rates can explain the negligible impact of learning modes here.

Table 2.9: Effect of In-Person Schooling on Pass Rates by Region in Colorado Schools

ELA Pass Rates	Rural	Suburb	City
% In-Person	-0.060	-0.027	0.014
	(0.064)	(0.032)	(0.034)
% Hybrid	-0.068	-0.073*	0.022
	(0.067)	(0.041)	(0.054)
Participation Rate	0.116**	-0.000	0.042
	(0.051)	(0.038)	(0.045)
Observations	3095	3171	2862
Math Pass Rates	Rural	Suburb	City
% In-Person	0.068*	0.060**	0.019
	(0.035)	(0.026)	(0.036)
% Hybrid	0.055	0.036**	0.073
	(0.044)	(0.017)	(0.060)
Participation Rate	0.040	-0.001	0.181
	(0.066)	(0.048)	(0.123)
Observations	2294	2307	2058

Results from specification (2.6). Table 2.7 incorporates the same controls used in Table 2.6. All regressions are weighted by school enrollment. Robust Standard errors clustered at the District level in parentheses.

2.5 Conclusion

The results from both district and school level analyses suggest that Colorado may have been an outlier in the sample studied by Jack et al. (30), but these results could be affected by limited performance data for minority students. Specifically, the findings here suggest that school board members in suburban schools could reduce learning losses by implementing more in-person instruction during future pandemics. However, outside of this specific subgroup, further data analysis would be needed to provide recommendations. Additional analysis with student-level administrative data may further reinforce the findings of Jack et al. (30). The findings indicate that higher participation rates are positively related with higher pass rates. In future disaster related events it would be important for schools to ensure that all students can easily access and participate in the assessments to gain accurate data and reduce the muting of school-level observations. This will be particularly important for more vulnerable student groups who experienced participation declines and exhibit less representation in the schools overall.

To expand on this research, future studies should include a disaggregated school-level analysis for the states with the largest coefficients reported in Table 2.7. For example, Mississippi, which has a student population composed of 51.5% Black and Hispanic students, could be a great starting point. Although such states may have also experienced significant participation declines, the diverse racial composition would mitigate the limitations imposed by insufficient data, enabling analysis specific to different racial groups. This would enable a direct estimation of the impact of learning modes on the pass rates of Black and Hispanic students, thus advancing our understanding of how learning modes can affect more vulnerable populations. In addition to these outlined areas, future research should include analysis on other performance outcomes. Studying the impact of these learning modes on on-time graduation rates would help to contribute to the literature on school disruptions and how they impact major milestones in student's academic careers.

Future analysis would also benefit from school-level instruction mode data to allow for better policy prescription to school administrators. Detailed breakdowns of the hybrid learning modes would provide better insight as well. The fact that some hybrid modes are largely in-person and some are largely virtual makes the hybrid learning mode coefficients difficult to interpret and provides little in the way of learning mode recommendations. The variation in hybrid learning modes could also explain the smaller magnitude and lack of statistical significance in the results. If hybrid modes are largely composed of 4 days of in-person instruction then this would mean that districts hybrid mode look very similar to fully in-person instruction modes. It would also help to better understand the quality of the instruction (22). Virtual learning may become a realistic alternative given schools have the resources and training to provide high quality instruction. In Colorado's case, city schools experienced higher levels of virtual learning relative to other regions but city schools also experienced larger participation declines. Diliberti and Kaufman (19) note that high-poverty schools may have had limited access to internet and technology resources and this could be a primary determinant of the participation declines.

Chapter 3

Student Athlete Support Services and Performance: Evidence from Colorado State University

3.1 Introduction

Colorado State University enrolls approximately 380 student athletes each year and on average the University allocates 214.1 athletic scholarships annually (46). Many of the student athletes at the University participate in the Student Athlete Support Services (SASS) program to assist them in their academic careers. SASS programs are prevalent across universities and these programs recognize the unique student athlete population (26).

The facilitation of National Collegiate Athletics Association (NCAA) athletics depends on the consistent signing of talented student athletes. Student athletes enter the University with varying levels of academic preparation. Policies like Proposition 48 mandate minimum ACT/SAT scores and require a minimum grade point average (GPA) for students to be admitted. The goal of these policies is to prevent the recruitment of athletes who are woefully unprepared for the academic setting in a university (27). The NCAA introduced the Bylaw 14.3.1.1.1 in 2003 which allows college coaches to recruit student athletes who achieve a minimum score of 400 on the SAT and a high school GPA of 3.55. This Bylaw serves as a subject of controversy as high-school GPA's could be inflated to ensure a student athlete's eligibility (42). An NCAA study conducted from 1975 to 1980 found that less than half (42.9%) of Division 1A football student athletes graduated college (16). This implies that there are still a number of student athletes that enter universities with an incomplete tool-kit to succeed. In 1991, NCAA Division I membership adopted a proposal mandating academic counseling and tutoring services for all Division I student-athletes. The list of services expanded in 2002 permitting institutions to finance any academic-support services determined to be appropriate and necessary for student athletes' academic success (42). Colorado State University serves as an excellent candidate for a case study of student-athletes and academic performance.

There is a rich literature on the positive effect of support services (53; 44; 39; 15; 13), however, this literature focuses on the general student population. Hollis (28) provides a theoretical framework for SASS programs. She states that programs like SASS are justified by equal opportunity theory. Mithaug (43) defines equal opportunity theory through two interventions: "The first intervention builds individual capacity to self determine—that is, make choices. The second intervention builds opportunity by decreasing obstacles." Hollis (28) notes that SASS programs are built on the second intervention and studies the relationship between these support services and graduation rates. She finds an inverse relationship between services offered by institutions in the study and the student-athlete graduation rate. She describes the development of SASS programs as a response to the admission of students with weaker academic profiles and finds that high-school preparation and college entrance exam scores are stronger predictors of college performance than support services offered. Dilley-Knoles et al. (20) look at the impact of support services on cumulative GPA and find that academic support programs were successful in terms of producing a 3.0 GPA for female student-athletes, but not for male student-athletes. Routon and Walker (49) uses propensity score matching to identify the impact of athletic participation on average GPA and find that participation generally has a negative effect.

I build off of the student athlete analysis from Dilley-Knoles et al. (20) and Hollis (28) by using fixed effect regressions and including semester GPA as my outcome variable. This provides better insight into the immediate impact of support services. I include individual and semester fixed effects in all of my regressions to control for time invariant characteristics and the major institutional changes that occur over the student athlete's time at Colorado State University. The two-way fixed effect model explores the within student relationship between support services and semester GPA as opposed to team variation. In this paper I use two administrative data sets to create a sample with two cohorts (Fall 2016 and 2017) of student athletes and develop a student semester panel that follows these cohorts through the spring semester in 2021. I demonstrate the variation in academic preparedness by reporting baseline characteristics for student athletes like entrance exam scores, high-school GPA, and minority composition and compare these baseline characteristics with the general student population.

I study the relationship between the four types of support services objective-based study hall, mentoring, tutoring, and academic skill development and student athletes' semester GPA and credit earned ratio at Colorado State University. I estimate regressions with student and semester fixed effects to identify the relationship between the support services and semester GPA. I find that during semesters in which student athletes receive intensive support services, like tutoring, they earn lower GPAs. In semesters where student athletes receive less intensive support services, like mentoring, they earn higher GPAs. The student athlete sample exhibits negative selection in this study because the student athletes are more likely to seek out or receive support services in semesters when they are struggling. This means that support service estimates are likely to be biased downward.

The following sections describe the support services. I outline the data and methods used in the paper. I present summary statistics for baseline characteristics and discuss the selection issues present in the paper. I then present the results of the analysis and conclude with discussion of results and ideas to extend the research in the future.

3.1.1 SASS

Colorado State University follows the structure of several different Division 1 Universities that offer support services. SASS department offers a variety of services that are designed to assist the student athletes in their academic pursuits. The University helps maintain the eligibility of athletes by hiring a variety of learning strategists in the department of student athlete support services. I describe the four types of support services in order from least intensive to most intensive including:

Objective-Based Study Hall: Objective-based study hall provides all student athletes an opportunity to build good study habits and skills through independent work. This is coordinated by a study hall monitor who navigates the room and encourages students to complete certain objectives throughout the session.

Mentoring: Mentoring is designed to support student-athletes who have individual learning challenges as they develop academic skills and progress towards academic independence. It consists of one-on-one meetings in which student athletes are closely monitored by an academic mentor to ensure that the student athlete remains diligent and completes the tasks at hand. Mentors are assigned by academic coordinators to student athletes who have the capacity to complete their assignments but need some assistance with paying attention to all details of an assignment.

Tutoring: Tutoring utilizes specialists in specific content areas to assist the student athletes understanding and retention of course-specific material. Tutoring in SASS is distinct from other course tutoring as tutors cannot assist in completion of course assignments. Tutors must develop problem sets that mimic homework or exam questions or use supplementary textbook problems. Tutoring differs from mentoring as students generally would not have the capacity to complete their assignments without the additional course specific training.

Academic Skill Development: Learning specialists offer highly intensive monitoring in which student athletes engage in one-on-one sessions to focus on academic skill development. Learning specialists have an educational background and focus on implementing skill specific goals throughout the semester with student athletes. This type of monitoring is assigned by academic coordinators and assists student athletes who are diagnosed with a learning disorder, cognitive challenges, or academic under-preparedness.

The services are set in place to meet the student's needs and ensure GPA requirements are met. Student athletes vary with respect to support services received on the extensive margin (i.e. whether they receive support services or not) and the intensive margin (i.e. how much support they receive). Academic coordinators will determine whether student athletes require more or less support and in some cases student athletes will communicate to their coordinator that they would like to receive additional support. The process is continuous and students will be subject to an increase or decrease in support services in subsequent years depending on their performance each semester. Students may receive multiple services simultaneously. For example, a student could have academic skill development and tutoring sessions in the same week.

3.2 Data and Methods

3.2.1 Data

I use two sources of administrative data from Colorado State University. The first source of data comes from Colorado State University's SASS and contains data on support services for student athletes. I observe each student athlete's meeting type, frequency, date, and duration of each meeting from the Fall 2016 semester to the Spring 2021 semester. I aggregate this meeting data to semester-student observations by summing each meeting type and summing student athlete's total hours in meetings for each semester. I use the SASS data to build a set of indicator variables for reception of each type of service in the semester and these serve as the primary variables of interest in the analysis. The SASS data has dates for each meeting that I use to construct an indicator variable for reception of services in the first 4 weeks of the semester. This variable captures the early intervention of support services. The early reception indicator is generated for each type of support service (i.e. early mentoring, early tutoring etc.).

The second source of data comes from Colorado State University's Institutional Research, Planning and Effectiveness (IRPE). The IRPE data includes data on all students at

Colorado State University. It includes an indicator variable for student athlete but this variable is incomplete and thus my sample for each cohort is smaller than the average 380 student athlete population that enters each year. I use the athlete indicator variable from IRPE to include student athletes who do not ever participate in the SASS. The IRPE data includes my main dependent variables of semester GPA and credit earned ratio each semester. It also includes individual characteristics like minority status, gender, high-school GPA, ACT/SAT score, and Pell grant eligibility. Time invariant individual characteristics are captured in my individual fixed effects.

I merge the student athlete data from SASS with IRPE data from 2016-2021. The 2016-2021 sample period allows me to observe the Fall 2016 and Fall 2017 cohorts for a four year period so I restrict the sample to these two cohorts. The students from each cohort do not necessarily appear in each semester. Some students leave the University throughout the sample. I have a panel data set of 100 student athletes in the Fall 2016 cohort and 96 student athletes from the Fall 2017 cohort.

3.2.2 Summary Statistics and Selection Issues

Student athletes are not ignorant to the college wage premium, but many student athletes recognize that a large determinant of their acceptance lies in their athletic performance. This means that SASS is more likely to be subject to negative selection. Examples of selection bias are common in health studies that recruit participants directly from clinics and miss all the cases who do not attend those clinics or seek care during the study. In this case, student athletes' ability is related to their reception of services. Some student athletes are more likely to receive support services. Students who are less academically prepared are more likely to struggle during the semester and receive support services. Student athletes with strong academic skills and high levels of motivation may not receive any support services. The selection issues imply that the student athlete support services act as a tool to assist those who are already struggling and as a result the relationship between support services and student's semester GPA is likely to be negative.

I present summary statistics of the individual background characteristics for the two student athlete cohorts and the Fall 2016 and Fall 2017 general student cohort in Table 3.1.

Table 3.1: Student Athlete and General Student Cohort Comparison

	FA16 Athletes	FA17 Athletes	FA16 Student	FA17 Student
Minority	0.38	0.40	0.24	0.26
First Generation	0.22	0.24	0.23	0.26
SAT	1072	1073	1209	1204
ACT	23.7	23.7	25.0	25.0
HS GPA	3.55	3.60	3.58	3.60
N	100	96	6370	6362

Table provides mean characteristics for the entering Fall 2016 and Fall 2017 student athlete with the general student population.

Entrance exam scores are imperfect, but provide a standardized measure of academic preparedness. Student athletes' average SAT score is approximately 130 points lower than that of the general student population and the mean ACT score is 1.3 points lower relative to the general student population. Student athletes have more minority representation as well. The Fall 2016 cohort has 12 percentage points more minority students than the general students in 2016 and the Fall 2017 cohort increases the gap in minority representation to 14 percentage points.

In Figure 3.1[3], I plot the two student athlete cohorts' semester GPA over time. The student athlete cohorts are designated by the dashed lines and the general student cohorts are solid lines. The two cohorts demonstrate variation in their first semester with the 2017 cohort beginning with an average cohort GPA of 3.0 and the 2016 cohort beginning with a 2.7 GPA.

[3]Figure 3.1 tracks mean cohort GPA through 2021 and uses the 5th year for the Fall 2016 cohorts.

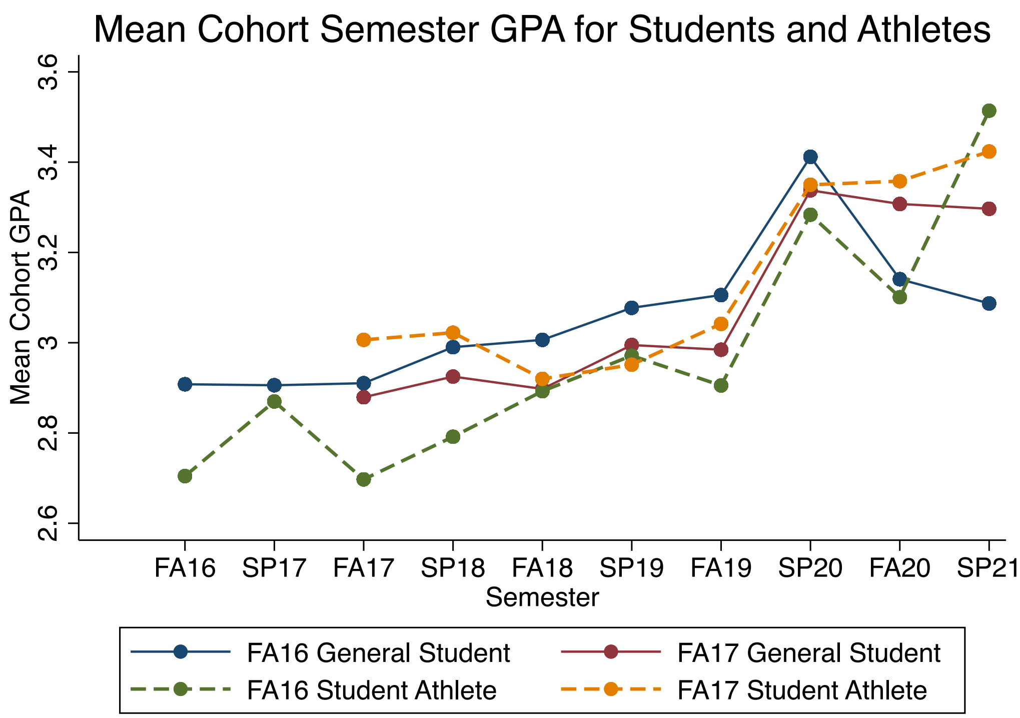

Figure 3.1: General Student versus Student Athlete Average Cohort Semester GPA Over Time

I plot the general student cohort's semester GPA over time for comparison. The general student cohort demonstrates a starting average of 2.9 for their semester GPA and each cohort gradually trends upward from this point. However, prior to the pandemic, both groups see a steady increase in mean cohort GPA and this increase coincides with students taking more upper level courses at the University. The sharp increase in mean cohort GPA from Fall 2019 to Spring 2020 is due to a change in the University's grading policy during the COVID-19 pandemic. When the pandemic forced alternative instruction modes, the University allowed students to opt in to satisfactory/unsatisfactory grading. Neither satisfactory nor unsatisfactory grades are counted in the calculation of GPA. This boosted the semester GPA for all students. The semester fixed effects should control this change in policy, but I nevertheless include analysis that ends in Fall 2019 as robustness check in Table 3.9 in the appendix and the results are largely unchanged.

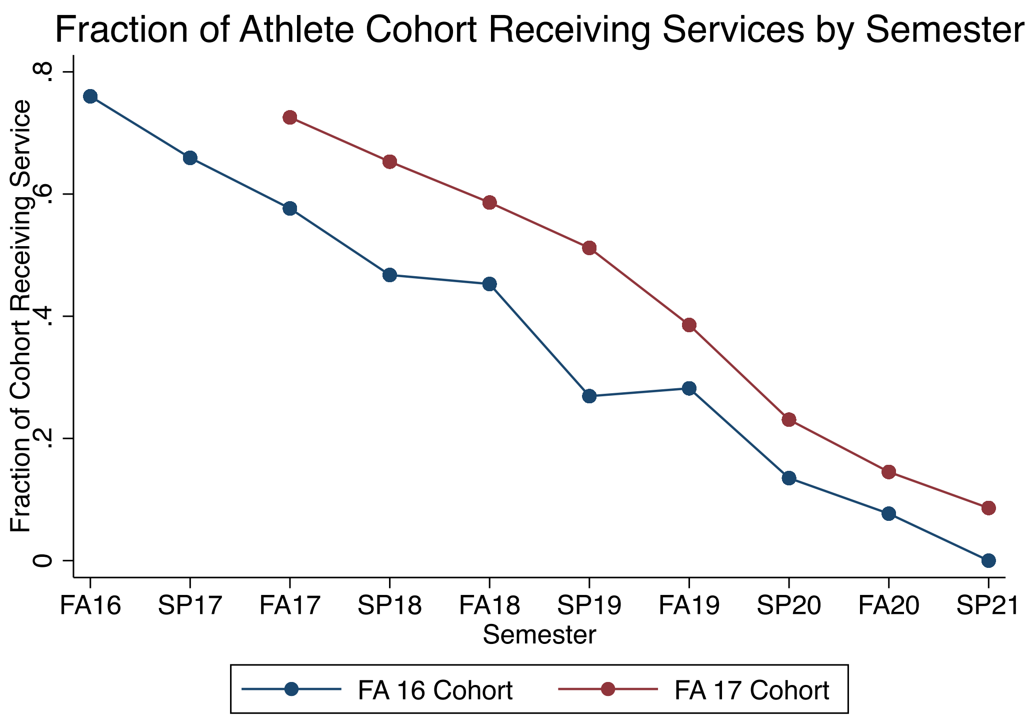

Figure 3.2: Support Service Meetings Over Time by Cohort

Figure 3.2 illustrates the declining support service meetings over time for the 2016 and 2017 student athlete cohort. All services drop from 76% in their first year to less than 20% in their final semester. Both cohorts demonstrate a similar decline in meetings. I report the trends by support service type in the appendix in Figures 3.3 and 3.4. Intensive support services like tutoring and academic skill development see steady declines over the student athlete's time at the University. Less intensive support services like mentoring see fluctuations over time. These initial summary statistics highlight two main features. First, student athletes demonstrate slightly lower academic preparedness relative to the general student population based on entrance exam scores. This is also present in the high proportion of student athletes receiving support services in their first years. Second, student athletes demonstrate learning by doing as their mean semester GPA improves and they see a steady decline in support services each semester.

The analysis breaks students into two groups. Students who receive any support service in a semester fall into the treatment group for that semester, and students who do not receive any support service fall into the control group for that semester. This means that the analysis will focus on the relationship between support services for the individual student's semester GPA when they receive services versus when they do not receive support services. In Table 3.2, I illustrate the entrance characteristics for student athletes who receive support services and student athletes who do not. Student athletes who receive support services have mean SAT scores approximately 120 points lower than student athletes who do not receive support services. Student athletes with meetings are at least two times the minority composition of students who do not receive support services and have higher levels of Pell recipients and first generation students as well.

Table 3.2: Athletes in Semesters with and without Services

	Meetings		No Meetings	
	FA 16	FA 17	FA 16	FA17
Minority	0.47	0.45	0.08	0.26
First Generation	0.25	0.28	0.08	0.15
SAT	1038	1050	1158	1223
ACT	22.9	22.9	26.4	25.9
HS GPA	3.52	3.50	3.62	3.90
Pell Recipient	0.28	0.24	0.04	0.19
N	284	292	278	347

Mean entrance characteristics for the Fall 2016 and Fall 2017 student athlete cohorts. N illustrates the number of semester-student observations I have for each cohort. For example, I see 284 semester-student observations in which students from the 2016 cohort receive support services. "Meetings" indicates the group of student athletes who received support services and "No meetings" is the group of student athletes who did not receive support services.

Table 3.2 illustrates the negative selection at hand. Student athletes who have lower mean entrance characteristics and lower levels of academic preparedness are more likely to receive support services every semester they are enrolled at Colorado State University.

This means that any positive impact of the support services is likely to illustrate the lower bound for the benefit to the student athlete.

3.2.3 Methods

I first analyze the relationship between the reception of each support service and student athlete's semester GPA and credit earned ratio. I estimate the following:

$$Y_{it} = \beta_1 D_{it} + \lambda_i + \mu_t + \varepsilon_{it}. \tag{3.1}$$

where Y_{it} is student i's semester GPA or credit earned ratio in semester t and D_{it} is a binary indicator for reception of services. The model includes individual fixed effects with λ_i to capture observed time-invariant individual characteristics like the students race, gender, high-school GPA, Pell-grant-eligibility, and first generation status, which may affect the student's GPA as well and unobserved time-invariant characteristics like academic preparation. I also include the semester fixed effects μ_t to control for common time-varying shocks and the steady increase in GPA illustrated in Figures 3.1 and 3.2. This sample includes the onset of the Coronavirus Disease 2019 (COVID-19) Pandemic, which impacted all students in the sample. I use fixed effect estimation because the unobserved heterogeneity is assumed to be constant. Independent variables are not random and the model exhibits no multicollinearity. I use robust standard errors clustered at the individual level in ε_{it}. The standard errors are clustered to account for heteroskedasticity. I cluster standard errors at the student level and because the treatment occurs at the individual level in this sample. Abadie et al. (1) note that clustering is not appropriate when sampling assignments are random. The negative selection discussed in the previous subsection indicates that the samples derived here are not random and therefore I cluster at the student level. I do not have serial correlated errors because I cluster the errors on the individual variable.

Some support services are more likely to be subject to negative selection. I construct indicator variables for reception of each support service separately and estimate the following:

$$Y_{it} = \beta_1 D_{1it} + \beta_2 D_{2it} + \beta_3 D_{3it} + \beta_4 D_{4it} + \lambda_i + \mu_t + \varepsilon_{it}. \tag{3.2}$$

where Y_{it} is student i's semester GPA is semester t. Here D_{it} is a binary indicator for reception of a particular service. The model includes individual fixed effects with λ_i to capture unobserved time-invariant individual characteristics. I include μ_t which denotes the semester fixed effects to control for time-varying shocks that may occur over the sample period and I use robust standard errors clustered at the individual student level in ε_{it}.

Even though my analysis includes individual fixed effects, it could still underestimate the impact of support services in grades if students are more likely to receive services in semesters that are particularly challenging. I attempt to identify the causal effect of support services by exploiting data on when students receive their first support service session each semester. I observe the date for each support service and I differentiate between (1) early intervention support services and (2) support services in general. I construct an indicator variable that turns on if the student athlete's first support service session takes place between weeks 1 and 4 of the semester. The estimation strategy will take the following form:

$$\begin{aligned} Y_{it} = {} & \beta_1 D_{1it} + \beta_2 e * D_{1it} + \beta_3 D_{2it} + \beta_4 e * D_{2it} + \beta_5 D_{3it} \\ & + \beta_6 e * D_{3it} + \beta_7 D_{4it} + \beta_8 e * D_{4it} + \lambda_i + \mu_t + u_{it}. \end{aligned} \tag{3.3}$$

where Y_{it} is student i's semester GPA and credit earned ratio in semester t. The variables $e{*}D_{it}$ are indicator variables that turn on if students receive the particular service within the first four weeks of the semester. The model includes individual and semester fixed effects. The variable $e{*}D_{it}$ allows me to look at the temporal effect of support services. The selection issues might be mitigated by early support service intervention. Early intervention could indicate that student athletes are taking initiative to optimize performance that semester, or that their academic coordinators assign intervention early to minimize the probability of failing a course. Students who first receive services after week 4 are more likely to be subject to negative selection, where the student receives support services after struggling on one or two exams.

I provide additional analysis to estimate the impact of support services at the intensive margin by including continuous variables measuring the number of meetings for a particular service in a semester. The indicator variables estimate the effect of service reception versus no reception, and the continuous variables estimate the effect of the number of meetings each semester. I estimate the following:

$$\begin{aligned} Y_{it} = \beta_1 Ct_{1it} + \beta_2 e * Ct_{1it} + \beta_3 Ct_{2it} + \beta_4 e * Ct_{2it} + \beta_5 Ct_{3it} \\ + \beta_6 e * Ct_{3it} + \beta_7 Ct_{4it} + \beta_8 e * Ct_{4it} + \lambda_i + \mu_t + u_{it}. \end{aligned} \tag{3.4}$$

where Y_{it} is student i's semester GPA or credit earned ratio in semester t. The variables Ct_{jit} are continuous variables that measure the count of meetings in a particular service that the student attends in the semester. The variables $e * Ct_{kit}$ are interactions between indicator variables that turn on if students receive the particular service within the first four weeks of the semester and the continuous variable that measures the count of meetings for a particular service the student receives in a semester. The model includes individual and semester fixed effects. The early intervention and continuous indicator measure the relationship between early meeting count and the outcome variables.

3.3 Results

Table 3.3 illustrates the negative selection with reception of any service illustrating a negative relationship with the student's semester GPA and a negligible relationship with their credit earned ratio. The results suggest that students' GPAs decline by 0.16 during the semesters they receive any support services. This effect is presumably driven by negative selection, in which student athletes who are struggling in a given semester receive support services to maintain a passing grade.

Table 3.3: Effect of Receiving Any Support Services on Semester GPA and Credit Earned Ratio

	Semester GPA	Ratio
Reception of any Service	-0.164***	-0.015
	(0.041)	(0.009)
N	1187	1201

Results are from specification (3.1). Table 3.3 includes individual and semester fixed effects. Robust Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4 presents the results from specification 3.2. Separating the indicator by support service type highlights tutoring's negative relationship with both the student's semester GPA and their credit earned ratio. This support service is more intensive in that it requires content specific engagement and detailed practice that cannot be related to specific homework, exams, or writing assignments. This type of support service is more likely to be driven by negative selection, because students who are struggling in a course are more likely to receive an intensive support service like tutoring. Mentoring has a positive relationship with the student's semester GPA. This is an example of a less intensive service. Student's receive mentoring support services when they are generally comfortable with the content for specific courses but may need assistance with checklist completion. It is important to note that there could be positive selection present in the sample as well. Without qualitative data it is impossible to capture the student athlete's motivation. It may be the case that more proactive student athletes seek out less intensive support services like mentoring in order to keep themselves on task and focused throughout the semester.

I isolate the timing for the reception of support services and present the results in Table 3.5. The indicator variable for the reception of the service without the early intervention indicator turns on for reception of services at all. This means that the overall impact of a particular service can be interpreted as additive. The intensive support services like tutoring are robust to each specification indicating that this support service in particular is subject to the negative selection issue. The early mentoring support service

Table 3.4: Effect of Receiving Support Services on Semester GPA and Credit Earned Ratio

	Semester GPA	Ratio
Mentoring	0.174**	0.11
	(0.080)	(0.021)
Study Hall	0.021	-0.20
	(0.066)	(0.024)
Tutoring	-0.222***	-0.018**
	(0.037)	(0.007)
Academic Skill Dev	-0.003	0.018
	(0.091)	(0.021)
Observations	1187	1201

Results are from specification (3.2). Table 3.4 includes individual and semester fixed effects. Robust Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

intervention exhibits a strong positive relationship with a student athlete's semester GPA. The overall effect of mentoring is captured by adding the coefficients for mentoring and early mentoring meaning the overall impact of receiving mentoring support services is a $-0.175 + 0.413 = 0.238$ increase in GPA. Comparing the results in Tables 3.4 and 3.5 demonstrates that the benefit from mentoring comes primarily from early intervention. The results from the early intervention specification further illustrate the heterogeneity in support services. When student athlete's feel comfortable with the course content and receive additional support the combination can mitigate the negative selection issues.

I explore the intensive margin in Table 3.6. Tutoring remains robust to each specification in that there exists a negative relationship between the number of tutoring sessions and the student athlete's semester GPA. The magnitude of the tutoring coefficient declines substantially here indicating that the extensive margin specifications may be capturing some late semester sessions where the student's grade is already low. The early mentoring count coefficient demonstrates a 0.03 increase in the semester GPA with an additional meeting early in the semester. The mentoring count variable is negative here without early intervention meaning that a large portion of the benefit student athlete's gain from

Table 3.5: Effect of Receiving Early Support Services on Semester GPA and Credit Earned Ratio

	Semester GPA	Ratio
Mentoring	-0.175	-0.045
	(0.197)	(0.074)
Early Mentoring	0.413*	0.069
	(0.210)	(0.070)
Study Hall	-0.225	-0.94
	(0.556)	(0.230)
Early Study Hall	0.256	0.081
	(0.545)	(0.223)
Tutoring	-0.200***	-0.008
	(0.049)	(0.011)
Early Tutoring	-0.050	-0.020
	(0.055)	(0.013)
Academic Skill Dev	-0.652	-0.120
	(0.652)	(0.110)
Early Academic Skill Dev	0.665	0.142
	(0.654)	(0.112)
Observations	1187	1201

Results are from specification (3.3). Table 3.6 includes individual and semester fixed effects. Robust Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

support services comes early in the semester. Increasing the number of meetings late in the semester has no additional benefit.

3.4 Conclusion

Student athletes are a subset of the higher education population that have seen little focus in the economic literature. Intercollegiate athletics and support service institutions are costly and the efficacy of these support service institutions should be an area of concern. Here I identify two cohorts of student athletes at Colorado State University and find that negative selection is present in the sample because struggling students are more likely to receive support services. I find that less intensive support services like academic mentoring are positively related to a student's semester GPA. Extending this analysis to include course specific grades will provide better insight into the efficacy of tutoring

Table 3.6: Effect of Meeting Count on Semester GPA and Credit Earned Ratio

	Semester GPA	Ratio
Mentoring Count	-0.021**	-0.007
	(0.010)	(0.006)
Early Mentoring Count	0.031***	0.008
	(0.011)	(0.005)
Study Hall Count	0.022	0.019
	(0.051)	(0.021)
Early Study Hall Count	-0.024	-0.020
	(0.050)	(0.021)
Tutoring Count	-0.016***	0.000
	(0.005)	(0.001)
Early Tutoring Count	0.007	-0.001
	(0.005)	(0.001)
Academic Skill Dev Count	-0.042	-0.008
	(0.028)	(0.005)
Early Academic Skill Dev Count	0.043	0.009*
	(0.029)	(0.005)
Observations	1187	1201

Results are from specification (3.4). Table 3.6 includes individual and semester fixed effects. Robust Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

which is robust here and demonstrates a negative relationship with semester GPA. Course specific grades could improve for students who receive tutoring for these courses. As students allocate more time to studying a specific subject this could lead them to allocate less time to their additional courses and their GPA could fall.

Maloney and McCormick (41) finds that the relatively poor performance of student-athletes in high-profile sports (in comparison with non-athlete students) was isolated to the playing season; these student athletes tended to fare better than their non-athlete counterparts in the off-season. Collaboration between SASS programs and IRPE institutions could allow for the development of complete athlete indicator variables that capture the athletes sport. This would allow for additional analysis to estimate the relationship between support service variables and outcomes in-season versus not. The sport indicator would allow for support service analysis at the team level as well. Routon and Walker (49) find that participation in intercollegiate sports has a minor negative impact on performance for the majority of student athletes relative to the general student population and that this impact is more pronounced for male basketball players and football players. There could be heterogeneity between sport when examining this relationship as well.

Future research here should focus on developing better identification strategies. Identifying scholarship versus non-scholarship student athletes would be a good first step. Non-scholarship student athletes are "walk-ons" and this means that they received acceptance to the university based on their academic preparedness alone. Being able to initially identify these student groups would reduce the selection issues at present. Clotfelter et al. (13) studies the support service program offered at the University of North Carolina at Chapel Hill and finds that improving graduation rates for disadvantaged students requires both financial and non-financial support. Identifying "walk-ons" would allow for differentiation between student athletes who have financial support versus those who do not. Additionally, identifying mandated services versus voluntary services or identifying specific teams that might have different restrictions would help reduce the selection issues as well. Perfect identification is difficult to present with a subset of the student population here but each meeting contains a qualitative assessment of the meeting that occurred. Incorporating some qualitative analysis of the assessments could help to provide better evidence of the quantitative results reported here.

References

[1] Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? Technical report, National Bureau of Economic Research.

[2] Act, E. S. S. (2015). Every Student Succeeds Act (ESSA). *Pub. L*, pages 114–95.

[3] Amrein-Beardsley, A., Berliner, D. C., and Rideau, S. (2010). Cheating in the First, Second, and Third Degree: Educators' Responses to High-Stakes Testing. *Education Policy Analysis Archives*, 18:14–14.

[4] Beaver, J. K., W. L. . S. J. (2014). The Potential Effects of Opting out of State Tests in Pennsylvania. *Report 201*.

[5] Bennett, R. E. (2016). Opt Out: An Examination of Issues. Research Report. ets rr-16-13. *ETS Research Report Series*.

[6] Betebenner, D. W. and Linn, R. L. (2010). Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis, and Accountability. *Center for K-12 Assessment and Performance Management*.

[7] Bourdieu, P. (1999). Cultural Reproduction and Social Reproduction. *Modernity: Cultural Modernity*, 2(351):9781351018142–3.

[8] Bourdieu, P. (2011). The Forms of Capital.(1986). *Cultural Theory: An Anthology*, 1:81–93.

[9] CDE (2015). Colorado department of education. *Colorado Department of Education*.

[10] Center, K. P. (2021). *Assessing Colorado's K-12 Performance in the Wake of COVID*. Keystone Policy Center.

[11] Chingos, M. (2015). Who Opts Out of State Tests? *Research Report No. 113*.

[12] Clayton, G., Bingham, A. J., and Ecks, G. B. (2019). Characteristics of the Opt-Out Movement: Early Evidence for Colorado. *Education Policy Analysis Archives*, 27:33–33.

[13] Clotfelter, C. T., Hemelt, S. W., and Ladd, H. F. (2018). Multifaceted Aid for Low-Income Students and College Outcomes: Evidence from North Carolina. *Economic Inquiry*, 56(1):278–303.

[14] Cokley, K. (2006). The Impact of Racialized Schools and Racist (Mis) Education on African American Students' Academic Identity. *Addressing racism: Facilitating Cultural Competence in Mental Mealth and Educational Settings*, pages 127–144.

[15] Cook, P. J., Dodge, K., Farkas, G., Fryer, R. G., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., Steinberg, L., et al. (2014). The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in chicago. Technical report, National Bureau of Economic Research.

[16] Covell, D. and Barr, C. A. (2001). The Ties that Bind: Presidential Involvement with the Development of NCAA Division I Initial Eligibility Legislation. *The Journal of Higher Education*, 72(4):414–452.

[17] Dee, T., Huffaker, E., Phillips, C., and Sagara, E. (2021). The Revealed Preferences for School Reopening: Evidence from Sublic-School disenrollment. Technical report, National Bureau of Economic Research.

[18] Deslile, D. S. (2015). Letter to the Honorable Robert Hammond, Colorado Commissioner of Education. *United States Department of Education*.

[19] Diliberti, M. and Kaufman, J. H. (2020). Will This School Year Be Another Casualty of the Pandemic? Key Findings from the American Educator Panels Fall 2020 COVID-19 Surveys. Data Note: Insights from the American Educator Panels. research report. rr-a168-4. *Rand Corporation*.

[20] Dilley-Knoles, J., Burnett, J. S., and Peak, K. W. (2010). Making the Grade: Academic Success in Today's Athlete. *The Sport Journal*, 13(1).

[21] Domina, T., Hashim, A., Kearney, C., Pham, L., and Smith, C. (2022). COVID-19 and the System Resilience of Public Education: A View from North Carolina. an Essay for the Learning Curve. *Urban Institute*.

[22] Dorn, E., Hancock, B., Sarakatsannis, J., and Viruleg, E. (2020). COVID-19 and Learning Loss—Disparities Grow and Students Need Help. *McKinsey & Company, December*, 8:6–7.

[23] Engzell, P., Frey, A., Verhagen, M. D., et al. (2020). Learning Inequality During the COVID-19 Pandemic.

[24] Figlio, D. N. (2006). Testing, Crime and Punishment. *Journal of Public Economics*, 90(4-5):837–851.

[25] Grossmann, M., Reckhow, S., Strunk, K. O., and Turner, M. (2021). All States Close but Red Districts Reopen: The Politics of In-Person Schooling During the COVID-19 Pandemic. *Educational Researcher*, 50(9):637–648.

[26] Gunn, E. L. and Eddy, J. P. (1989). Student Services for Intercollegiate Athletes. *College Student Affairs Journal*, 9(3):36–44.

[27] Heck, R. H. and Takahashi, R. (2006). Examining the Impact of Proposition 48 on Graduation Rates in Division 1a Football and Program Recruiting Behavior: Testing a Policy Change Model. *Educational Policy*, 20(4):587–614.

[28] Hollis, L. P. (2001). Service Ace? Which Academic Services and Resources Truly Benefit Student Athletes. *Journal of College Student Retention: Research, Theory & Practice*, 3(3):265–284.

[Jack et al.] Jack, R., Halloran, C., Okun, J., and Oster, E. Community Case Rate Data by District. *COVID-19 School Data Hub [CSDH]*.

[30] Jack, R., Halloran, C., Okun, J., and Oster, E. (2022). Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts. *American Economic Review: Insights*.

[31] Johnston, M. (2014). Giving Colorado's Teacher Evaluation Bill Time to Succeed. *The Denver Post*.

[32] Kaufman, J. H. and Diliberti, M. K. (2021). Divergent and Inequitable Teaching and Learning Pathways During (and Perhaps Beyond) the Pandemic: Key Findings from the American Educator Panels Spring 2021 COVID-19 surveys. data note: Insights from the american educator panels. research report. rr-a168-6. *RAND Corporation*.

[33] Kendi, I. X. (2016a). Why Standardized Tests have Standardized Postracial Ideology. *Academe*, 102(6):26–29.

[34] Kendi, I. X. (2016b). Why the Academic Achievement Gap is a Racist Idea. *Black Perspectives*.

[35] Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., and Liu, J. (2020). Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement. *Educational Researcher*, 49(8):549–565.

[36] Kurmann, A. and Lalé, E. (2022). School Closures and Effective In-Person Learning During COVID-19: When, Where, and for Whom. *Where, and for Whom (February 4, 2022)*.

[37] Lamont, M. and Lareau, A. (1988). Cultural Capital: Allusions, Gaps, and Glissandos in Recent Theoretical Developments. *Sociological Theory*, pages 153–168.

[38] Lareau, A. and Calarco, J. M. (2012). Class, Cultural Capital, and Institutions: The Case of Families and Schools. *Facing social class: How societal rank influences interaction*, pages 61–86.

[39] Lavecchia, A. M., Oreopoulos, P., and Brown, R. S. (2020). Long-Run Effects from Comprehensive Student Support: Evidence from Pathways to Education. *American Economic Review: Insights*, 2(2):209–24.

[40] Linn, R. L., Baker, E. L., and Betebenner, D. W. (2002). Accountability Systems: Implications of Requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6):3–16.

[41] Maloney, M. T. and McCormick, R. E. (1993). An Examination of the Role that Intercollegiate Athletic Participation Plays in Academic Achievement: Athletes' Feats in the Classroom. *Journal of Human Resources*, pages 555–570.

[42] Meyer, S. K. (2005). NCAA Academic Reforms: Maintaining the Balance Between Academics and Athletics. In *Phi Kappa Phi Forum*, volume 85, page 15. National Forum: Phi Kappa Phi Journal.

[43] Mithaug, D. E. (1996). *Equal Opportunity Theory*. Sage.

[44] Munley, V. G., Garvey, E., and McConnell, M. J. (2010). The Effectiveness of Peer Tutoring on Student Achievement at the University Level. *American Economic Review*, 100(2):277–82.

[45] National Academies of Sciences, E., Medicine, et al. (2020). *Reopening K-12 schools during the COVID-19 pandemic: Prioritizing health, equity, and communities*. National Academies Press.

[46] Phifer, T. (2015). CSU Will Provide Stipends to Student-Athletes.

[47] Pizmony-Levy, O. and Cosman, B. (2017). How Americans View the Opt Out Movement.

[48] Pizmony-Levy, O. and Green Saraisky, N. (2016). Who opts out and why? Results from a National Survey on Opting Out of Standardized Tests.

[49] Routon, P. and Walker, J. K. (2015). Student-Athletes? The Impact of Intercollegiate Sports Participation on Academic Outcomes. *Eastern Economic Journal*, 41(4):592–611.

[50] The Leadership Conference on Civil and Human Rights (2015). Civil Rights Groups: "We Oppose Anti-Testing Efforts". *Retrieved from http://www.civilrights.org/press/2015/anti-testing- efforts.html?referrer=http://www .usnews.com/opinion/articles/2016-05-09/who-does-the- movement-to-opt-out-of-standardized-testing-help*.

[51] USDE (2015). Fact sheet: Testing Action Plan. *USDE*.

[52] Wang, Y. (2017). The Social Networks and Paradoxes of the Opt-Out Movement Amid the Common Core State Standards Implementation. *Education Policy Analysis Archives*, 25.

[53] Weiss, M. J., Ratledge, A., Sommo, C., and Gupta, H. (2019). Supporting Community College Students from Start to Degree Completion: Long-Term Evidence from a Randomized Trial of CUNY's ASAP. *American Economic Journal: Applied Economics*, 11(3):253–97.

[54] Woodhall, M. (1987). Human Capital Concepts. In *Economics of education*, pages 21–24. Elsevier.

Appendix

Table 3.7: The Effect of In-Person Schooling on District-Level Pass Rates with Demographic Interactions

		Math			ELA	
	Pass Rate	Pass Rate	Pass Rate	Pass Rate	Pass Rate	Pass Rate
% IP * ’21	0.036	0.055	0.028	-0.004	0.005	0.005
	(0.044)	(0.058)	(0.060)	(0.030)	(0.040)	(0.044)
% Hyb * ’21	0.007	-0.018	-0.060	0.001	-0.035	-0.068
	(0.044)	(0.072)	(0.0803)	(0.039)	(0.060)	(0.069)
% Black * IP * ’21	-0.044			0.619		
	(0.785)			(0.932)		
% Black * Hyb * ’21	0.178			0.013		
	(0.681)			(0.821)		
% Hisp * IP * ’21		-0.096			-0.040	
		(0.129)			(0.090)	
% Hisp * Hyb * ’21		-0.068			-0.040	
		(0.171)			(0.150)	
% FRPL * IP * ’21			0.005			-0.000
			(0.114)			(-0.01)
% FRPL * Hyb* ’21			0.106			0.117
			(0.165)			(0.156)
CZ X Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	614	614	614	634	634	634

Results from specification (2.5). Table 3.7 incorporates the same controls used in specification (2.4) but now I allow effects of in-person shares to vary by race. Standard errors are clustered by district and are reported in parentheses. Robust Standard errors clustered at the District level in parentheses.

Table 3.8: The Effect of In-Person Schooling on School-Level Pass Rates with Demographic Interactions

	Math			ELA		
	Pass Rate	Pass Rate	Pass Rate	Pass Rate	Pass Rate	Pass Rate
% IP * '21	0.062**	0.078**	0.055	0.021	0.037	-0.003
	(0.024)	(0.036)	(0.048)	(0.039)	(0.051)	(0.060)
% Hyb * '21	0.032	0.023	0.031	-0.013	-0.007	0.003
	(0.020)	(0.029)	(0.026)	(0.040)	(0.035)	(0.038)
% Black * IP * '21	-0.339*			-0.340		
	(0.173)			(0.359)		
% Black * Hyb * '21	0.107			0.231		
	(0.272)			(0.330)		
% Hisp * IP * '21		-0.032			-0.071	
		(0.041)			(0.068)	
% Hisp * Hyb * '21		0.099***			-0.064*	
		(0.031)			(0.038)	
% FRPL * IP * '21			0.008			0.023
			(0.048)			(0.051)
% FRPL * Hyb* '21			0.030			-0.053*
			(0.033)			(0.029)
Observations	6659	6659	6659	9128	9128	9128

Table 3.8 incorporates the same controls used in Table 2.5 but now I allow effects of race to vary by in-person shares in earlier years. Standard errors are clustered by district and are reported in parentheses. All regressions are weighted by school enrollment. Robust Standard errors clustered at the District level in parentheses.

Table 3.9: Effect of Receiving Support Services on Semester GPA and Credit Earned Ratio (Before COVID-19)

	Semester GPA	Ratio
Mentoring	0.152**	0.001
	(0.084)	(0.026)
Study Hall	0.060	-0.001
	(0.067)	(0.021)
Tutoring	-0.224***	-0.019*
	(0.037)	(0.007)
Academic Skill Dev	-0.001	0.006
	(0.096)	(0.023)
Observations	935	938

Results are from specification (3.2). Table 3.9 includes individual and semester fixed effects. The table includes the analysis for the sample period that ends in Fall 2019 to account for the changing grading policy in the Spring 2020 semester. Robust Standard errors clustered at the individual level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
.

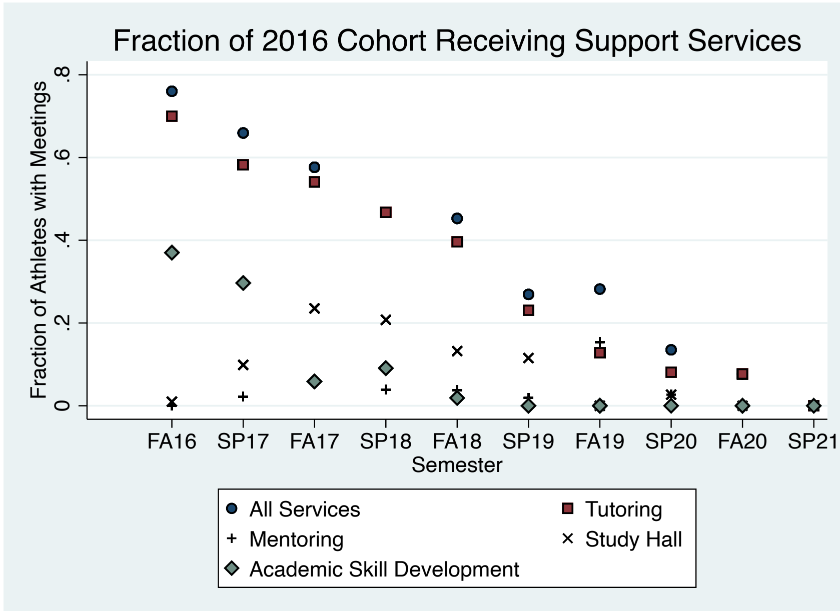

Figure 3.3: Fall 16 Support Service Meetings by Type Over Time

Figures 3.3 and 3.4 illustrate the percentage of each student athlete cohort who receive each type of support service over the course of the sample

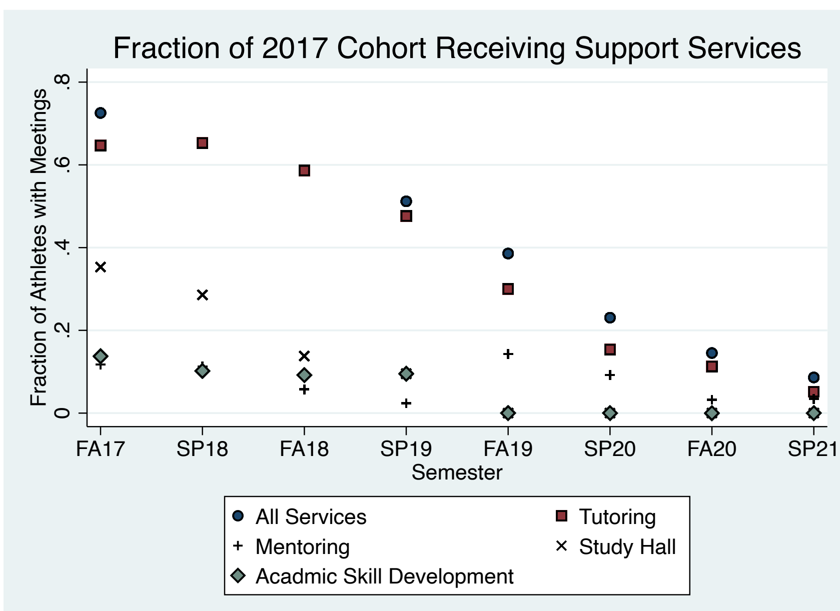

Figure 3.4: Fall 17 Support Service Meetings by Type Over Time